



Turun yliopisto  
University of Turku

# **KONKURSSIN ENNUSTAMISMALLIN LAATIMINEN TILASTOLLISIN MENETELMIN**

**Lineaarinen erotteluanalyysi, logistinen regressio ja yleistetty lineaarinen sekamalli**

Liiketaloustiede, rahoituksen ja laskenta-  
toimen pro gradu -tutkielma

Laatija  
Olli Laakso

Ohjaajat:  
KTT Erkki Vuorenmaa  
KTM Minna Vähäsalo

3.10.2016  
Pori



Turun kauppakorkeakoulu • Turku School of Economics



## SISÄLLYS

1	JOHDANTO .....	7
1.1	Tutkielman taustaa .....	7
1.2	Tutkielman tavoitteet ja rajaukset .....	9
1.3	Tutkimusote.....	11
2	KONKURSSIN ENNUSTAMINEN.....	12
2.1	Tilinpäätösraportointi ja konkurssi.....	12
2.2	Tilastollisen konkurssin ennustamistutkimuksen taustaa.....	14
2.2.1	William H. Beaver (1966).....	14
2.2.2	Edward I. Altman (1968) .....	17
2.2.3	Logit-mallia ja sen muunnelmia käyttäneet tutkimukset .....	20
2.3	Konkurssin ennustamisen vaihtoehtoiset menetelmät ja tutkimusten meta-analyysit .....	23
2.4	PK-yritykset ja niiden konkurssin ennustaminen.....	27
3	TUTKIELMAN METODIT JA AINEISTO .....	31
3.1	Lineaarinen erotteluanalyysi .....	31
3.2	Logistinen regressio .....	35
3.3	Yleistetty lineaarinen sekamalli (GLMM) .....	38
3.4	Aineiston keruu ja muokkaus.....	41
3.5	Ennustamismallien laatimisessa käytetyt potentiaaliset tunnusluvut.....	44
3.6	Estimointiryhmän ja testiryhmän laatiminen .....	48
4	TUTKIMUSTULOKSET .....	49
4.1	Aineiston kuvailu sekä konkurssiyritysten ja toimivien yritysten vertailu ..	49
4.2	Ennustamismallien laatiminen .....	53
4.2.1	Laaditut mallit ja ennustamistarkkuudet .....	55
4.2.2	ROC-analyysi.....	62
4.2.3	Altmanin Z-mallin soveltaminen aineistossa.....	66
5	JOHTOPÄÄTÖKSET JA YHTEENVETO .....	68

LÄHTEET 74

## LIITTEET

LIITE 1	SPSS-TULOSTEET TAULUKON 6 ENNUSTAMISMALLEISSA .....	78
LIITE 2	TUTKIELMAN AINEISTON KONKURSSIYRITYKSET .....	98
LIITE 3	TUTKIELMAN AINEISTON TOIMIVAT YRITYKSET .....	102

## **KUVIOT**

Kuvio 1	Erotteluanalyysin Z-arvojen jakauma kahdelle ryhmälle.....	32
Kuvio 2	Kahden ryhmän erotteluanalyysin graafinen esitys.....	33
Kuvio 3	Esimerkit logistisen regression erottelukyvystä.....	37
Kuvio 4	Ennustamismallien ROC-käyrät testiryhmässä.....	63

## **TAULUKOT**

Taulukko 1	Suomalaisten yritysten liikevaihdot 2014.....	29
Taulukko 2	Alustavien tunnuslukujen laskentakaavat.....	45
Taulukko 3	Aineiston yritysten kokoluokat.....	49
Taulukko 4	Aineiston yritysten toimialaluokitukset.....	50
Taulukko 5	Estimointiryhmän tunnuslukuvertailu toimivien ja konkurssiyri- tysten välillä.....	51
Taulukko 6	Laadittujen ennustamismallien vakiotermit, kertoimet ja tarkkuudet estimointi- ja testiryhmissä.....	56
Taulukko 7	Testiryhmässä parhaiten suoriutuneen ennustamismallin tarkkuuden ristiintaulukointi.....	57



# 1 JOHDANTO

## 1.1 Tutkielman taustaa

Yrityksen ajautuminen konkurssiin on ilmiö, joka on kiinnostanut tutkijoita jo vuosikymmenien ajan. Yrityksen konkurssin ennustaminen on vallitseva osa-alue konkurssija käsittelevissä tutkimuksissa ja kirjallisuudessa. Mielenkiinto ilmiön ennustamista kohtaan ei ole yllättävää, sillä yrityksillä on useita sidosryhmiä, joilla on erilaisia intressejä yrityksiin liittyen. Oman ja vieraan pääoman ehtoiset rahoittajat, potentiaaliset uudet rahoittajat, valtio, asiakkaat, tavaran ja palveluiden toimittajat, henkilöstö, tilintarkastajat ja analyytikot ovat kaikki kiinnostuneita konkurssin tai rahoituskriisin uhasta.

Mahdollista konkurssiin ajautumista on tarpeen arvioida esimerkiksi liikepankin lainapäätöksen yhteydessä sekä tilintarkastuksessa *going concern* -periaatteen toteutumisen arvioinnin yhteydessä. Lisäksi yrityksen konkurssiin asettamisesta päätettäessä tuomioistuimen on tässä ennustamistilannetta muistuttavassa kontekstissa syytä arvioida yrityksen konkurssiin asettamisen tarpeellisuutta, jonka pääsääntöisenä edellytyksenä on muu kuin tilapäinen maksukyvyttömyys.

Konkurssin ennustamismalleja laativissa tutkimuksissa esiintyy monenlaisia menetelmiä valintoja, joista yhtenä keskeisimmistä on käytettävä tilastollinen tai muu menetelmä. Konkurssin *ennustamismenetelmällä* tarkoitetaan käytettävää menetelmää kokonaisuutena. Konkurssin *ennustamismallilla* puolestaan tarkoitetaan menetelmän tuottamaa funktiota tai algoritmia, joka jakaa yritykset konkurssiyrityksiin ja toimiviin yrityksiin esimerkiksi vuoden aikahorisontilla nykyhetkestä lähtien. Yhteistä useimmille menetelmille on se, että ne tekevät yritysten jaottelun yhden tai useamman selittävän muuttujan perusteella. Selittävät muuttujat ovat olleet enimmäkseen kirjanpidosta johdettuja taloudellisia tunnuslukuja, mutta myös markkinaperusteisia selittäviä muuttujia on käytetty tutkimuksissa.

Konkurssin ennustaminen on aihealue, jossa tutkimusta voidaan tehdä mielekkäästi määrällisesti paljon. Uusia menetelmiä voidaan kehittää ja soveltaa uusilla tavoilla, mutta jo olemassa olevia menetelmien tuottamia malleja voidaan myös parannella vaihtamalla esimerkiksi selittäviä muuttujia tai niiden painoarvoja. Lisäksi yritysten toimintaympäristö muuttuu ajan kuluessa, jolloin samat lainalaisuudet eivät välttämättä enää päde kuin silloin, kun joku tietty ennustamismalli on laadittu. On syytä huomioida esimerkiksi yleisen korkotason vaikutus eri ajankohtina. 10 %:n *omavaraisuusaste* yhdistettynä 3 %:n *kokonaispääoman tuotto prosenttiin* ei ole (*ceteris paribus*) yhtä huolestuttavaa yrityksen elinkelpoisuuden kannalta silloin, kun yleinen korkotaso on 2 %:n tasolla verrattuna esimerkiksi 4 %:n korkotasoon. Lisäksi eri ennustamismalleja voidaan verrata erilaisissa aineistoissa, sillä mallit ovat aina riippuvaisia käytetystä otannasta,

joka voi vaihdella esimerkiksi suomalaisista tietyn toimialan pk-yrityksistä amerikkalaisiin pörssiyrityksiin.

Tutkielmassa myöhemmin käsiteltävässä William H. Beaverin (1966) tutkimuksessa, jossa ennustetaan konkurssia tilinpäätöksistä johdettujen tunnuslukujen avulla, todetaan kyseisen tutkimuksen olevan perimmältä olemukseltaan pelkän tunnuslukututkimuksen sijaan pikemminkin taloudellisen raportoinnin ennustamiskyvyn empiirinen testaaminen tiettyä ajanjaksona, jolloin tietyt taloudellisen raportoinnin säännökset ovat voimassa. Taloudellista raportointia voidaan arvioida sen luoman hyödyn avulla ja sen luomaa hyötyä voidaan arvioida sen kyvyllä ennustaa tulevia olennaisia tapahtumia, joista yksi on yrityksen toiminnan lakkaaminen esimerkiksi konkurssin myötä. (Beaver 1966, 72–99.)

Näillä samoilla perusteilla voidaan arvioida Beaverin tutkimuksen lisäksi myös muita konkurssin ennustamistutkimuksia, joiden havaintoaineistot ja tutkimusmenetelmät ovat varsin vaihtelevia tarjoten informaatiota erilaisista metodisista ja empiirisistä konteksteista. Lisäksi konkurssin ennustamistutkimuksella voi olla instrumentaalinen arvo, mikäli siinä laaditaan konkurssin ennustamismalli, joka on toimivuudeltaan parempi tietyssä yrityspopulaatiossa kuin aikaisemmat mallit.

Vertailemalla konkurssin ennustamistutkimuksissa saavutettuja ennustamistarkkuuksia on selvästi havaittavissa, että pienten yritysten konkurssin ennustaminen on haastavampaa kuin keskikokoisten tai suurten yritysten (ks. esim. Edmister 1972 ja Pompe & Bilderbeek 2005). Ei ole kuitenkaan selvää, missä kokoluokassa on mahdollista saavuttaa esimerkiksi aikaisemmissa tutkimuksissa saavutettu keskimääräinen ennustamistarkkuus tietyllä ennustamismenetelmällä. Tämän selvittämiseksi on tarpeen valita selkeä kokorajaus, joka on mahdollisimman alhainen, mutta kuitenkin riittävän korkea toimivan ennustamismallin laatimiseksi. Tällainen kokorajaus voi olla esimerkiksi tilintarkastusvelvollisuuden rajat ylittävät yritykset.

Kyseinen rajaus on kahdesta näkökulmasta merkittävä. Ensinnäkin tilintarkastusvelvollisten yritysten tilinpäätösinformaation voidaan olettaa olevan luotettavampaa, sillä ulkopuolinen taho on tarkastanut tilinpäätöksen. Toiseksi rajaus tuo mukanaan lisärelevanssia tilintarkastajien päätöksentekijäryhmälle konkurssin ennustamisen kontekstissa, sillä tällöin malli on laadittu pääsääntöisesti tilintarkastusvelvollisten yritysten tilinpäätösinformaation perusteella.

Konkurssin ennustamistutkimuksista tehtyjen meta-analyysien (ks. esim. Aziz & Dar 2006 ja Jackson & Wood 2013) perusteella käytetyimpiä ennustamismenetelmiä ovat olleet *lineaarinen erotteluanalyysi* (linear discriminant analysis) ja *logistinen regressio* (logistic regression) eli *logit-malli*. *Yleistetty lineaarinen sekamalli* (generalized linear

mixed model)<sup>1</sup> on yläkäsite edistyneille tilastollisille menetelmille, jotka huomioivat epälineaarisia satunnaisvaikutuksia mallien matemaattisen mallintamisen yhteydessä. GLMM-menetelmässä on mukana hyvin yleisesti käytetyn logistisen regression hyvät puolet, mutta lisäksi sen matemaattisessa rakenteessa on ominaisuuksia, jotka puoltavat sen soveltuvuutta konkurssin ennustamiseen. Sen avulla on mahdollista esimerkiksi mallintaa eri toimialoihin liittyvää heterogeenisyyttä ennustamismallin laatimisen yhteydessä. Sitä on kuitenkin käytetty hyvin rajoitetusti konkurssin ennustamisessa eikä sen kaikkia ominaisuuksia ole hyödynnetty.

Konkurssin ennustamista käsittelevässä kirjallisuudessa esiintyy lukuisia vaihtoehtoisia konkurssin ennustamismenetelmiä, joilla on omat vahvuutensa ja heikkoutensa. Meta-analyysien perusteella esimerkiksi *neuroverkot* (neural networks) ovat olleet keskimäärin hieman tarkempia kuin lineaarinen erotteluanalyysi ja logistinen regressio, mutta neuroverkkojen perusongelmana on ennustamismallin matemaattinen monimutkaisuus, jonka seurauksena menetelmä ei tuota siirrettävissä olevaa ja intuitiivisesti tulkittavissa olevaa funktiota toisin kuin yleisemmin käytetyt menetelmät lineaarinen erotteluanalyysi ja logistinen regressio. Näistä kahdesta menetelmästä logistinen regressio lienee nykyään suosituimpi, joten sitä voidaan ainakin jossain määrin pitää konkurssin ennustamisen yhteydessä menetelmällisenä normina tai standardina.

GLMM-menetelmällä voidaan laajentaa tavallista logistista regressiota säilyttämällä kuitenkin sen toimintaperiaatteen perusulottuvuudet. Näin ollen, mikäli GLMM-menetelmällä on saavutettavissa parempi ennustamistarkkuus kuin logistisella regressiolla, on vaikea perustella perinteisen logistisen regression käyttöä tilanteessa, jossa GLMM-menetelmän käyttö on mahdollista. Tästä johtuen GLMM-menetelmän testaaminen empiirisesti on erityisen tarpeellista, sillä toimiessaan sillä on edellytykset muodostua uudeksi menetelmälliseksi normiksi konkurssin ennustamisen yhteydessä.

## 1.2 Tutkielman tavoitteet ja rajaukset

Tutkielman tavoitteet voidaan jakaa kahteen erilaiseen tasoon. Ensimmäinen tavoitteena on selvittää, kuinka hyvin suomalaisten tietyn kokoluokan ylittävien yritysten viime vuosina esiintyneitä konkursseja on mahdollista ennustaa tilinpäätöksistä johdettujen taloudellisten tunnuslukujen avulla. Toiseksi tavoitteena on verrata kolmella erilaisella tilastollisella metodilla laadittavia ennustamismalleja ja niiden ennustamistarkkuuksia.

---

<sup>1</sup> Tutkielmassa käytetään toiston välttämiseksi termiä *GLMM-menetelmä* tai pelkkä *GLMM* kuvaamaan luvussa 3.3 määriteltyä tapaa soveltaa yleistettyä lineaarista sekamallia.

Yksi tutkielmassa käytettävistä tilastollisista menetelmistä on GLMM-menetelmä. Aikaisemmissa konkurssin ennustamistutkimuksissa tätä menetelmää ei ole tutkielman laatijan käsityksen mukaan käytetty kuin kerran aikaisemmin (Dakovic ym. 2010) ja siinäkin sen ominaisuuksia suppeammin hyödyntäen kuin tässä tutkielmassa tehdään. Menetelmää ei ole myöskään mainittu konkurssitutkimuksia käsittelevissä meta-analyyseissa, joten sen testaamisella empiirisen aineiston avulla on oma aikaisempaan kirjallisuuteen kontribuutiota luova vaikutuksensa. GLMM-menetelmän lisäksi tutkielmassa käytetään laajasti käytössä olevia menetelmiä eli lineaarista erotteluanalyysia ja logistista regressiota, jotka tarjoavat vertailukohtaan uutuusarvoa omaavalle erityisen mielenkiinnon kohteena olevalle GLMM-menetelmälle.

Ennustamismallien laatiminen on keskeinen osa tutkielmaa, sillä selittäviä muuttujia ei ole ennalta määrätty minkään aikaisemman mallin mukaan. Sen sijaan ne selvitetään aineiston ja tilastollisten analyysien avulla. Tavoitteena on laatia matemaattinen funktio, jonka avulla perusjoukon muodostavien yritysten konkurssesja voi ennustaa mahdollisimman hyvin.

Empiirisen aineiston koostuessa suomalaisista yrityksistä on tarpeen huomioida perusjoukon yksikköjen ominaisuuksia. Esimerkiksi suomalaisten pörssiyritysten konkurssien ennustaminen ei ilmiön harvinaisuudesta johtuen onnistu uskottavasti tilastollisin menetelmin. Toisaalta hyvin pienten yritysten konkurssien ennustaminen on myös ongelmallista johtuen esimerkiksi siitä, että omistajavetoisten mikroyritysten rahoitusrakenteet poikkeavat usein suuremmista yrityksistä ja niissä tapahtuvia muutoksia on vaikeampi ennakoita.<sup>2</sup> Lisäksi hyvin pienet yritykset eivät ole tilintarkastusvelvollisia, joten niiden tilinpäätösten luotettavuus ei ole lähtökohtaisesti samalla tasolla kuin tilintarkastusvelvollisilla yrityksillä.

Tutkielmassa rajataan aineiston havaintoyksikköjen eli yritysten tilinpäätösten määrää siten, että kaikkien tilintarkastuslaissa määritettyjen kokorajojen tulee täytyä sinä tilikautena, jonka perusteella konkurssin ennustamisessa käytettävä tilinpäätös on laadittu. Konkurssiin ajautuneiden yritysten tilinpäätösten saatavuus on toinen otoskokoa rajaava tekijä. Aineisto kerätään Suomen Asiakastieto Oy:n ylläpitämästä Voitto+tietokannasta. Aineiston keruuta käsitellään tarkemmin luvuissa 3.4–3.6.

Tutkielmassa testataan lisäksi tunnetuimman konkurssin ennustamismallin eli Altmanin (1968) Z-mallin toimivuutta tutkielman aineistossa. Näin tehdään, jotta tutkielmassa laadittavien mallien ennustamistarkkuudet saavat vertailukohtaan mahdollistaen niiden toimivuuden kokonaisvaltaisemman arvioinnin.

---

<sup>2</sup> Aihetta käsitellään tarkemmin luvussa 2.4.

### 1.3 Tutkimusote

Tutkielma on luonteeltaan kvantitatiivinen, empiirinen sekä tilinpäätösanalyttinen konkurssin ennustamistutkimus. Tutkielmassa laaditaan ja testataan erilaisia matemaattisia funktioita, joiden avulla yrityksiä jaotellaan sekä *ex post* (mallin rakentaminen) että *ex ante* (mallin testaaminen) toimiviksi yrityksiksi ja konkurssiyrityksiksi perustuen niiden kaupparekisteriin toimitettuihin tilinpäätöstietoihin ja niistä johdettuihin taloudellisiin tunnuslukuihin.

*Ex post* -jaottelulla tarkoitetaan matemaattisen funktion – tai konkurssin ennustamismallin – laatimista ja sen jälkeen tämän funktion avulla niiden samojen yritysten luokittelua toimiviksi yrityksiksi ja konkurssiyrityksiksi, joiden taloudellisten tunnuslukujen perusteella kyseinen funktio on laadittu. Tätä yritysjoukkoa kutsutaan yleensä konkurssin ennustamisen yhteydessä *estimointiryhmäksi*.

*Ex ante* -jaottelu puolestaan viittaa laaditun funktion testaamiseen mallin laatimisesta riippumattomiin yrityksiin sen selvittämiseksi, miten hyvin mallin voidaan odottaa toimivan estimointiryhmän ulkopuolella. Tästä aineistosta käytetään yleensä nimitystä *testiryhmä*. Toimivuutta testiryhmässä voitaneen pitää tärkeämpänä kuin estimointiryhmässä, sillä se indikoi paremmin sitä, kuinka hyvin mallin voidaan odottaa toimivan tutkimuksen aineiston ulkopuolella. Tutkielmassa käytettyä jakoperustetta estimointi- ja testiryhmiin käsitellään luvussa 3.6.

Metodisessa mielessä tutkielma on kolmen vaihtoehdoisen tilastollisen konkurssin ennustamismenetelmän komparatiivinen analyysi. Siinä kuvaillaan menetelmien erilaisia toiminnallisia ominaisuuksia ja verrataan niiden toimivuutta tutkielman empiirisessä aineistossa.

## 2 KONKURSSIN ENNUSTAMINEN

### 2.1 Tilinpäätösraportointi ja konkurssi

Ulkoisen laskentatoimen kokonaisvaltaista teoreettista viitekehystä ei ole tähän päivään mennessä saatu kehitettyä. Sen sijaan aihetta käsittelevässä kirjallisuudessa esiintyy useita eri näkökulmista lähteviä keskitason teorioita, joiden kehittelyä jatketaan yhä. (Riahi-Belkaoui 2004, 84.)

Tilinpäätökset ovat keskeinen osa kokonaisuutta ulkoisessa laskentatoimessa ja taloudellisessa raportoinnissa. Siten myös tilinpäätösten funktiot ja niiden painopisteet on eri teorioissa nähty eri tavoin. Tilinpäätösraportoinnin voidaan katsoa olevan kolmen eri ryhmän vuorovaikutuksen ja niiden erilaisten intressien tulos. Nämä ryhmät ovat: (1) tilinpäätöksiä laativat yritykset, (2) tilinpäätöksiä käyttävät tahot, kuten osakkeenomistajat, vieraan pääoman rahoittajat ja valtion virastot sekä (3) ulkoisen laskentatoimen ammattikunta, jolla tarkoitetaan tässä yhteydessä erityisesti tilintarkastajia, jotka varmentavat tilinpäätösten säännösten mukaisuuden. (Riahi-Belkaoui 2004, 164–165.)

Tietyn ryhmän tarpeita korostettaessa esimerkiksi tilinpäätösstandardien laadinnassa saattaa toisen ryhmän tilinpäätösraportoinnista saatavat hyödyt vähentyä tai haitat lisääntyä. Esimerkiksi todella tiukat vaatimukset raportoinnin sisällön suhteen myös pienten yritysten kohdalla saattaisi palvella lainapäätöstä harkitsevan rahoittajan etua, mutta se voisi lisäksi aiheuttaa kohtuuttomasti kustannuksia pienille yrityksille taloudellisen raportoinnin viedessä enemmän resursseja.

International Accounting Standards Board (IASB) on parhaillaan päivittämässä omaa taloudellisen raportoinnin käsitteellistä viitekehystään. Ottaen huomioon IASB:n rooli kansainvälisten tilinpäätösstandardien laatijana, voidaan järjestön laatimaa viitekehystä pitää relevanttina arvioitaessa, mitä taloudelliselta raportoinnilta ja sen myötä myös tilinpäätöksiltä yleisesti odotetaan. Toukokuussa 2015 julkaistussa käsitteellisen viitekehysten luonnoksessa IASB luonnehtii taloudellisen raportoinnin tavoitteita seuraavasti:

*”The objective of general purpose financial reporting is to provide financial information about the reporting entity that is useful to existing and potential investors, lenders and other creditors in making decisions about providing resources to the entity. Those decisions involve buying, selling or holding equity and debt instruments, and providing or settling loans and other forms of credit.”*

Kuvauksesta on luontevaa johtaa tulkinta, jonka mukaan tilinpäätöksen yksi käyttötarkoitus on auttaa oman ja vieraan pääoman nykyisiä ja potentiaalisia uusia sijoittajia

arvioimaan sitä, onko yritys vaarassa ajautua konkurssiin. Konkurssin uhan toteutuessa, rahoittajat jäävät yleensä ilman sijoittamaansa pääomaa kokonaan tai osittain. Mikäli konkurssi pystytään ennustamaan ennen sen toteutumista, voidaan taloudellisia (ja henkisiä) resursseja säästää suuremmilta menetyksiltä verrattuna siihen, että lopulta konkurssiin johtavaa toimintaa jatketaan ja rahoitetaan edelleen. Näin ollen konkurssit voidaan nähdä niin taloudellisten kuin henkistenkin resurssien uudelleenallokoinnin impulssina.

Beaver ym. (2005, 94–118) havaitsivat listattujen yhtiöiden tilinpäätöksistä johdettujen taloudellisten tunnuslukujen konkurssin ennustamiskyvyn heikentyneen hieman ajan myötä aikavälillä 1962–2002, mutta markkinaperusteisilla selittäville muuttujilla tämä lievä heikentyminen oli kompensoitavissa. Tunnuslukujen ennustamiskykyyn on Beaverin ym. mukaan saattanut vaikuttaa tilinpäätösstandardien kehitys, aineettomien hyödykkeiden ja johdannaisten lisääntynyt merkitys sekä lisääntynyt harkinnanvaraisuus tilinpäätöksiin sisältyvien erien arvostuksissa. Tutkimus osoittaa ainakin listattujen yritysten osalta, että tunnuslukujen selityskyky ei ole vakio eri ajankohtina.

Konkurssin käsite on määritelty suomalaisessa lainsäädännössä konkurssilaissa (KonkL 1:1). Sen mukaan konkurssi on velallisen kaikkia velkoja koskeva maksukyvyttömyysmenettely, jossa velallisen omaisuus käytetään konkurssisaatavien maksuun. Lisäksi pykälässä todetaan, että velallisen omaisuus siirtyy konkurssin alkaessa velkojien määräysvaltaan ja että velallisen omaisuuden hoitamista ja myymistä sekä muuta konkurssipesän hallintoa varten on tuomioistuimen määräämä pesänhoitaja.

Konkurssilakia koskevan hallituksen esityksen (HE 26/2003) mukaan konkurssilain määritelmä ei ole legaalimääritelmä vaan kuvaileva. Konkurssimenettelyn käytännön kannalta määritelmää on syytä laajentaa toteamalla sen olevan lakisääteinen järjestely, jossa tuomioistuimen määräämä pesänhoitaja muuttaa velallisen omaisuuden nimenomaisesti rahaksi ja maksaa kertyneillä varoilla kaikki velallisen velat tai ainakin siinä määrin kuin ne siihen riittävät.

Konkurssilain 2:1.1:ssä säädetään konkurssiin asettamisen edellytykseksi maksukyvyttömyys, jota täsmennetään 2:1.2:ssä siten, että velallisen tulee olla muuten kuin tilapäisesti kykenemätön maksamaan velkojaan niiden erääntyessä. Tämä edellytys on voimassa siitä riippumatta, hakeeko konkurssia velallinen itse vai tämän velkoja. Maksukyvyttömyyden arvioiminen ei ole täysin yksiselitteistä, joten tilannekohtainen harkinta on tarpeen (Koulu 2009, 153–154). Maksukyvyttömyyden arvioinnissa olisi mahdollista käyttää konkurssin ennustamismallia ainakin tukena kokonaisvaltaisessa arvioinnissa. Parhaimmillaan malliin sisältyy runsaasti relevanttia empiiristä dataa helposti hyödynnettävän funktion muodossa, jota päätöksen maksukyvyttömyydestä tekevällä taholla ei muuten välttämättä olisi käytettävissään hyödyllisellä tavalla valmiiksi analysoituna. Muun muassa insolvenssioikeuden professorina toiminut Risto Koulu (2009, 103) toteaa, että taloustieteen konkurssitutkimus on tuottanut suoraan oikeustieteessä

hyödynnettäviä tuloksia erityisesti maksukykyanalyysien muodossa, johon myös tämän tutkielman painopiste kohdistuu.

## 2.2 Tilastollisen konkurssin ennustamistutkimuksen taustaa

Tässä luvussa käsiteltävät konkurssin ennustamistutkimukset<sup>3</sup> on valittu laajasta tutkimusten kirjosta niiden tutkielman tutkimusasetelmaan liittyvän relevanssin perusteella. Kronologisessa järjestyksessä etenevä tutkimusten käsittely kattaa sen metodisen pohjan, jonka varaan tutkielman lähestymistapa on rakennettu näiden aikaisempien tutkimusten metodisia valintoja yhdistäen, muokaten ja laajentaen.

Eryityisesti Beaverin (1966) ja Altmanin (1968) tutkimukset ovat alan keskeisimpiä julkaisuja, joita käsitellään tästä syystä tutkielmassa tarkemmin. Näiden tutkimusten tutkimusasetelmiin sisältyi monia valintoja, joista osa perustui tutkijoiden omaan harkintaan ja osa aineiston saatavuuteen. Monet näistä valinnoista ovat esiintyneet vastavina myöhemmissä tutkimuksissa, mutta jotkut niistä ovat olleet myös enemmän tai vähemmän tutkimuskohtaisia. Nämä lukuisat tutkimusasetelmaan liittyvät valinnat – joille on olemassa myös vaihtoehtoisia tapoja – ovat aiheen mielenkiintoisuuden ohella yksi syy siihen, että aihetta on voitu tutkia edelleen ja sen kirjallisuuteen on pystytty tuomaan jotain uutta kontribuutiota useassa tutkimuksessa myös näiden urauurtavien tutkimusten jälkeen.

### 2.2.1 *William H. Beaver (1966)*

Viime vuosikymmeninä tilinpäätösanalyysiin perustuvissa konkurssin tai rahoituskriisin ennustamistutkimuksissa on mainittu lähes poikkeuksetta William H. Beaverin (1966) tutkimus. Aihetta oli tutkittu myös aikaisemmin muiden muassa Merwinin (1942) toimesta, mutta Beaverin tutkimuksessa oli tiettyjä uusia metodisia valintoja, jotka ovat tehneet siitä laajasti viitatus myöhemmissä konkurssin ennustamistutkimuksissa. Erityisesti *vastinparimenettelyn* käyttäminen aineiston keräämisessä ja kerätyn aineiston ja-

---

<sup>3</sup> Ennustettavasta ilmiöstä käytettävä termi ja ilmiön määritelmä vaihtelevat jonkin verran tutkimuksittain. Joissain tutkimuksissa ennustetaan yrityksen epäonnistumista tai rahoituskriisiä. Lisäksi joissain tutkimuksissa saattaa ennustettavana olla konkurssin lisäksi muita maksukyvyttömyyttä indikoivia tilanteita, kuten lakisääteinen saneeraus. Tutkielmassa puhutaan toiston välttämiseksi *konkurssin ennustamistutkimuksista* tarkoittaen kaikkia näitä variaatioita ennustettavan ilmiön termistössä ja sen määritelmässä.

kaminen *estimointi-* ja *testiryhmiin*<sup>4</sup> olivat sellaisia ratkaisuja, joita on sittemmin käytetty toistuvasti alan tutkimuksissa.

Beaver käytti yksittäisiä tilinpäätöksistä johdettuja tunnuslukuja yhdysvaltalaisen listattujen teollisuusyritysten epäonnistumisen ennustamiseen. Yritys määriteltiin epäonnistuneeksi, mikäli se oli konkurssissa, ei ollut kyennyt maksamaan joukkovelkakirjan sitoumuksia, ei ollut pystynyt maksamaan etuoikeutettua osinkoa tai se oli ylittänyt pankkitilinsä. Tutkimuksen havaintoaineisto koostui 79 epäonnistuneesta yrityksestä<sup>5</sup> ja 79 vastinparimenettelyllä valitusta toimivasta yrityksestä.

Epäonnistuneille yrityksille valittiin vastinpari samalta Yhdysvaltain kauppaministeriön määrittämältä toimialalta (Standard Industrial Classification system of the United States Department of Commerce) ja samasta kokoluokasta käyttäen koon mittarina taseen loppusummaa. Taseen loppusumman keskiarvo oli epäonnistuneilla yrityksillä 6,3 miljoonaa dollaria ja toimivilla yrityksillä 8,5 miljoonaa dollaria. Aineiston saatavuudesta johtuen Beaverin tutkimuksen ulkopuolelle jäivät muun muassa yksityisomistuksessa olevat yritykset ja muut kuin teolliset yritykset, jonka hän myöntää olevan vastahakoisesti tehty valinta. Samalla hän muistuttaa tulosten yleistettyden pätevän lähtökohtaisesti vain samankaltaiseen yrityspopulaatioon kuin tutkimuksessa käytettiin. (Beaver 1968, 71–75.)

Tutkimuksessa käytettiin 30 tilinpäätöksestä johdettavissa olevaa tunnuslukua, joiden valintaperusteena Beaver käytti kolmea kriteeriä, joista yhden piti täytyä. Tunnusluvun piti olla yleisesti käytetty alan kirjallisuudessa, sen piti olla suoriutunut hyvin aikaisemmissa tutkimuksissa tai sen piti perustua kassavirtaan.<sup>6</sup> Tunnuslukujen keskiarvoja vertailtiin toimivien ja epäonnistuneiden yritysten välillä viitenä vuotena ennen epäonnistuneen yrityksen epäonnistumisen hetkeä. Viimeisin tilinpäätös epäonnistuneella yrityksellä sai olla (tarkemmin määrittelemättömästä syystä) enintään kuusi kuukautta vanha epäonnistumisen hetkellä. Käyttämällä vastinparimenettelyä pyrittiin vähentämään toimialan ja yrityksen koon vaikutusta tunnuslukuihin. Tunnuslukujen keskiarvoissa oli huomattavia eroja toimivien ja epäonnistuneiden yritysten välillä, erityisesti käytettäessä viimeisen tilinpäätöksen tunnuslukuja ennen epäonnistumista. (Beaver 1968, 78–82.)

---

<sup>4</sup> Osaryhmistä käytettävät termit vaihtelevat jonkin verran. Tässä tutkielmassa käytetään Laitisen (1990, 40) käyttämiä suomenkielisiä termejä *estimointiryhmä* ja *testiryhmä*.

<sup>5</sup> Epäonnistuneista yrityksistä 59 oli ajautunut konkurssiin, 16 ei ollut maksanut etuoikeutettua osinkoa, 3 ei kyennyt maksamaan joukkovelkakirjan sitoumuksia ja 1 oli ylittänyt pankkitilinsä.

<sup>6</sup> Beaverin käyttämä kassavirran käsite ei vastaa rahoituslaskelman kassavirtaa, sillä se on johdettu suori-teperusteisesta kirjanpidosta lisäämällä nettotulokseen poistot, arvonalentumiset ja luonnonvarojen vähentymiset. Suomalaisessa termistössä kyseessä on rahoitustulos (Yritystutkimus ry 2011, 62).

Tunnuslukujen keskiarvojen vertailun sijaan suurempi kontribuutio alan kirjallisuuteen Beaverin tutkimuksessa liittyi siinä käytettyyn kaksiluokkaiseen luokittelutestiin, jossa hän jaotteli yritykset toimiviin yrityksiin ja epäonnistuviin yrityksiin perustuen tutkimuksessa käytettyjen tunnuslukujen arvoihin. Hän määrittä keskiarvovertailussa hyvin suoriutuneille tunnusluvuille (suuret erot toimivien ja epäonnistuvien yritysten välillä) kriittisen raja-arvon, jota käyttämällä yritykset saatiin luokiteltua toimiviksi ja epäonnistuneiksi yrityksiksi siten, että mahdollisimman harva yritys tuli luokiteltua väärään ryhmään. Tämä oli Beaverin mukaan intuitiivisesti vetoava tapa simuloida päätöksentekotilannetta esimerkiksi pankin lainapäätöstilanteessa, jossa vaihtoehtoina on joko myöntää tai olla myöntämättä lainaa yritykselle. (Beaver 1968, 83–84.)

Jotta tutkimuksen analyysit muistuttaisivat mahdollisimman läheisesti todellista edellä mainittua päätöksentekotilannetta, Beaver jakoi aineiston satunnaisesti kahteen toimivia ja epäonnistuneita yrityksiä sisältävään osaryhmään, joille hän ensin määrittä tunnuslukujen kriittiset raja-arvot ryhmien sisällä. Tämän jälkeen hän käytti osaryhmissä määritettyjä raja-arvoja toisen ryhmän yritysten – jotka olivat toisesta osaryhmästä riippumattomia – luokitteluun toimiviksi ja epäonnistuneiksi yrityksiksi. (Beaver 1966, 84.) Tämänkaltainen aineiston jakaminen kahteen osaan on ollut Beaverin tutkimuksen jälkeen varsin yleistä alan tutkimuksissa.

Yritys voi tulla luokiteltua väärin kahdella eri tavalla: epäonnistunut yritys luokitellaan toimivaksi yritykseksi tai toimiva yritys luokitellaan epäonnistuneeksi yritykseksi. Ensin mainitusta virheestä Beaver käytti termiä virhetyyppi I ja jälkimmäisestä virhetyyppi II. Kokonaisluokitteluvirhe puolestaan saadaan laskemalla virhetyypit I ja II yhteen, joka voidaan ilmoittaa joko absoluuttisena tai suhteellisena osuutena kaikista analyyseissa mukana olleista yrityksistä.<sup>7</sup> Tutkimuksessa päästiin pienimpään kokonaisluokitteluvirheeseen käyttämällä yritysten luokitteluun tunnuslukua *rahoitustulos/vieras pääoma* (cash flow/total debt) tai *nettotulos/taseen loppusumma* (net income/total assets). Kokonaisluokitteluvirhe oli ensin mainitulla tunnusluvulla 13 % vuotta ennen

---

<sup>7</sup> Kaikissa tutkimuksissa ei ole yhtä suuri määrä epäonnistuneita (tai konkurssissa olevia) yrityksiä ja toimivia yrityksiä. Jos toimivia yrityksiä on selvästi enemmän, niiden ennustamistarkkuus on paljon merkittävämpi, mikäli kokonaisennustamistarkkuudella tarkoitetaan sitä, kuinka suuri osa *kaikista* yrityksistä luokitellaan oikein. Tutkielmassa tarkoitetaan *kokonaisennustamistarkkuudella* ja *ennustamistarkkuudella* epäonnistuneiden (tai konkurssissa olevien) yritysten ja toimivien yritysten ennustamistarkkuuksien keskiarvoa, ellei asiayhteydestä käy toisin ilmi.

epäonnistumista ja 21 % kahta vuotta ennen epäonnistumista.<sup>8</sup> Vastaavat prosentit jälkimmäisellä tunnusluvulla olivat 13 % ja 20 %. (Beaver 1968, 85–88.)

Konkurssia ennustavissa tutkimuksissa, kuten Beaverinkin tutkimuksessa, on lukuisia ennustamistarkkuuteen vaikuttavia tekijöitä. Keskeisiä eroja tähän tutkielmaan verrattuna ennustamismenetelmän lisäksi olivat muun muassa havaintoyksiköiden kokoluokat ja toimialat, aineiston käsittämä ajanjakso sekä Beaverin käyttämä sääntö, jonka mukaan tilinpäätös ei saanut olla yli kuusi kuukautta vanha yrityksen epäonnistumisen hetkellä, kun tarkastelussa oli viimeinen tilinpäätös ennen konkurssia.<sup>9</sup>

### 2.2.2 Edward I. Altman (1968)

Ehkä kaikkein tunnetuimmassa konkurssin ennustamistutkimuksessa Edward I. Altman (1968) käytti tiettävästi ensimmäisenä tilastotieteellistä monimuuttujamenetelmää konkurssin ennustamiseen tilinpäätösinformaation avulla. Aikaisemmin Wall ja Duning (1928) olivat arvioineet yrityksen suoriutumista seitsemästä tunnusluvusta laaditun painotetun summan avulla, joka muodosti siten lineaarisen funktion. Tunnuslukujen määritetyt painokertoimet perustuivat kokemukseen ja intuition, joten menetelmällä laadittua funktiota ei voida pitää tieteellisen tutkimuksen tuloksena. (Laitinen & Laitinen 2004, 84.)

Altmanin käyttämä tilastotieteellinen menetelmä oli *lineaarinen erotteluanalyysi* (linear discriminant analysis). Sitä oli hänen mukaansa käytetty pääasiassa biologiassa ja käyttäytymistieteissä sen kehittämisen lähtien 1930-luvulla, mutta myöhemmin sitä oli käytetty onnistuneesti myös muun muassa kuluttajien lainakelpoisuuden määrittämiseen (Durand 1941) ja sijoitusten luokitteluun (Walter 1959). (Altman 1968, 591.)

Lineaarista erotteluanalyysia käsitellään tarkemmin tutkielman luvussa 3.1. Lyhyesti kuvailtuna kyseessä on menetelmä, jossa konkurssin ennustamisen yhteydessä havaintoyksikön (yritys) selitettävä muuttuja (konkurssiyritys/toimiva yritys) on kategorinen ja selittävät muuttujat (yrityksen tilinpäätöksestä johdetut tunnusluvut) ovat yleensä numeerisia ja niitä voi olla useampia. Useampaan selittävään muuttujaan sisältyvä informaatio yhdistetään yhdeksi havaintoyksikkökohtaiseksi yhdistelmä-tunnusluvuksi, jonka

---

<sup>8</sup> Virhetyypin I osuus oli 22 % vuotta ennen epäonnistumista ja 34 % kahta vuotta ennen epäonnistumista. Vastaavat virhetyypin II osuudet olivat 5 % ja 8 %. Nettotulos/taseen loppusumma -tunnusluvun osalta virhetyyppien osuuksia ei mainittu.

<sup>9</sup> Tutkielmassa puhutaan eri yhteyksissä ennustamistarkkuudesta vuotta ennen konkurssia tai viimeisen tilinpäätöksen perusteella ennen konkurssia, joka voi tarkoittaa sisällöllisesti hieman eri asiaa eri tutkimuksissa johtuen tietyistä seikoista, joita käsitellään myöhemmin luvuissa 2.2.3 ja 3.4.

arvo määrittää sen, mihin selitettävän muuttujan ryhmään havaintoyksikkö luokitellaan. Konkurssin ennustamisen yhteydessä tämä tarkoittaa sitä, että yrityksen konkurssistatus (konkurssissa/toimiva yritys) pyritään selittämään mahdollisimman hyvin optimaalisella yhdistelmällä tunnuslukuja painottaen niitä optimaalisessa suhteessa. Lineaarinen erotteluanalyysi määrittää nämä painotukset muuttujille käytetyn aineiston datan perusteella.<sup>10</sup> Analyysin tuloksena on erottelufunktio, jota voidaan käyttää myös käytetyn analyysintuotteen ulkopuolisiin yrityksiin niiden konkurssin ennustamiseksi.

Altman näki tarpeen kehittää erotteluanalyysin mahdollistaman yhdistelmätunnusluvun johtuen toisaalta aikaisempien yksittäisiin tunnuslukuihin perustuvien tutkimusten ansioista ja toisaalta niiden puutteista. Näin siksi, että aikaisemmissa tutkimuksissa oli havaittu selviä eroja tunnuslukujen arvoissa toimivien yritysten ja konkurssiin ajautuneiden tai muilla tavoin heikosti suoriutuneiden yritysten välillä, mutta käytettäessä yksittäisiä tunnuslukuja yritysten suorituskyvyn arviointiin tulokset voivat olla harhaanjohtavia ja ristiriitaisia. Parhaat indikaattorit lähestyvistä konkurssista olivat olleet aikaisemmissa tutkimuksissa tunnuslukuja, jotka mittaavat kannattavuutta, lyhyen aikavälin maksuvalmiutta tai vakavaraisuutta. Altmanin mukaan näitä osa-alueita ei voitu asettaa tärkeysjärjestykseen, sillä tutkimustulokset olivat olleet tältä osin varsin vaihtelevia. (Altman 1968, 590–591.)

Tutkimuksessa mainitaan hypoteettinen tilanne, jonka tarkoituksena on havainnollistaa, miten yksittäisten tunnuslukujen käyttö jättää ulkopuolelle olennaista tietoa yrityksistä. Tämän vuoksi on perusteltavissa tarve kehittää menetelmä, joka vastaa tähän ongelmaan. Oletetaan, että yrityksellä on huono kannattavuus ja/tai vakavaraisuus, mutta sillä on keskimääräistä parempi lyhyen aikavälin maksuvalmius. Hyvästä maksuvalmiudesta johtuen tilanne ei Altmanin mukaan ehkä vaikuta niin vakavalta kuin silloin, jos se jätetään huomioimatta.<sup>11</sup> Hän sanoo olennaista olevan sen, mitkä tunnusluvut parhaiten ennakoivat tulevaa konkurssia ja miten näitä tunnuslukuja pitäisi painottaa ennustamismallin laatimisessa. (Altman 1968, 591.)

Altmanin estimointiryhmä koostui 66 yhdysvaltalaisesta listatusta teollisuusyrityksestä, joista 33 oli ajautunut konkurssiin aikavälillä 1946–1965. Konkurssiyri-tysten taaseen loppusumma oli keskimäärin 6,4 miljoonaa dollaria vaihdellen 0,7 miljoonasta 25,9 miljoonaan dollariin. Toimivat yritykset valittiin estimointiryhmässä samankaltaisella vastinparimenetelmällä kuin Beaverin (1966) tutkimuksessa siten, että konkurssiyri-tyksen pariin valittiin toimiva yritys samalta toimialalta ja samasta kokoluokasta,

<sup>10</sup> Vaikka tavoitteena on optimaalinen yhdistelmä tunnuslukuja ja niiden painotuksia, se ei tarkoita sitä, että menetelmällä on välttämättä mahdollista saavuttaa tämä tavoite.

<sup>11</sup> Tämä sinänsä intuitiivisesti vetoava näkemys ei välttämättä ole aivan yksiselitteinen, kuten tutkielman tuloksista luvussa 4.2.1 käy ilmi.

mutta muilta osin valinta tehtiin satunnaisesti. Toimivien yritysten taseen loppusumma oli keskimäärin 9,6 miljoonaa dollaria vaihdellen 1 miljoonasta 25 miljoonaan dollariin. Lisäksi aineiston pohjana olevat tilinpäätökset vastasivat ajankohdiltaan toisiaan vastinpareittain. Viimeisin analyysiin sisällytetty konkurssiyrityksen tilinpäätös oli sen vuoden tilinpäätös, jonka jälkeisellä tilikaudella yritys ajautui konkurssiin. (Altman 1968, 593.)

Testiryhmään Altman valitsi 25 konkurssiyritystä samalta aikaväliltä ja samasta koluokasta kuin estimointiryhmässä sekä 66 toimivaa yritystä, joiden nettotulos oli ollut negatiivinen vuoden 1958 tai 1961 tilinpäätöksessä. Hän perusteli tätä negatiivisen nettotuloksen kriteeriä sillä, että näin laadittu ennustamismalli joutui kovempaan testiin, sillä näiden yritysten suorituskyvyn voidaan katsoa olleen keskimääräistä heikommalla tasolla. Vuosien 1958 ja 1961 valinta perustui näiden vuosien suhteellisen heikkoon bruttokansantulon kasvuun. Kaikki yritykset olivat myös testiryhmässä yhdysvaltalaisia teollisuusyrityksiä, mutta 66 toimivaa yritystä valittiin muuten satunnaisesti, joten näiden yritysten kokoluokalla ei ollut merkitystä testiryhmää muodostettaessa. (Altman 1968, 601–602.)

Altman oli valinnut alustavasti 22 potentiaalista selittävää muuttujaa eli tunnuslukua, joiden joukosta tavoitteena oli löytää mahdollisimman hyvin toimiva yhdistelmä tunnuslukuja ja niiden painotuksia konkurssin ennustamiseen. Alustavat tunnusluvut valittiin perustuen niiden käyttöön aikaisemmassa kirjallisuudessa tai mahdollisella relevanssilla tutkimukselle. Lisäksi Altman kehitti muutamia uusia tunnuslukuja tutkimusta varten. Hän jakoi tunnusluvut viiteen eri ryhmään, jotka olivat maksuvalmius (liquidity), kannattavuus (profitability), vipuvaikutus (leverage), vakavaraisuus (solvency) ja toiminta (activity). Lukuisten tilasto-ohjelmiston suorittamien analyysien jälkeen parhaiten suoriutuvaksi erottelufunktioksi tutkimuksessa muodostui seuraava edelleen hyvin tunnettu *Altmanin Z-malli*:<sup>12</sup>

$$(kaava 1) \quad Z = 0,012X_1 + 0,014X_2 + 0,033X_3 + 0,006X_4 + 0,999X_5$$

missä

$X_1$  = nettokäyttöpääoma / taseen loppusumma

$X_2$  = kertyneet voittovarot / taseen loppusumma

$X_3$  = tulos ennen korkoja ja veroja / taseen loppusumma

$X_4$  = oman pääoman markkina-arvo / vieraan pääoman kirjanpitoarvo

---

<sup>12</sup> Z-malli on esitetty tässä alkuperäisessä muodossaan. Tätä mallia käytettäessä neljän ensimmäisen muuttujan kohdalla yrityksen tunnusluvut ovat prosenttimuodossa eli suhdeluvut on kerrottu 100:lla ennen kaavan käyttöä.

$X_5$  = myynti / taseen loppusumma

$Z$  = konkurssin uhkaa mittaava yhdistelmätunnusluku

Tunnusluvut mittaavat Altmanin mukaan kaavan mukaisessa järjestyksessä lyhyen aikavälin maksuvalmiutta, kannattavuutta pitkällä aikavälillä, kannattavuutta lyhyellä aikavälillä, vakavaraisuutta sekä yrityksen varojen kykyä tuottaa tuloja. (Altman 1968, 594–595.)

Z-mallin painokertoimet tunnusluville määritettiin aiemmin kuvatun estimointiryhmän tilinpäätösinformaation<sup>13</sup> perusteella. Estimointiryhmässä erottelufunktio sai jaoteltua yritykset 95 %:n tarkkuudella toimiviksi yrityksiksi ja konkurssiyrityksiksi käyttäen viimeisen julkaistun tilinpäätöksen dataa yrityskohtaisten Z-tunnuslukujen laskemiseen. Virhetyypin I osuus oli 6 % ja virhetyypin II osuus 3 %.<sup>14</sup> Testiryhmässä kokonaisluokitteluvirhe oli 13 % virhetyyppien I ja II luokitteluvirheiden ollessa 4 % ja 21 %. Altmanin testiryhmän toimivat yritykset koostuivat keskimääräistä heikoimmin suoriutuvista yrityksistä, jonka hän katsoi kasvattavan virhetyypin II osuutta. Altman testasi lisäksi mallin ennustamiskykyä viitenä vuotena ennen konkurssia estimointiryhmän konkurssiyrityksissä. Ennustamiskyky heikkeni merkittävästi siirryttäessä aikaisempiin tilinpäätöksiin.<sup>15</sup> (Altman 1968, 599–604.)

### 2.2.3 *Logit-mallia ja sen muunnelmia käyttäneet tutkimukset*

Chesser (1974) käytti *logit-mallia* eli *logistista regressiota* kaupallisten lainojen laiminlyömisestä ennustamiseen ja Martin (1977) pankkien epäonnistumisen todennäköisyyden arviointiin, mutta James R. Ohlsonin (1980) tutkimusta pidetään logit-mallin käytön osalta keskeisimpänä metodisesti urauurtavana tutkimuksena konkurssin ennustamisen

---

<sup>13</sup> Muuttujan  $X_4$  osalta oli lisäksi tarpeen selvittää yrityksen markkina-arvo. Altman laati myöhemmin myös sellaisen mallin, jossa oman pääoman markkina-arvo korvattiin oman pääoman kirjanpitoarvolla. Painokertoimet muuttuivat jonkin verran, mutta erottelutarkkuus säilyi lähes ennallaan. (Altman 2000, 25–26.) Asiaa käsitellään lisää luvussa 4.2.3.

<sup>14</sup> Altman käytti eri luokitteluvirheistä Beaverin (1966) tapaan nimityksiä virhetyypin I ja virhetyypin II, jossa virhetyypin I tarkoittaa väärin luokiteltua konkurssiyritystä ja virhetyypin II väärin luokiteltua toimivaa yritystä. Myös tässä tutkielmassa käytetään näitä samoja termejä.

<sup>15</sup> Tässä tutkielmassa mielenkiinnon kohteena on viimeisimmän konkurssia edeltävän tilinpäätöksen informaatio ja ennustamiskyky. Mikäli laadittu malli ennustaisi konkurssiyrityksen konkurssin esimerkiksi kolme vuotta ennen konkurssia, silloin se ei ennustaisi oikein tilanteessa, jossa arvioidaan konkurssiin ajautumista viimeistä tilinpäätöstä seuraavan ajanjakson aikana.

yhteydessä. Logit-malli ei aseta tilastotieteellisessä mielessä selittäville muuttujille niin rajoittavia vaatimuksia kuin lineaarinen erotteluanalyysi, ja lisäksi sen antama yrityksen konkurssin todennäköisyyttä mittaava arvo on intuitiivisesti helpommin tulkittavissa. Näihin asioihin palataan tutkielman luvussa 3.2 logit-mallin käsittelyn yhteydessä.

Uuden tilastollisen menetelmän käyttämisen lisäksi Ohlson toi esille erään aineiston keräämiseen liittyvän epäkohdan, jota aikaisemmissa tutkimuksissa ei ollut huomioitu. Joihinkin tutkimuksiin sisältyi Ohlsonin mukaan implisiittinen oletus, jonka mukaan tilinpäätös on ollut saatavilla tilinpäätöksen päiväyksen hetkellä, eli useimmiten 31.12.19XX.<sup>16</sup> Tämä ei tietenkään ole realistinen oletus, sillä tilinpäätös pitää ensin laatia ja tilintarkastajan pitää suorittaa tilintarkastus ennen sen julkaisua. Tutkimuksen tavoite lopulta määrittää sen, onko kyseinen oletus ongelmallinen. (Ohlson 1980, 110–113.)

Jos tarkoituksena on arvioida virallisen tilinpäätösinformaation kykyä ennustaa konkurssia, ongelma on ilmeinen. Yritys saattaa ajautua konkurssiin tilinpäätöspäivän jälkeen, mutta ennen tilinpäätöksen julkaisua. Mikäli tällaisen yrityksen tilinpäätös on asianmukaisesti laadittu, ei ole kovin yllättävää, jos sen perusteella yrityksen voidaan "ennustaa" ajautuvan konkurssiin myös tunnuslukujen perusteella sen lisäksi, että tilintarkastuskertomuksessa on todennäköisesti maininta konkurssiin ajautumisesta.

Ohlson totesi, että noin joka kuudennessa konkurssiyrityksen konkurssia edeltävän tilinpäätöksen tilintarkastuskertomuksessa oli maininta konkurssiin ajautumisesta. Näiden yritysten osalta hän otti aineistoon kyseistä tilinpäätöstä edeltävän tilinpäätöksen. Ilman tätä ratkaisua Ohlsonin laatima malli olisi näennäisesti ennustanut konkurssia noin 2 prosenttiyksikköä pienemmällä luokitteluvirheellä, mutta tutkimuksen validiteetti olisi ollut heikompi. (Ohlson 1980, 115–127.)

Tutkimuksen aineisto koostui 105 konkurssiyrityksestä ja 2 058 toimivasta yrityksestä. Siinä ei käytetty Beaverin (1966) ja Altmanin (1968) tapaan vastinparimenettelyä toimivien yritysten valitsemiseen. Ohlson ei myöskään käyttänyt testiryhmää estimointiryhmän datan perusteella laaditun ennustamismallin testaamiseen. Hän laati kolme ennustamismallia, joista yksi arvioi konkurssin todennäköisyyttä vuotta ennen konkurssia, yksi kahta vuotta ennen konkurssia ja yksi vuotta tai kahta vuotta ennen konkurssia. Selittävinä muuttujina oli yhdeksän eri muuttujaa, joista kaksi oli logit-mallin mahdollistamana kategorisia muuttujia. Käyttäen parhaan luokittelutuloksen antavaa kriittistä raja-arvoa luokittelukriteerinä Ohlsonin mallin virhetyyppien osuudet olivat 12,4 % virhetyypin I osalta ja 17,4 % virhetyypin II osalta, joten kokonaisluokitteluvirhe oli 14,9 % estimointiryhmässä vuotta ennen konkurssia. (Ohlson 1980, 110–126.)

---

<sup>16</sup> Hän arvioi syyn olevan tutkimuksissa käytetyssä aineistolähteessä (Moody's Manual), josta ei ilmene, milloin tilinpäätösinformaatio on tullut julkisesti saataville.

Jones ja Hensher (2004) kritisoivat normaalia logit-mallia sekä sen laajennusta *multinomiaalista logit-mallia*. Multinomiaalinen malli sallii useamman tuleman selitettävälle muuttujalle toisin kuin normaali binaarinen malli, jossa tulemien vaihtoehtoja on vain kaksi.<sup>17</sup> Heidän mukaansa perinteisten logit-mallien oletukset liittyen niillä laadittujen mallien havaitsematta jättämiin virhetermeihin ovat liian rajoittavia. Tiivistettynä tällä tarkoitetaan sitä osaa koko aineiston tasolla, mitä laadittu malli ei pysty selittämään. Ratkaisuksi he esittivät *mixed logit* -menetelmän käyttöä, jonka he kertovat tavoittavan paremmin havaittavissa ja havaitsemattomissa olevan heterogeenisyyden sekä yritysten sisällä että niiden välillä käyttäen ylimääräisiä parametreja verrattuna perinteisiin logit-malleihin. (Jones & Hensher 2004, 1012–1015.)

Mixed logit-mallilla on Jonesin ja Hensherin (2004) käyttämällä tavalla yhtäläisyyksiä tässä tutkielmassa käytettävään GLMM-menetelmään, mutta olennaisia erojakin löytyy. Jones ja Hensher (2004) eivät laatineet siirrettävissä olevaa funktiota, jota voisi soveltaa yksittäisiin yrityksiin aineiston ulkopuolella. Sen sijaan heidän mallinsa validointi – joka tyypillisesti tehdään käyttämällä estimointiryhmän perusteella laadittua mallia testiryhmään – tapahtui otostasolla yrityskohtaisen soveltamisen sijaan, joka edellytti tilinpäätösinformaatiota kaikista validointi-/testiryhmässä olevista yrityksistä otostasolla jo ennen testausta. (Jones & Hensher 2004, 1032–1034.) Tästä johtuen kyseisessä tutkimuksessa saavutettuja todella korkeita lähes 100 %:n ennustamistarkkuuksia ei voida verrata esimerkiksi tämän tutkielman ennustamistarkkuuksiin, sillä se ei olisi mielekästä johtuen hyvin erilaisista mallien testaamis- ja soveltamistavoista. Myöhemmin mixed logit -mallia konkurssin ennustamisessa ovat käyttäneet ainakin Trabelsi ym. (2015).

Luvussa 3.3 tarkemmin käsiteltävää GLMM-menetelmää konkurssin ennustamisessa ovat tutkielman laatijan käsityksen mukaan käyttäneet vain Dakovic ym. (2010). Heidän aineistonsa muodostui 98 421 norjalaisesta yrityksestä, joista 2 270 oli konkurssiyrityksiä. He käyttivät *GAM-mallia* (generalized additive model) epälineaaristen suhteiden selvittämiseen selittävien ja selitettävän muuttujan välillä. Tämän jälkeen he sisällyttivät ylimääräisiä muuttujia normaaliin logit-malliin<sup>18</sup> löydösten perusteella, joiden tarkoituksena oli mallintaa epälineaarisia suhteita taloudellisten tunnuslukujen ja konkurssiin ajautumisen välillä. GLMM-menetelmää käytettiin ottamalla logit-malliin mukaan toi-

---

<sup>17</sup> Konkurssin ennustamisen yhteydessä multinomiaalista mallia voidaan soveltaa siten, että yritys voi olla toimiva yritys, maksukyvytön jollain muulla (lievemällä) perusteella kuin konkurssin kautta tai se voi olla ajautunut konkurssiin. Jones ja Hensher (2004, 1020) tekivät juuri näin, jonka vuoksi kritiikki osoitettiin ensisijaisesti multinomiaaliseen logit-malliin.

<sup>18</sup> Tutkimuksessa käytetään termiä GLM (generalized linear model), joka oli tutkimuksessa käytetyssä yhteydessä verrattavissa normaaliin logit-malliin.

mialakohtainen satunnaismuuttuja vakiotermin osalta, jonka tarkoituksena oli mallintaa heterogeenisyyttä toimialojen välillä. Käyttäen mallin estimointiin dataa vuosilta 1996–1999 ja ennustaen vuonna 2000 tapahtuneita konkurssseja tutkimuksessa saatiin GLMM-menetelmällä hieman parempi ennustamistarkkuus *ROC*- (receiver operating characteristic) ja *CAP-käyräanalyysillä*<sup>19</sup> (cumulative accuracy profile) mitattuna kuin tavallisella logit-mallilla tai GAM-mallilla. He vertasivat kolmea itse kehittämäänsä mallia myös Altmanin (1968) *Z*-malliin, joka suoriutui heikommin kuin kaikki edellä mainitut mallit. Dakovic ym. eivät mainitse tutkimuksessaan varsinaisia ennustamistarkkuuksia, mutta ne voidaan päätellä *ROC*- ja *CAP*-kurvien perusteella olevan optimaalisessa kriittisen raja-arvon kohdassa hieman alle 80 % Altmanin *Z*-mallilla ja muutaman prosenttiyksikön yli 80 % muilla malleilla.<sup>20</sup> (Dakovic ym. 2010, 1741–1745.)

### 2.3 Konkurssin ennustamisen vaihtoehtoiset menetelmät ja tutkimusten meta-analyysit

Konkurssin ennustamiseen on käytetty runsaasti eri lähestymistapoja. Yksi karkea jako voidaan tehdä siten, että eritellään varhaisiin ja myöhäisiin varoittajiin keskittyvät lähestymistavat. Varhaiset varoittajat ovat konkurssiin tai rahoituskriisiin johtaneita syitä, jotka ovat luonteeltaan usein kvalitatiivisia, kuten heikoksi arvioitu yritysjohto. Myöhäiset varoittajat ovat yleensä kvantitatiivisia, kuten heikot taloudelliset tunnusluvut. Valtaosa konkurssin ennustamistutkimuksista hyödyntää vain myöhäisiä varoittajia, sillä varhaisten varoittajien selvittäminen on huomattavasti työläämpää ja käytännössä usein mahdotonta ilman erityistä laadullisen aineiston keräämisen mahdollistavaa asemaa tutkittavassa tai tutkittavissa yrityksissä.

John Argentin (1976) tutkimus edusti varhaisiin varoittajiin keskittyvää lähestymistapaa, jossa hän eritteli kolme erilaista tyyppiä konkurssiyritykselle sekä syitä, jotka kussakin tyypissä voivat olla konkurssiin johtaneita tekijöitä. Yritystyyppit olivat lyhyesti kuvailtuna seuraavat: (1) yritykset, jotka epäonnistuvat muutaman vuoden sisällä,

---

<sup>19</sup> *ROC*-analyysia käsitellään luvussa 4.2.2.

<sup>20</sup> Dakovicin ym. tutkimusasetelmassa ennustamistarkkuuteen vaikuttavia tekijöitä olivat todennäköisesti muun muassa: tilinpäätösten julkaisupäivien huomioimatta jättäminen; ääriarvojen poistaminen aineistosta; yritysten poistaminen aineistosta, mikäli tilinpäätöstä ei ollut julkaistu konkurssiin ajautumista edeltävänä vuotena (tai mikäli tilinpäätös puuttui joltain vuodelta) sekä yritysten keskimääräisesti pieni koko päätellen aineiston yritysten suuresta lukumäärästä suhteessa sen kohdemaahan Norjaan. (Dakovic ym. 2010, 1740–1741.) Viimeksi mainittu seikka todennäköisesti vaikutti mallin ennustamistarkkuuteen heikentävästi (verrattuna suurempien yritysten konkurssin ennustamiseen) ja muut seikat sitä parantavasti.

eivätkä koskaan tule kannattaviksi; (2) yritykset, jotka yrittäjä aloittaa varsin menestyneesti, mutta tulevat liian suuriksi kontrolloida sekä (3) vanhat hiipuvat yritykset, jotka ovat olleet menestyksekkäitä menneisyydessä. (Argenti 1976, 15.) Argenti (1983) kehitti myöhemmin niin sanotun *A-mallin*, jolla voidaan mitata yrityksen riskiä ajautua taloudelliseen kriisiin. Mallissa arvioitiin muun muassa yritysjohtoa, yrityksen laskentajärjestelmää, yrityksen suhtautumista muutoksiin, yrityksessä tehtyjä virheitä sekä oireita, kuten heikkoja taloudellisia tunnuslukuja. (Argenti 1983, 20.)

Myöhäisiin varoittajiin perustuvia konkurssin ennustamistutkimuksia on tehty määrällisesti niin paljon, että niistä on tehty myös meta-analyttisiä tutkimuksia, joissa tutkimuskohteena ovat olleet konkurssia ennustavat tutkimukset ja niissä käytetyt ennustamismenetelmät. Aziz ja Dar (2006, 23) jaottelivat 89 konkurssitutkimusta kattavassa meta-analyysissään menetelmät kolmeen kategoriaan, jotka olivat tilastolliset menetelmät, keinoälyyn perustuvat tietokoneavusteiset menetelmät sekä teoreettiset menetelmät. Selvästi suurin kategoria oli tilastolliset menetelmät (64 % tutkimuksista), johon tämäkin tutkielma lukeutuu. Tilastollisten menetelmien kategoriassa puolestaan selvästi käytetyimmät menetelmät olivat lineaarinen erotteluanalyysi ja logistinen regressio, jotka muodostivat vastaavassa järjestyksessä 29 % ja 22 % kaikista ennustamismalleista mukana olleissa tutkimuksissa. Jackson ja Wood (2013, 187) kartoittivat yli 350 tutkimusta ja päätyivät samankaltaiseen tulokseen menetelmien suhteellisten käyttöfrekvenssien osalta.

Ensin mainitussa meta-analyysissä menetelmiä oli yhteensä 16 ja jälkimmäisessä jopa 25. Menetelmien tarkempi kuvailu on rajattu tämän tutkielman ulkopuolelle,<sup>21</sup> koska niiden kirjo on niin laaja ja tutkielman kannalta olennaisia menetelmiä ovat lineaarinen erotteluanalyysi, logistinen regressio sekä GLMM-malli, joka ei esiinny meta-analyseissa lueteltujen menetelmien joukossa. Eri menetelmillä on luonnollisesti omat vahvuutensa ja heikkoutensa. Tutkielmassa käytettyjen edellä mainittujen menetelmien etuna verrattuna esimerkiksi keinoälyyn perustuviin tietokoneavusteisiin neuroverkkoihin<sup>22</sup> on se, että ne tuottavat suhteellisen helposti tulkittavan funktion, joka on siirrettävissä vaivatta mihin hyvänsä käyttöympäristöön, jossa yrityksen tarpeelliset tilinpäätöstiedot ovat saatavilla. Tavalla tai toisella ihmisajattelun rajoitteisiin vastaavien tyypillisten konkurssin ennustamistutkimusten lisäksi tutkimusta on tehty myös päätöksentekijän käyttäytymisestä konkurssin ennustamisessa, josta käytetään nimitystä *HIP-tutkimus* (Human Information Processing) (ks. esim. Libby 1975 ja Casey 1980).

<sup>21</sup> Ks. erinomaiset yleiskuvaukset yleisimmistä menetelmistä esim. Laitinen & Laitinen 2004, 115–168.

<sup>22</sup> Neuroverkot olivat mainituissa meta-analyseissa kolmanneksi käytetyin menetelmä konkurssin ennustamisessa lineaarisen erotteluanalyysin ja logistisen regression jälkeen.

Aziz ja Dar (2006) vertailivat meta-analyyseissaan myös eri menetelmien ennustamistarkkuuksia, joita ne olivat saavuttaneet niitä käyttäneissä tutkimuksissa keskimäärin. Lineaarista erotteluanalyysia (25 tutkimusta), logistista regressiota (19 tutkimusta), neuroverkkoja (7 tutkimusta) ja *rekursiivista osittamista* (recursive partitioning algorithm) (5 tutkimusta) lukuun ottamatta muita menetelmiä hyödyntävien tutkimusten lukumäärä oli niin vähäinen, että niiden ennustamistarkkuuksista ei voida näiden tutkimusten perusteella vielä tehdä kovin vahvoja johtopäätöksiä. Edellä mainittujen menetelmien ennustamistarkkuudet olivat vastaavassa järjestyksessä keskimäärin 85,1 %, 86,7 %, 87,4 % ja 86,4 %. (Aziz & Dar 2006, 26–27.)

Eri menetelmien ennustamistarkkuuksien vertailu eri tutkimuksissa saavutettujen tarkkuuksien perusteella on parhaimmillaan vain suuntaa antavaa, sillä tutkimusasetelmiin liittyy useita muuttujia. Näitä ovat muun muassa käytetyn aineiston havaintoyksikköjen lukumäärä, konkurssien ajankohta (sijoittuvat eri vuosikymmenille), yrityspopulaation kokoluokka ja toimialarajaus, aineiston keräämistapa (huomioidaanko esimerkiksi tilinpäätöksen julkaisupäivän ja tilinpäätöspäivämäärän ero) sekä ennustamismallin laatimistapa, joka vaikuttaa selittävien muuttujien valikoitumiseen.

Toinen tapa vertailla eri ennustamismenetelmien keskinäisiä eroja ennustamistarkkuuksissa on käyttää eri menetelmiä samassa aineistossa, jolloin edellä mainittuja eri tutkimusten tutkimusasetelmiin liittyviä muuttujia saadaan vakioitua. Hamer (1983, 297) testasi neljällä eri tunnuslukuyhdistelmällä logit-mallin, lineaarisen erotteluanalyysin sekä *kvadraattisen erotteluanalyysin*<sup>23</sup> ennustamistarkkuuksia. Tunnuslukuyhdistelmät valikoituivat aikaisempien tutkimusten ennustamismallien perusteella. Logit-mallin ennustamistarkkuus oli kolmessa neljästä tunnuslukuyhdistelmästä parempi kuin lineaarisella erotteluanalyysillä. Tilastollisesti merkitseviä eroja logit-mallin ja lineaarisen erotteluanalyysin ennustamistarkkuuksissa vuotta ennen konkurssia ei havaittu pienellä 75 yrityksen aineistolla<sup>24</sup> käyttämällä ennustamismallien tarkkuuksien selvittämiseen *Lachenbruchin validointimenetelmää*.<sup>25</sup> Ainoastaan kvadraattinen erotteluanalyysi

---

<sup>23</sup> Kvadraattinen erotteluanalyysi on selvästi vähemmän käytetty versio erotteluanalyysista kuin lineaarinen. Menetelmää kokeilivat Hamerin (1983) lisäksi ainakin Altman ym. (1977), mutta se ei toiminut toivotulla tavalla. Ongelmana oli liiallinen sopeutuminen estimointiryhmän tunnuslukuihin, joka näkyi heikkona toimivuutena testiryhmässä. Ks. esim. Johnson & Wichern (2002, 596–598) liittyen kvadraattisen erotteluanalyysin matemaattisiin ominaisuuksiin.

<sup>24</sup> Hamerin aineiston muodostaneet yritykset olivat listattuja teollisia yrityksiä. Konkurssiyritysten konkurssit sijoittuivat aikavälille 1966–1975. (Hamer 1983, 295.)

<sup>25</sup> Ks. Lachenbruch (1975, 36–37). Lachenbruchin menetelmän käyttö sopii tilanteisiin, joissa aineiston pieni koko tekee jakamisen estimointiryhmään ja testiryhmään epämielekkääksi tilastollisessa mielessä.

suoriutui tilastollisesti merkitsevästi heikommin tietyillä yhdistelmillä kuin lineaarinen erotteluanalyysi ja logit-malli. (Hamer 1983, 297–301.)

Laitinen ja Kankaanpää (1999) käyttivät vastaavanlaista lähestymistapaa tutkimuksessaan, jossa he vertailivat lineaarisen erotteluanalyysin, logit-mallin, rekursiivisen osittamisen, *eloonjäämisanalyysin* (survival analysis), neuroverkkojen sekä HIP-analyysin ennustamistarkkuuksia. Heidän aineistonsa koostui 38 konkurssiyrityksestä ja 38 toimivasta vastinparimenettelyllä valitusta yrityksestä. Konkurssiyritykset valittiin satunnaisotannalla suomalaisista yrityksistä, joiden konkurssit sijoittuivat aikavälille 1986–1989. Tutkijoiden mukaan yritykset edustivat eri toimialoja ja olivat enimmäkseen pieniä ja keskisuuria yrityksiä.<sup>26</sup> Selittäviä muuttujia oli kolme ja ne olivat kaikilla menetelmillä samat. (Laitinen & Kankaanpää 1999, 68–78.)

Laitisen ja Kankaanpään tutkimuksen tulosten mukaan parhaiten suoriutui logit-malli 10,5 %:n kokonaisluokitteluvirheellä lineaarisen erotteluanalyysin ja neuroverkkojen ollessa seuraavaksi tarkimpia menetelmiä kumpikin 13,2 %:n luokitteluvirheellä (vuotta ennen konkurssia). Eloojäämisanalyysin tarkkuus oli heikoin 23,7 %:n kokonaisluokitteluvirheellä muiden menetelmien sijoituessa mainittujen tarkkuuksien väliin. Menetelmien ennustamistarkkuuksissa ei ollut tilastollisesti merkitseviä eroja lukuun ottamatta parhaiten suoriutuneen logit-mallin ja heikoiten suoriutuneen eloonjäämisanalyysin välillä, kun tarkkuuksia vertailtiin *kaksisuuntaisella suhteellisten osuuksien vertailutestillä*. Tutkimuksen pienekö aineisto ei ollut suotuisa ajatellen havaintoyksikköjen jakamista estimointi- ja testiryhmiin, joten ennustamistarkkuudet perustuvat Lachenbruchin validointimenetelmään. (Laitinen & Kankaanpää 1999, 79–84.)

Toteuttamalla tutkimus Hamerin (1983) tai Laitisen ja Kankaanpään (1999) tavalla voidaan vertailla eri ennustamismenetelmien keskinäisiä ennustamistarkkuuksien eroja suoraviivaisemmin kuin konkurssin ennustamistutkimuksia käsittelevien meta-analyysien avulla, mutta lopullista varmuutta jonkin menetelmän yliveritaisuudesta tai heikkoudesta ei tälläkään tavalla saavuteta. Tämä johtuu erityisesti siitä, että kaikkien menetelmien käytön toteutus voidaan tehdä monella eri tavalla vaihtamalla esimerkiksi selittäviä muuttujia. Lisäksi ennustamistarkkuuksien erot saattavat vaihdella aineistoittain, vaikka selittävät muuttujat olisivat samoja.

---

<sup>26</sup> Laitisen ja Kankaanpään (1999) tutkimuksessa käytetty aineisto on ominaisuuksiltaan melko lähellä tässä tutkielmassa käytettyä aineistoa. Siinä ei tosin kuvailtu aineiston yritysten kokoluokkia tarkemmin, esimerkiksi taseen loppusumman keskiarvon tai mediaanin muodossa. Näin ollen ei voida sanoa, onko aineisto ollut täysin satunnainen otanta suomalaisista yrityksistä – jolloin suurin osa yrityksistä olisi todennäköisesti hyvin pieniä – vai käytettiinkö esimerkiksi jonkinlaista rajausta koon suhteen. Lisäksi tilinpäätöspäivän ja tilinpäätöksen julkaisupäivän eroa ei tutkimusraportissa tuotu esille, joten aineiston muodostamisen menetelmät ovat mahdollisesti eronneet tältä osin tutkielman metodeista (käsitellään luvussa 3.4).

Yhteenvedon voidaan todeta, että lineaarinen erotteluanalyysi on ollut konkurssin ennustamistutkimuksissa käytetyin menetelmä ja logit-malli toiseksi käytetyin menetelmä. Logit-malli puolestaan on ollut hieman tarkempi näistä kahdesta menetelmästä kokonaisuutena arvioiden.

## 2.4 PK-yritykset ja niiden konkurssin ennustaminen

Pienten tai pienten ja keskisuurten yritysten konkurssin ennustaminen on ollut alan kirjallisuudessa harvinaisempaa kuin suurten yritysten. Edmisterin (1972) tutkimus toi esille pienten yritysten epäonnistumisen ennustamisen vaikeuden. Hän ei onnistunut laatimaan lineaarisen erotteluanalyysin avulla ennustamismallia, joka olisi toiminut pienten yritysten epäonnistumisen ennustamisessa käyttäen yhden tilinpäätöksen sisältämää informaatiota. Sen sijaan sellaisten pienten yritysten,<sup>27</sup> joiden tilinpäätöstiedot oli saatavilla kolmelta tilikaudelta ennen viimeistä toimintapäivää, epäonnistumista oli mahdollista ennustaa estimointiryhmässä käyttäen hyväksi kolmen tilinpäätöksen sisältämää informaatiota tunnuslukuihin sisältyvien trendien ja keskiarvojen avulla. Edmisterin otanta kolmen tilinpäätöksen vaatimuksen täyttävissä yrityksissä oli kuitenkin niin pieni (42 yritystä), että aineiston jakaminen estimointi- ja testiryhmiin ei ollut mahdollista. (Edmister 1972, 1484–1491.)

Edmisterin jälkeen esimerkiksi Keasey ja Watson (1987), Laitinen (1992) sekä Huyghebaert ym. (2000) ovat tutkineet pienten yritysten konkurssin ennustamista. Aineiston saatavuus on ollut etenkin vanhemmissa tutkimuksissa ongelma, sillä edellä mainituissa tutkimuksissa konkurssiyritysten määrät niiden aineistoissa olivat vain 73, 20 ja 81 vastaavassa järjestyksessä.

Pompen ja Bilderbeekin (2005) belgialaisten pienten ja keskisuurten yritysten konkurssia ennustavan tutkimuksen aineistossa oli jopa 1 369 konkurssiyritystä, joiden konkurssit sijoittuivat aikavälille 1986–1994. Yritykset jaoteltiin vanhoihin ja nuoriin alle 8 vuotta toimineisiin yrityksiin. Vanhojen yritysten taseen loppusumman mediaani oli noin 636 000 euroiksi muutettuna ja nuorten yritysten 251 000. He saavuttivat lineaarisella erotteluanalyysillä 80 %:n ennustamistarkkuuden vanhoissa yrityksissä ja 76 %:n ennustamistarkkuuden nuorissa yrityksissä käyttäen analyyseissa viimeistä tilinpäätöstä ennen konkurssia. Vastaavat tarkkuudet käyttäen ennustamismenetelmänä neuroverkkoja olivat 81 % ja 77 %. (Pompe & Bilderbeek 2005, 851–857.) Pompen ja Bilderbeekin tutkimuksessa esiintyy riittävän suuresta aineistosta huolimatta kuitenkin sama ongelma kuin monissa muissakin alan tutkimuksissa, eli tilinpäätöksen julkista-

---

<sup>27</sup> Taseen loppusumman keskiarvo yrityksillä oli noin 165 000 dollaria (Edmister 1972, 1485).

misajankohdan huomioimatta jättäminen valikoitaessa viimeistä analyysieihin mukaan otettavaa tilinpäätöstä.<sup>28</sup>

Pienet ja keskisuuret yritykset voivat muodostaa tutkimuksen pääasiallisen aineiston myös silloin, kun tutkimuksen keskeisin tavoite tai kirjallinen kontribuutio liittyy johonkin muuhun kuin pieniin ja keskisuuriin yrityksiin tutkittavana populaationa itsessään. Näin on esimerkiksi silloin, kun tutkimuksen aineiston määrittämä perusjoukko koostuu jonkin maan kaikista yrityksistä ilman rajauksia koon suhteen ja otanta perustuu satunnaiseen otantaan. Näissä tutkimuksissa aineisto on saattanut valikoitua esimerkiksi saatavuuden perusteella, jolloin käytetyn aineiston käyttöä ei välttämättä ole perusteltu sen tarkemmin. Tällaisia tutkimuksia ovat muun muassa Dakovicin ym. (2010) sekä Laitisen ja Kankaanpään (1999) tutkimukset, joista ensin mainittua on käsitelty luvussa 2.2.3 ja jälkimmäistä luvussa 2.3.

Yhteenvedona voidaan todeta, että pienten yritysten konkurssien ennustaminen tilinpäätöksistä johdettujen tunnuslukujen avulla on vaikeampaa kuin suurempien yritysten ja informaation saatavuus ja laatu on heikompaa. Keasey ja Watson (1987, 336) toteavat tämän johtuvan ensinnäkin siitä, että suurimmalla osalla pieniä yrityksiä ei ole systemaattista sisäistä kontrollia. Toiseksi pienten yritysten tilinpäätösten julkinen rekisteröinti tapahtuu usein viiveellä. Kolmanneksi pienillä yrityksillä on kevennetyt vaatimukset tilinpäätöksen sisällön suhteen.

Herranz ym. (2009) tutkivat pienten yritysten ominaisuuksia liittyen omistajien päätöksiin koskien yritysten lakisääteistä muotoa, kokoa, pääomarakennetta sekä omistajien omia sijoituksia yrityksiin. Heidän aineistonsa pohjana toimi yhdysvaltalainen SSBF-tutkimusaineisto,<sup>29</sup> joka koostui noin 5 miljoonasta yrityksestä. Yritykset rajattiin alle 500 työntekijän yrityksiin, mutta esimerkiksi yli 50 työntekijän yritysten lukumäärä oli suhteellisesti varsin vähäinen sen ollessa noin 0,5 % rajoittamattoman vastuun yrityksissä ja noin 5 % rajoitetun vastuun yrityksissä.<sup>30</sup> (Herranz ym. 2009, 342–349.)

Tutkimuksessa havaittiin rajoitetun vastuun pienyritysten oman pääoman olevan negatiivinen lähes joka neljännessä yrityksessä riippuen vuodesta (esimerkiksi vuonna 1993 osuus oli 17,8 % ja vuonna 1998 26,3 %). Henkilökohtaisten sijoitusten osuus yritysten nettovarallisuudesta vuonna 1998 oli keskimäärin 22,2 % rajoitetun vastuun yrityksissä. Negatiivisen oman pääoman yritysten suhteellisesti suuri osuus saattoi tutkijoiden mukaan aluksi vaikuttaa yllättävältä ottaen huomioon, että hakeutumalla kon-

---

<sup>28</sup> Aihetta on käsitelty luvuissa 2.2.3 ja 3.4.

<sup>29</sup> Survey of Small Business Finance (SSBF) administered by The Board of Governors of the Federal Reserve System and the US Small Business administration (Herranz ym. 2009, 342).

<sup>30</sup> Suomalaisessa yritysjärjestelmässä rajoittamattoman vastuun yritys on esimerkiksi yksityinen elinkeinon harjoittaja ja rajoitetun vastuun yritys esimerkiksi osakeyhtiö.

kurssiin henkilökohtaisia tappioita voitaisiin välttää. He viittaavat aikaisempaan tutkimukseensa (Herranz ym. 2008), jonka tulosten perusteella tämä voi olla selitettävissä suurilla tuotto-odotuksilla ja omistajan valmiudella sijoittaa lisää pääomaa yritykseen säilyttääkseen mahdollisuuden tulevaisuudessa saataviin suuriin tuottoihin. Lisäksi suurin osa (79 % vuonna 1993 ja 89 % vuonna 1998) pienyritysten omistajista työskenteli omistamissaan yrityksissä, joten konkurssin toteutuessa he menettävät mahdollisesti sekä sijoitetut varansa että työpaikkansa. (Herranz ym. 2009, 343–353.)

Herranzin ym. (2009) tutkimus tuo hyvin esille pienelle yritykselle ominaisia piirteitä, joilla voidaan odottaa olevan vaikutuksia myös konkurssin ennustamismallin laati- miseen. Esimerkiksi suurten teollisten yritysten kohdalla omavaraisuusasteen merkityk- sen tulkinta poikkeaa hyvin pienestä omistajavetoisesta yrityksestä edellä mainituista syistä. Täten ensin mainitun yritysryhmän tilinpäätösinformaatioon perustuvan datan perusteella laaditun konkurssin ennustamismallin käyttökelpoisuus toiseksi mainitun yritysryhmän konkurssien ennustamiseen ei todennäköisesti tuota optimaalista tulosta.

Tilastokeskuksen julkaiseman tilaston pohjalta on laadittu taulukko 1, josta nähdään vuoden 2014 tilinpäätösten perustella suomalaisten yritysten jakauma liikevaihdon suu- ruuden perusteella jaoteltuina.

Taulukko 1 Suomalaisen yritysten liikevaihdot 2014 (Suomen virallinen tilasto (SVT) 2015)

Liikevaihto (1000 €)	Yritysten lukumäärä	Suhteellinen osuus
0 - 39	175878	48,37 %
40 - 99	68649	18,88 %
100 - 399	69231	19,04 %
400 - 1 999	35192	9,68 %
2 000 - 9 999	11066	3,04 %
10 000 - 39 999	2579	0,71 %
40 000 - 199 999	761	0,21 %
200 000 -	231	0,06 %
<b>Yhteensä</b>	<b>363587</b>	<b>100,00 %</b>

Taulukon 1 perusteella on luontevaa päätellä, että valtaosa esimerkiksi lainapäätök- sistä kohdistuu muihin kuin suuriin yrityksiin. Yritykset, joiden vuotuinen liikevaihto on alle 100 000 euroa, muodostivat yli kaksi kolmasosaa kaikista suomalaisista yrityk- sistä vuonna 2014. Alle 2 miljoonan liikevaihdon yritysten osuus oli jo lähes 96 % kai- kista yrityksistä.

Tutkielmassa käytetyn tilintarkastuslakiin perustuvan vähintään 200 000 euron liike- vaihdon rajauksen ylittävien yritysten osuutta suomalaisista yrityksistä ei voi taulukosta suoraan nähdä, sillä raja-arvo on kokokategorian 100 000 - 399 000 sisällä. Sen sijaan siitä on nähtävissä, kuinka olennainen vaikutus käytetyllä rajauksella on. Mikäli rajausta

ei tehtäisi lainkaan, mukana olisi pääasiassa vain todella pieniä yrityksiä, joissa korostuisi edellä käsitellyn Herranzin (2009) tutkimuksen esille tuomat seikat omistajavetoisissa hyvin pienissä yrityksissä. Rajauksesta huolimatta jäljelle jääneet yritykset ovat edelleen pääosin varsin pieniä verrattuna suuriin tai edes keskisuuriin yrityksiin. Suurimmassa osassa konkurssin ennustamistutkimuksia on käytetty keskimäärin suurempia yrityksiä kuin tässä tutkielmassa, joten niissä laadittujen mallien toimivuutta ei voida yleistää tutkielmassa rajattuun yrityspopulaatioon. Näin ollen käytetty rajaus tekee analysoitavasta aineistosta käytännössä tietynlaisen kokonsa puolesta välimaastoon sijoittuvan yrityspopulaation, jonka voidaan ajatella olevan esimerkiksi tilintarkastajien mielenkiinnon kohteena, sillä harvoja poikkeuksia lukuun ottamatta nämä yritykset ovat tilintarkastusvelvollisia.

## 3 TUTKIELMAN METODIT JA AINEISTO

### 3.1 Lineaarinen erotteluanalyysi

Lineaarinen erotteluanalyysi on tilastollinen menetelmä, joka soveltuu tilanteisiin, joissa selitettävä muuttuja on kategorinen ja selittävät muuttujat ovat numeerisia. Normaali regressioanalyysi ei tällaisessa tilanteessa tule kyseeseen selitettävän muuttujan kategorisuudesta johtuen. Erotteluanalyysiä voidaan käyttää myös tilanteissa, joissa selitettävä muuttuja koostuu useammasta kuin kahdesta ryhmästä. Eli toisin sanoen se voi saada useampia arvoja kuin kaksi (esim. huono, neutraali tai hyvä). (Hair ym. 2010, 339.)

Konkurssitutkimuksessa selitettävä muuttuja kuitenkin on binaarinen, sillä yritys on joko konkurssiyritys tai toimiva yritys. Tämän vuoksi tässä tutkielmassa keskitytään erotteluanalyysin käyttöön tapauksessa, jossa selitettävä muuttuja on kaksiluokkainen.

Erotteluanalyysin keskeinen käsite on erottelufunktio, jonka tarkoituksena on jakaa havaintoyksiköt (esim. yritykset) ryhmiin (esim. konkurssiyritykset ja toimivat yritykset) parhaalla mahdollisella tavalla käytetyn menetelmän puitteissa hyödyntäen havaintoyksikköjen selittävien muuttujien arvojen ja selitettävän muuttujan arvon välistä riippuvuutta. Menetelmän tavoitteena on laatia sellainen erottelufunktio, josta johdettujen numeeristen niin sanottujen Z-lukujen keskiarvo eroaa mahdollisimman paljon eroteltavien ryhmien välillä ja lisäksi ryhmien sisällä kyseisten arvojen varianssi on mahdollisimman pieni (Hair ym. 2010, 340). Konkurssin ennustamisen kontekstissa tämä tarkoittaa sitä, että funktio antaa konkurssiyrityksille ja toimiville yrityksille arvoja, jotka ovat keskimäärin mahdollisimman kaukana toisistaan kuitenkin siten, että myös arvojen hajonta on mahdollisimman pieni konkurssiyritysten saamien arvojen välillä sekä vastaavasti toimivien yritysten saamien arvojen välillä.

Erottelufunktion kaava konkurssin ennustamisen yhteydessä voidaan esittää seuraavalla tavalla:

$$(kaava 2) \quad Z_j = a + w_1 X_{1j} + w_2 X_{2j} + \dots + w_n X_{nj}$$

missä

$Z_j$  = erottelufunktion arvo yritykselle j

a = vakiotermi

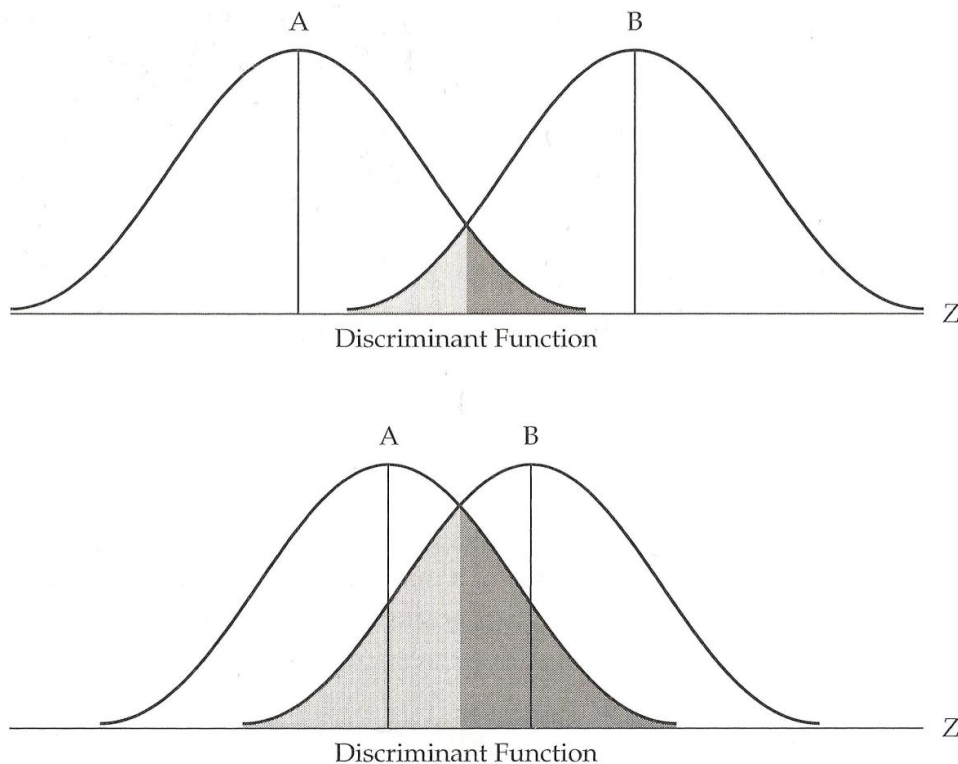
$w_i$  = selittävän muuttujan  $X_i$  painokerroin

$X_{ij}$  = selittävän muuttujan i arvo yritykselle j

Funktio on mahdollista laatia myös ilman vakiotermiä, jolloin kaavan muoto on muuten sama, mutta komponentti a jää pois. Selittävät muuttujat voivat olla konkurssia ennustettaessa esimerkiksi *nettotulos/koko pääoma*, *quick ratio* ja *omavaraisuusaste*.

Tutkimuksissa laadituille erottelufunktiolle ja muillekin kirjanpitoinformaatioon perustuvilla konkurssin ennustamismalleilla on ollut tyypillistä, että yrityksen kannattavuus, maksuvalmius ja vakavaraisuus ovat jonkun tai joidenkin selittävien muuttujien avulla huomioituna. Ajatuksena on, että useiden tunnuslukujen huomioiminen parantaa edellytyksiä ennustaa konkurssia verrattuna yksittäiseen tunnuslukuun.

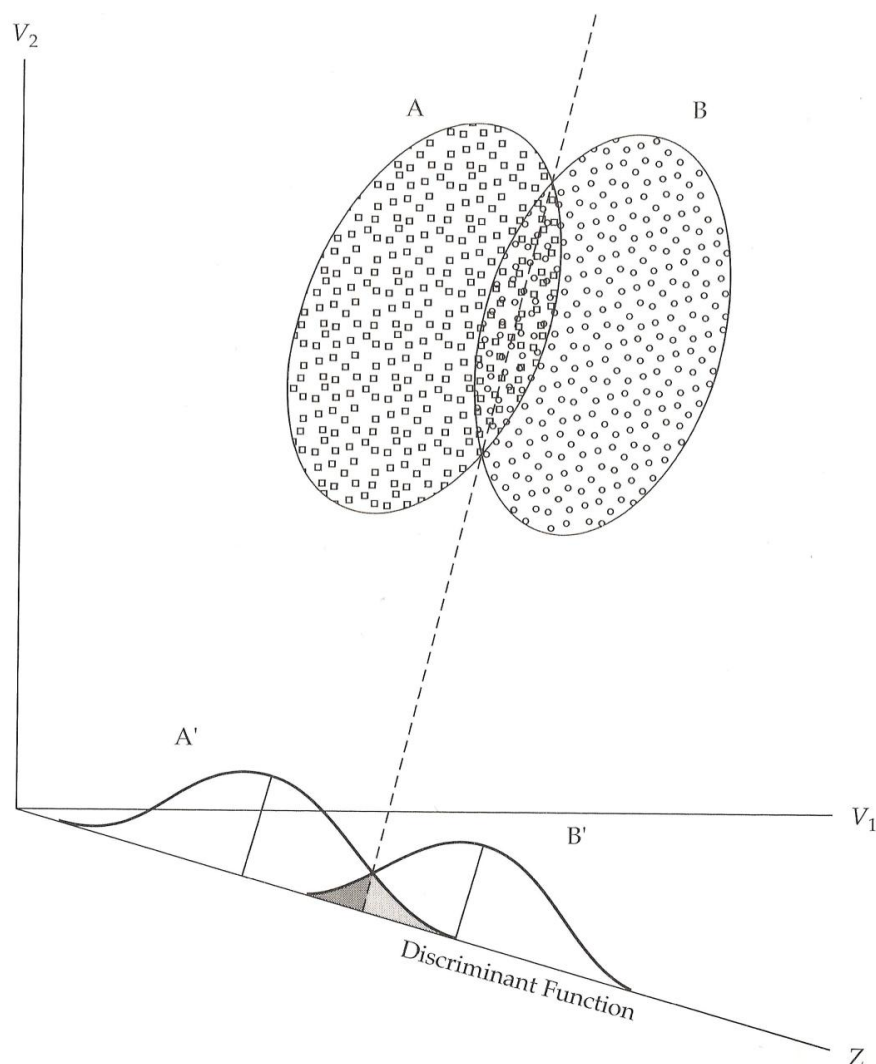
Kuviossa 1 on esitetty kahden toisensa poissulkevan ryhmän saamien erotteluanalyysin Z-arvojen jakauma kahdella erilaisella hypoteettisella erottelufunktiolla. A-ryhmän havaintoyksikköjen Z-arvot sijoittuvat vasemmanpuoleisen ja B-ryhmän oikeanpuoleisen käyrän alle. Vaaka-akseli kuvaa Z-arvoa ja pystyakseli havaintoyksikköjen frekvenssiä kunkin Z-arvon kohdalla. Konkurssin ennustamisen kontekstissa ryhmä A, jolla on keskimäärin pienempi Z-arvo, edustaa konkurssiyrityksiä ja ryhmä B toimivia yrityksiä. Asetelmasta riippuen on myös mahdollista, että suurempi Z-arvo tarkoittaisi suurempaa konkurssialttiutta, mutta käytännössä asetelma on useimmiten ensin mainitun lainen. Tämä johtuu siitä, että tutkimuksissa käytettävien selittävien tunnuslukujen suuri arvo on yleensä suotavaa yrityksen kannalta konkurssin välttämistä ajatellen (esim. omavaraisuusaste). Asia riippuu kuitenkin täysin tutkijan tekemistä valinnoista tutkimusasetelmassa.



Kuvio 1 Erotteluanalyysin Z-arvojen jakauma kahdelle ryhmälle (Hair ym. 2010, 341)

Ylemmän esimerkin erottelufunktio kuviossa 1 erottelee yritykset paremmin konkurssiyrityksiin ja toimiviin yrityksiin kuin alempi esimerkki. Yritysten Z-lukujen arvojen hajonnat ovat yhtä suuria kummassakin tapauksessa, mutta ylemmässä esimerkissä ryhmien Z-lukujen keskiarvot ovat kauempana toisistaan. Ryhmien Z-lukujen keskiarvot on merkitty pystysuorilla viivoilla. Harmaat alueet kuvioissa kuvaavat niitä havaintoyksiköjä, jotka erottelufunktio erottelisi väärin, mikäli kriittisenä raja-arvona Z-luvulle, jonka perusteella yritys luokitellaan konkurssiyritykseksi tai toimivaksi yritykseksi, pidettäisiin ryhmien Z-lukujen keskiarvojen keskiarvoa. Tämä tarkoittaa kuviossa kohtaa, jossa käyrät leikkaavat toisensa.

Kuviossa 2 on havainnollistettu graafisesti erotteluanalyysin toimintaperiaatetta tilanteessa, jossa selitettävä muuttuja on kaksiluokkainen ja selittäviä muuttujia on kaksi.



Kuvio 2 Kahden ryhmän erotteluanalyysin graafinen esitys (Hair ym. 2010, 346)

Vaaka-akseli kuvaa havaintoyksikön selittävän muuttujan  $V_1$  arvoa ja pystyakseli vastaavasti toisen selittävän muuttujan  $V_2$  arvoa. A- ja B-ellipsien sisällä olevat pisteet

kuvaavat havaintoyksikköjä, joiden saamat edellä mainittujen muuttujien arvot ovat nähtävissä etäisyyksinä vaaka- ja pystyakseleilta. A-ellipsin sisällä voisivat olla esimerkiksi konkurssiyritykset ja B-ellipsin sisällä toimivat yritykset. Z-akseli kuvaa optimaalista painotusta selittävien muuttujien  $V_1$  ja  $V_2$  välillä, jota soveltamalla yritykset voidaan erotella mahdollisimman hyvin konkurssiyrityksiin ja toimiviin yrityksiin erottelufunktiolla näiden kahden muuttujan arvojen perusteella. Kuvion 2 esimerkissä muuttujalla  $V_1$  on suurempi painoarvo kuin muuttujalla  $V_2$ . Muuttujalla  $V_1$  on positiivinen etumerkki erottelufunktiossa ja muuttujalla  $V_2$  negatiivinen. Suuremman painoarvon pystyy päättämään Z-akselin alle 45 asteen kulmasta  $V_1$ -akseliin. Muuttujien etumerkit erottelufunktiossa voi päätellä siitä, onko Z-akseli yli 90 asteen, mutta alle 270 asteen kulmassa muuttujan akseliin nähden. Tällaisessa tapauksessa etumerkki on negatiivinen.  $V_1$  voisi konkurssikontekstissa tarkoittaa esimerkiksi *nettotulosprosenttia* ja  $V_2$  *velkaantumisastetta*. Tämä olisi loogista myös muuttujien etumerkkien kannalta. Kuvioista nähdään, että selittävien muuttujien huomioiminen yhdessä mahdollistaa paremman erottelun kuin pelkästään toista muuttujaa käytettäessä.

Kuviossa 2 A' ja B' ovat kuviossa 1 esitettyjä erotteluanalyysin Z-arvojen jakaumia vastaavat käyrät. Erotteluanalyysin toimintaperiaatteen voisi esittää graafisesti vielä kolmen selittävän muuttujan tapauksessa, jolloin selittävien muuttujien arvot muodostaisivat kuution kolme dimensiota ja erottelufunktio muodostaisi tason (tasainen kaksiulotteinen pinta), joka erottelisi havaintoyksiköt kahteen joukkoon. Kyseinen graafinen esitys olisi kuitenkin haastava toteuttaa ilman kolmiulotteista esitystapaa. Yli kolmen selittävän muuttujan tapauksessa graafinen esittäminen lienee mahdotonta ainakin havainnollistavalla tavalla, sillä erottelu tapahtuu tällöin n-ulotteisessa avaruudessa, jossa n on selittävien muuttujien määrä. Toimintaperiaate on kuitenkin sama kuin kuviossa 2 myös useamman selittävän muuttujan tapauksessa.

Erotteluanalyysiin, kuten myös useimpiin muihin konkurssin ennustamismenetelmiin, kuuluu olennaisena osana aineiston jakaminen estimointiryhmään ja testiryhmään. Erottelufunktion kertoimet ja mahdollinen vakiotermin johdetaan estimointiryhmän datan pohjalta. Testiryhmä puolestaan koostuu datasta, johon estimointiryhmän datasta johdettua erottelufunktiota testataan sen toimivuuden arvioimiseksi estimointiryhmän ulkopuolella. Näin menetellään siksi, että erottelufunktio tai muu ennustamismalli saattaa toimia hyvin aineistossa, jonka perusteella se on laadittu, mutta se ei takaa sen toimivuutta käytetyn estimointiryhmän ulkopuolella. Aineiston jakosuhteesta estimointi- ja testiryhmään päättää tutkija itse. Aineiston rajallinen määrä saattaa aiheuttaa sen, että aineiston jakaminen ryhmiin ei ole mielekästä, jolloin voidaan käyttää vaihtoehtoisia menetelmiä mallin testaamiseen, kuten luvussa 2.3 mainittua *Lachenbruchin validointimenetelmää*.

## 3.2 Logistinen regressio

Logistinen regressio on tilastollinen menetelmä, joka soveltuu tilanteisiin, joissa selitettävä muuttuja on kategorinen sekä binaarinen ja selittävät muuttujat ovat numeerisia tai kategorisia. Selittävät muuttujat voivat olla myös binaarisia. Logistisen regression perusmuodossa se soveltuu vain tilanteisiin, joissa selitettävä muuttuja on kaksiluokkainen. (Peng ym. 2002, 4.) Tämä asia ei kuitenkaan konkurssin ennustamisen yhteydessä ole ongelma, sillä selitettävä muuttuja, eli yrityksen konkurssistatus, on kaksiluokkainen muuttuja. Logistisella regressiolla on paljon yhtäläisyyksiä normaalin regressioanalyysin kanssa erityisesti tilastotieteellisestä näkökulmasta. Konkurssin ennustamisen kontekstissa olennaisempia ovat kuitenkin sen yhtäläisyydet lineaarisen erotteluanalyysin kanssa.

Logistisen regression tarkoituksena konkurssia ennustettaessa on erotteluanalyysin tavoin luokitella yritykset konkurssiyrityksiksi ja toimiviksi yrityksiksi ensin estimointiryhmän datan perusteella, ja sen jälkeen soveltaa saatua ennustamismallia testiryhmään kehitetyn mallin erottelukyvyn testaamiseksi. Selittävät muuttujat ovat myös logistisessa regressiossa yrityksiä kuvaavia tunnuslukuja, jotka voivat mitata esimerkiksi maksuvalmiutta, kannattavuutta ja vakavaraisuutta. Toisin kuin erotteluanalyysissä, selittävien muuttujien arvojen ei tarvitse olla normaalisti jakautuneita, ja ne voivat olla myös kaksitai useampiluokkaisia kategorisia muuttujia. Esimerkkinä tällaisesta muuttujasta käy yrityksen maantieteellinen sijainti, joka ei luonnollisesti voi noudattaa normaalijakaumaa.

Logistisen regression antama arvo yritykselle, joka on menetelmän yrityskohtaisena arvona verrattavissa erotteluanalyysin erottelufunktion  $Z$ -arvoon, sijoittuu aina välille 0 ja 1. Tämä luku mittaa todennäköisyyttä, että yritys kuuluu toiseen kahdesta ryhmästä, kuten esimerkiksi konkurssiyritysten ryhmään. Logistisen regression matemaattisesta rakenteesta johtuen kyseinen arvo ei kuitenkaan koskaan voi saavuttaa arvoa 0 tai 1, mutta se voi olla hyvin lähellä sitä. Yritykset, jotka saavat mallista arvon  $\sim 1$  estimointiryhmän datan pohjalta laaditun mallin perusteella ovat mallin mukaan lähes varmasti konkurssiyrityksiä olettaen, että konkurssiyritykset on koodattu arvolle 1 ja toimivat yritykset arvolle 0. Teoriassa paras erottelukyky konkurssiyrityksiin ja toimiviin yrityksiin saavutetaan asettamalla kriittiseksi raja-arvoksi 0,5. Jos logistisen regression avulla laadittu malli antaa yritykselle arvon, joka on suurempi kuin 0,5, luokitellaan yritys tällöin konkurssiyritykseksi ja päinvastoin.

Logistisen regression valmiin mallin kaava ei ole yhtä selkeästi esitettävissä kuin lineaarisen erotteluanalyysin. Keskeinen funktiomuodossa oleva osa logistisen regression mallista on hyvin samankaltainen erotteluanalyysin erottelufunktion kanssa. Tämän kaavan tulos ei ole kuitenkaan heti intuitiivisesti tulkittavissa, sillä sen tuloksena on logaritminen ristitulosuhde. Asia voidaan esittää matemaattisesti seuraavalla tavalla:

$$(kaava\ 3) \quad \ln(P_j / (1 - P_j)) = a + w_1 \cdot X_{1j} + w_2 \cdot X_{2j} + \dots + w_n \cdot X_{nj}$$

missä

$P_j$  = yrityksen  $j$  konkurssin todennäköisyys

$a$  = vakiotermi

$w_i$  = selittävän muuttujan  $X_i$  painokerroin

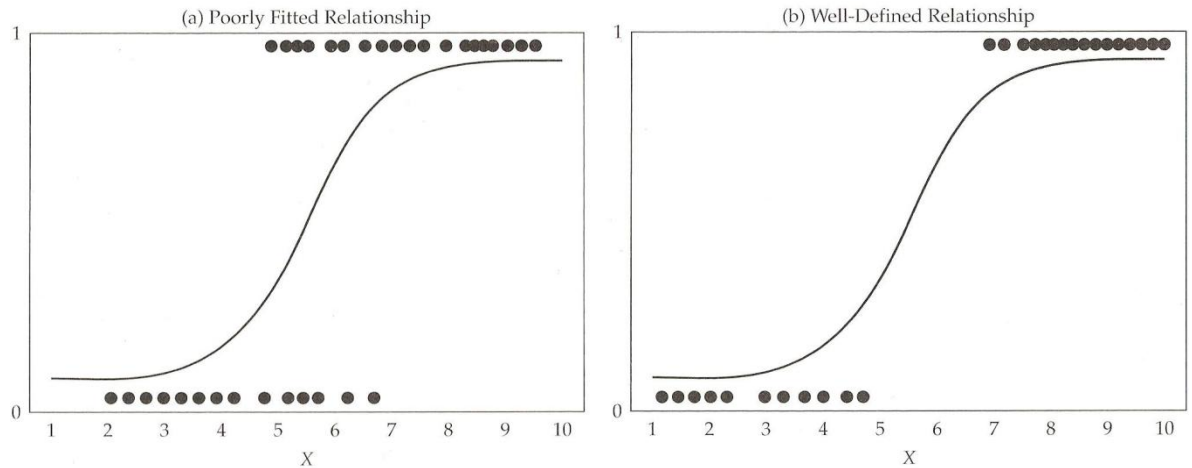
$X_{ij}$  = selittävän muuttujan  $i$  arvo yritykselle  $j$

Yhtälön oikea puoli on verrattavissa rakenteeltaan erotteluanalyysin erottelufunktion, mutta sen antama arvo ei ole verrattavissa erottelufunktion  $Z$ -arvoon. Kaavan vasemmasta puolesta nähdään, että tuloksena on logaritminen ristitulosuhde. Ristitulosuhde tarkoittaa kahden todennäköisyyden keskinäistä suhdetta. Esimerkiksi todennäköisyyksien 0,7 ja 0,3 ristitulosuhde on  $0,7/0,3 = 2,333$ . Tästä saadaan logaritminen arvo ottamalla luonnollinen logaritmi arvosta 2,333, jolloin tuloksena on noin 0,847. Näin ollen yritys, jonka konkurssin todennäköisyys on 70 %, saa kaavan 3 arvoksi noin 0,847. Jotta selville saadaan kiinnostuksen kohteena oleva yrityksen konkurssin todennäköisyyden arvo kaavan 3 oikean puolen antaman arvon perusteella, tulee se asettaa seuraavaan logistisen funktion kaavaan sulkeiden sisällä olevaan osaan:

$$(kaava\ 4) \quad P_j = \frac{1}{1 + e^{-(a + w_1 \cdot X_{1j} + w_2 \cdot X_{2j} + \dots + w_n \cdot X_{nj})}}$$

missä kaikkien komponenttien selitykset ovat samat kuin kaavassa 3 ja  $e$  = Neperin luku

Logistinen regressio ei ole lineaarinen, joten sen toimintaperiaatetta ei voida esittää graafisesti samalla tavalla kuin lineaarisen erotteluanalyysin luvussa 3.1. Linearisuuden puuttuminen ilmenee siten, että tietty lisäys jonkin selittävän muuttujan arvoon ei näy aina samansuuruisena lisäyksenä mallin antamaan ennusteeseen. Sen sijaan se riippuu siitä, missä kohdassa välillä 0 ja 1 mallin antama arvo yritykselle on ennen lisäystä jonkin selittävän muuttujan arvoon. Mikäli ennen lisäystä ollaan lähellä jompaakumpaa ääripäätä, on muutos pienempi kuin silloin, jos ollaan keskivaiheilla välillä 0 ja 1. Asia on havainnollistettavissa kuvion 3 avulla.



Kuvio 3 Esimerkit logistisen regression erottelukyvystä (Hair ym. 2010, 417)

Kuviossa on kaksi esimerkkiä logistisesta regressiosta, joista vasemmanpuoleinen erottelee ryhmät huonosti ja oikeanpuoleinen hyvin. Esimerkeissä mustat pallot ovat havaintoyksikköjä. Konkurssin ennustamisen kontekstissa havaintoyksiköt ovat yrityksiä ja toisensa poissulkevat ryhmät ovat konkurssiyrietykset ja toimivat yritykset. Ylä-laidassa olevat yritykset kuuluvat mallin ennusteen mukaan ryhmään 1 ja alalaidassa olevat yritykset ryhmään 0. Vaaka-akselilla on jokin tietty selittävä tunnusluku  $X$  tai tunnuslukuyhdistelmän painotettu summa. Tässä esimerkissä selittävä tunnusluku on "perusmuodossaan", joten tunnusluvun arvoa ei ole muokattu logaritmisestä ristitulo-suhteen muotoon.  $S$ :n muotoinen logistinen käyrä kuvastaa todennäköisyyttä  $0:n$  ja  $1:n$  välillä kussakin  $X:n$  arvon kohdassa. Esimerkiksi  $X:n$  arvo  $5,5$  tarkoittaa todennäköisyyttä  $0,5$  kuulua ryhmään 0 ja samaa todennäköisyyttä kuulua ryhmään 1.

Vasemmanpuoleisessa esimerkissä logistisella regressioanalyysillä laadittu malli erottelee yrityksiä väärin selittävän muuttujan  $X$  arvoilla, jotka ovat suunnilleen välillä  $4,9-6,8$ . Muuttujan  $X$  raja-arvo kuviossa on  $5,5$ . Alle kyseisen arvon saavat yritykset luokitellaan toiseen ryhmään ja loput toiseen ryhmään. Näin ollen malli erottelee ainakin kolme yritystä kummastakin ryhmästä väärin. Oikeanpuoleinen malli puolestaan erottelee kaikki yritykset oikein, sillä kaikkien ryhmään 0 kuuluvien yritysten selittävän muuttujan  $X$  arvo on alle  $5,5$  ja ryhmään 1 kuuluvien yli  $5,5$ . Mallin erottelukyvyn parantuessa mustat pallot siirtyvät entistä pienemmälle alueelle asteikon vasempaan alalaitaan ja oikeaan ylälaitaan.

Konkurssin ennustamisessa käytetään yleensä useampaa selittävää muuttujaa, jolloin kuviossa 3 esitetty  $X:n$  arvo on mielletävä usean selittävän muuttujan painotettuna summana kaavan 3 esittämällä tavalla. Tällöin asteikko on tosin erilainen kuin kuvion esimerkissä. Raja-arvona on logaritmisestä ristitulo-suhtetta käytettäessä 0, sillä kaavan 3 vasenta puolta soveltaen saadaan  $0,5$  todennäköisyydelle raja-arvoksi 0. Koska kaavan 3 oikea puoli on lineaarinen, voidaan se mieltää kuvion 3 muuttujaksi  $X$ . Kuvioista nähdään, että  $X:n$  kasvaessa esimerkiksi arvosta 2 arvoon 4, on muutos pystyakselin ku-

vaamassa todennäköisyydessä paljon pienempi kuin esimerkiksi  $X:n$  kasvaessa arvosta 5 arvoon 7. Selittävien muuttujien arvojen muutosten vaikutus mallin antamaan tulokseen ei siis ole yhtä suoraviivainen kuin lineaarisessa erotteluanalyysissä, jossa kaikki muutokset tietyssä selittävässä tunnusluvussa vaikuttavat aina samansuuruisesti mallin antamaan tulokseen riippumatta muiden selittävien tunnuslukujen arvoista. Logistisen regression lineaarisuuden puuttuminen sekä arvoväli nollan ja yhden välillä vaikuttavat olennaisesti kaavan 3 vakiotermin ja selittävien muuttujien painokertoimien määrittämiseen tehden sen matemaattisen rakenteen myös varsin erilaiseksi verrattuna lineaarisen erotteluanalyysiin.

### 3.3 Yleistetty lineaarinen sekamalli (GLMM)

*Yleistetty lineaarinen malli* (generalized linear model eli GLM) on tilastotieteellinen nimitys tietyntylaisille malleille, joiden avulla voidaan mallintaa muuta kuin normaalijakaumaa noudattavaa vastemuuttujaa eli selitettävää muuttujaa, kuten esimerkiksi yrityksen konkurssistatusta. Edellä käsitelty logistinen regressio (logit-malli) on yleistetyn lineaarisen mallin yksi sovellustapa. *Yleistetty lineaarinen sekamalli* (generalized linear mixed model eli GLMM) on puolestaan yleistetyn lineaarisen mallin laajennus, joka mahdollistaa kiinteiden vaikutusten lisäksi satunnaisten vaikutusten huomioimisen selittävien muuttujien ja vastemuuttujan välillä.<sup>31</sup> (Demidenko 2013, 331–332.)

Satunnaisten vaikutusten huomioiminen mallissa mahdollistaa hierarkkisen aineiston paremman analysoinnin ja siten selittävien muuttujien ja selitettävän muuttujan välisten suhteiden selvittämisen. Hierarkkisella aineistolla tarkoitetaan tässä yhteydessä sellaista aineistoa, jossa havaintoyksiköt muodostavat toisistaan poikkeavia ryhmiä jollakin selittävällä perusteella, jolla on selittävää vaikutusta selitettävään muuttuajaan.

Ryhmitteleviä tekijöitä voivat olla esimerkiksi lääkärit, markkina-alueet tai koulut, kun kiinnostuksen kohteena olevat ominaisuudet liittyvät havaintoyksikköihin, jotka voivat olla ryhmitteleviä tekijöitä vastaavassa järjestyksessä esimerkiksi potilaita, tuotteita tai oppilaita. Tiedossa saattaa olla, että ryhmittelevällä tekijällä on vaikutusta ennustettavaan ilmiöön, joka voi olla esimerkiksi potilaan saama diagnoosi, tuotteen kaupallinen menestys tai oppilaan akateeminen menestys. GLMM-menetelmän avulla on mahdollista huomioida ja eristää tämä ryhmittelevän tekijän vaikutus, jolloin on mahdollista tutkia suuremmin havaintoyksikköihin liittyvien ominaisuuksien vaikutusta ennustettavaan tai selitettävään ilmiöön.

---

<sup>31</sup> Tutkielman mielenkiinnon kohteena on vain logistisen regression ja sen sekamallilla laajennettujen versioiden soveltaminen, joten muita yleistetyn lineaarisen sekamallin soveltamistapoja ei käsitellä tutkielmassa.

Konkurssin ennustamisen kontekstissa toimiala saattaisi olla sellainen ryhmä, jolla voi olla vaikutusta yrityksen tulevan konkurssin todennäköisyyteen. Käytettäessä yleistettyä lineaarista sekamallia, jossa toimiala on luokiteltu havaintoyksikköjä ryhmittäväksi tekijäksi, voidaan selvittää monimutkaisempia suhteita selittävien ja selitettävän muuttujan välillä kuin vain esimerkiksi koodaamalla toimiala itsenäiseksi muuttujaksi, joka sisällytetään valmiiseen kaavaan.<sup>32</sup> Näin siksi, että GLMM-menetelmällä on mahdollista huomioida esimerkiksi vain tietyn selittävän muuttujan varianssien eroavaisuudet toimialoittain sen sijaan, että koko toimiala määriteltäisiin jonkin kiinteän kertoimen avulla konkurssiherkemmäksi tai vähemmän konkurssiherkäksi kuin jokin toinen toimiala. Kaikkien selittävien muuttujien konkurssin selittämiskyvyn mahdollinen riippuvuus toimialasta voidaan siis käsitellä erillisinä laadittaessa ennustamismalli GLMM-menetelmällä. Lisäksi valitun tunnuslukuyhdistelmän riippuvuus toimialasta kokonaisuutena on mahdollista huomioida ottamalla mallin määrittämisvaiheeseen mukaan vakiotermin satunnaisvaikutukset, jolloin tilanne muistuttaa läheisemmin toimialan koodaamista omana selittävänä muuttujana laadittavaan malliin.

GLMM:n toimintaperiaate voidaan esittää matemaattisesti seuraavalla tavalla:

$$(kaava 5) \quad y = X\beta + Zw + e$$

missä

- $y$  =  $N \times 1$  pystyvektori vastemuuttujia ( $N$  = havaintoyksikköjen lukumäärä)
- $X$  =  $N \times p$  matriisi selittävien muuttujien arvoja ( $p$  = selittävien muuttujien lukumäärä vakiotermin mukaan lukien)
- $\beta$  =  $p \times 1$  pystyvektori selittävien muuttujien ja vakiotermin kiinteitä kertoimia
- $Z$  = satunnaisten vaikutusten  $N \times q$  suunnittelumatriisi ( $q$  = satunnaisten vaikutusten lukumäärä)
- $w$  =  $q \times 1$  pystyvektori satunnaisten vaikutusten kertoimia (kiinteän  $\beta$  satunnainen vastine)<sup>33</sup>
- $e$  =  $N \times 1$  pystyvektori residuaaleja eli niitä osia, joita malli ei pysty selittämään

---

<sup>32</sup> Valmiin mallin käyttäminen käytännössä hankaloituu, mikäli toimiala koodataan omaksi selittäväksi muuttujaksi, sillä toimialoja on etenkin yksityiskohtaisella tasolla lukumääräisesti paljon. Lisäksi mielenkiinnon kohteena voi olla jokin yritys, jonka toimialalle ei ole määritetty kerrointa.

<sup>33</sup>  $w$ -komponenttiin sisältyy GLMM-menetelmän olennainen ero logit-malliin verrattuna. Menetelmien tarkempi matemaattinen tarkastelu on rajattu tutkielman ulkopuolelle (ks. esim. Demidenko 2013, 331–432). Tiiviisti ilmaistuna komponentti perustuu yleensä oletukseen satunnaisvaikutusten normaalijakautuneisuudesta sekä niiden varianssi-kovarianssimatriiseihin.

Jättämällä kaavasta pois  $Z_w$  saadaan normaalin logistisen regression toimintaperiaate.  $X\beta$  vastaa luvussa 3.2 esiintyvän kaavan 3 oikeaa puolta, joka on myös valmiin logit-mallin soveltamiskelpoinen kaava (joka syötetään kaavan 4 sulkeiden sisälle), kun kiinteät kertoimet on selvitetty logistisen regression avulla. Kaavassa 5 sama asia on esitetty koko aineiston tasolla, mistä johtuu  $X$ -matriisin  $N \times p$  muoto ja  $y$ -vektorin  $N \times 1$  muoto. Komponentti  $y$  koostuu konkurssin ennustamisen yhteydessä estimointiryhmän yritysten todellisista konkurssistatuksista (esim. 0 = toimiva yritys ja 1 = konkurssiyritys), jotka laadittava malli pyrkii selittämään valittujen selittävien muuttujien avulla. Sekä GLMM että logit-malli pyrkivät tekemään kaavan 5 e-komponentista – eli jäännöstermistä – mahdollisimman pienen koko aineiston tasolla.

Satunnaisten vaikutusten lukumäärä riippuu mallin laatimisessa käytetyistä oletuksista. Konkurssin ennustamisen yhteydessä niitä olisi 15, jos aineistossa on 15 toimialaa, jotka määritetään ryhmitteleväksi tekijäksi ja satunnaismuuttujaksi valitaan vain vakio-termi. Toisin sanoen vakio-termi saa vaihdella satunnaisesti toimialoittain mallin matemaattisessa laatimisessa vaikuttamatta kiinteisiin kertoimiin kokonaisuutena, joka puolestaan mahdollistaa kiinteiden kertoimien määräytymisen eri tavalla kuin ilman tätä vakio-termin satunnaisvaikutusta. Vakio-termi vastaa luvun 3.2 kaavassa 3 ikään kuin lähtökohtaa, joka on valmiin mallin antama tulos, mikäli kaikkien selittävien muuttujien eli tunnuslukujen arvo on nolla. Näin ollen edellä kuvailtu asetelma olettaisi, että tietyt toimialat ovat kokonaisuutena arvioiden herkempiä konkurssille kuin toiset, mutta tunnuslukujen suhteellinen vaikutus mallin antamaan ennusteeseen olisi kaikkien toimialojen yrityksillä sama.<sup>34</sup>

GLMM mahdollistaa satunnaismuuttujien sisällyttämisen mallin laatimisvaiheeseen varsin vapaasti. Jos esimerkiksi edelliseen esimerkkiin lisätään kolme tunnuslukua satunnaisiin vaikutuksiin, nousee satunnaisten vaikutusten lukumäärä vastaavasti. Tällöin niitä olisi jo 60 ( $4 \times 15$ ), jos vakio-termi säilytetään edelleen satunnaismuuttujana. Satunnaismuuttujien ei tarvitse olla samoja kuin kiinteiden vaikutusten muuttujat, mutta ne myös voivat olla. Yleisellä tasolla voidaan sanoa, että satunnaselementin sisällyttämisellä pyritään (ainakin tässä tutkielmassa) eristämään heterogeenisiin ryhmiin (toimialoihin) sisältyviä ominaispiirteitä, jotka eivät ole yleistettävissä koko perusjoukkoon. Näin on esimerkiksi silloin, jos jollain toimialalla *koko pääoman tuottoprosentin* varianssi ja sen selittävä vaikutus konkurssiin eroaa toisesta toimialasta. GLMM tuottaa matemaattisesti täysin samassa muodossa olevan funktion kuin normaali logit-malli (kaava 3), mutta sen matemaattinen rakenne funktion kertoimien selvittämiseksi on erilainen satunnaisten vaikutusten huomioimisesta johtuen. Komponentti  $Z_w$  vaikuttaa GLMM-menetelmän laskelmissa mallin laadinnan yhteydessä, mutta se ei näy itse en-

---

<sup>34</sup> Dakovic ym. (2010) käyttivät tällaista lähestymistapaa tutkimuksessaan.

nustamismallissa, jossa on vain kiinteitä vaikutuksia tunnuslukujen kertoimien ja vakio-termin muodossa. GLMM pitää siten sisällään logit-mallin hyväksi katsotut ominaisuudet, kuten selkeän ja intuitiivisesti helposti tulkittavan tuloksen antavan mallin tuottamisen, jota voidaan soveltaa aineiston ulkopuoliseen yksittäiseen tai useampaan yritykseen.

### 3.4 Aineiston keruu ja muokkaus

Tutkielman empiirinen aineisto kerättiin Suomen Asiakastieto Oy:n ylläpitämästä Voitto+ -tietokannasta. Tietokanta sisälsi 150 000 suomalaisen yrityksen tilinpäätöstiedot (Suomen Asiakastieto Oy 2016). Tietokannan käyttöliittymässä oli mahdollista käyttää erilaisia hakukriteerejä, joiden perusteella yrityshaun tuloksia voitiin rajata. Kuten jo mainittu, tämän tutkielman mielenkiinnon kohteena olivat yritykset, joiden kaikki tilintarkastuslain määrittämät kokorajat täyttyivät. Toisin sanoen yritysten liikevaihdon tuli olla vähintään 200 000 euroa, taseen loppusumman vähintään 100 000 euroa ja henkilöstön koon vähintään kolme henkilöä (TilintarkL 2:2).

Tosin tilintarkastuslain henkilöstön kokoon liittyvä kohta edellytti keskimäärin yli kolmea henkilöä. Hakukriteerit eivät muutenkaan täysin myötäilleet tilintarkastuslain tilintarkastusvelvollisuuden määrittämistä. Täten tutkielman havaintoaineisto on pikemminkin approksimaatio tilintarkastusvelvollisista yrityksistä kuin kirjaimellinen tulkinta. Tämä johtuu tutkimuksen toteutukseen liittyvistä käytännöllisistä syistä, kuten mahdollisimman suuren aineistokoon varmistamisesta sekä tutkimukseen käytettävissä olevista resursseista.

Pelkästään edellä mainittuja hakukriteerejä käyttämällä hakutulokseksi tuli odotetusti useita tuhansia yrityksiä. Tämän tutkielman kannalta oli olennaista löytää konkurssiin ajautuneet yritykset näiden yritysten joukosta. Voitto+ -käyttöliittymässä hakukriteerinä ei voinut käyttää konkurssiin ajautumista. Sen sijaan oli mahdollista rajata haku yrityksiin, joiden toiminta oli lakannut. Toiminnan lakkaamisen syynä saattoi olla fuusio tai muu yritysjärjestely, normaali toiminnan lopetus ilman konkurssimenettelyä tai konkurssimenettelyn kautta lopetettu toiminta. Näistä yrityksistä piti saada identifioitua yritykset, joiden toiminnan lakkaamisen syynä oli ajautuminen konkurssiin. Tämä onnistui käyttämällä Patentti- ja rekisterihallituksen ylläpitämää Virre-tietopalvelua ja sen kuulutushakua.

Virre-tietopalvelusta oli nähtävissä konkurssistatuksen lisäksi päivämäärä, jolloin yritys oli asetettu konkurssiin. Konkurssiin asettamisen päivämäärä oli tutkielman kannalta keskeinen, sillä sen perusteella määritettiin se tilinpäätös, minkä pohjalta analyysihin otettavat tunnusluvut on laskettu. Myös konkurssissa ollut yritys on saattanut

laatia tilinpäätöksen, joten automaattisesti viimeisimmän saatavilla olleen tilinpäätöksen käyttäminen ei ollut mahdollista tutkielman validiteettia ajatellen.

Aineistossa esiintyi myös sellaisia yrityksiä, joiden viimeisin tilinpäätös oli esimerkiksi tilikaudelta 2011 (tilikausi sama kuin kalenterivuosi) ja konkurssiin asettaminen oli tapahtunut vuoden 2012 maaliskuussa, mutta tilinpäätöksen julkaisupäivä oli vasta toukokuussa. Tällöin ei olisi ollut mielekasta käyttää tilikauden 2011 tilinpäätöstä aineiston lähteenä, sillä tuo tilinpäätös ei vielä konkurssiin ajautumisen hetkellä ollut julkista tietoa. Tällaisissa tapauksissa käytettiin tilikauden 2010 tilinpäätöstä tunnusluku-  
jen informaatiolähteenä. Aineistoon otettavat tilinpäätökset on pyritty valitsemaan siten, että se simuloisi mahdollisimman hyvin todellista päätöksentekotilannetta (esimerkiksi vieraan pääoman ehtoisen rahoituksen myöntämiseen liittyvä päätöksenteko). Toisin sanoen aineistoon valitun konkurssiyrityksen tilinpäätöksen on haluttu olevan sellainen, jonka julkistamishetkellä tuleva konkurssi ei vielä ollut julkisesti saatavilla olevan informaation perusteella tiedossa.

Tämä tilinpäätöksen julkaisupäivän merkitys on usein jätetty huomioimatta konkurssitutkimuksissa. Asiaa voidaan tosin tarkastella eri näkökulmista. Jos tutkimustulokset on ikään kuin suunnattu esimerkiksi tilintarkastajille, jotka näkevät alustavan tilinpäätöksen jo ennen sen julkaisua, olisi tällöin tietyin edellytyksin<sup>35</sup> validiteettia uhraamatta mahdollista käyttää edellisessä tekstikappaleessa kuvatussa tilanteessa 2011 tilinpäätöksen tunnuslukuja. Jos tutkimustulokset on kuitenkin suunnattu esimerkiksi vieraan pääoman ehtoisten rahoituspäätöstentekijöille tai analyytikoille, ei 2011 tilinpäätöksen tunnusluku-  
jen käyttö mallin rakentamisessa olisi ollut perusteltua, sillä nuo luvut eivät olisi olleet saatavilla yleisesti missään vaiheessa ennen konkurssia.

Muutamasta konkurssiyrityksestä puuttui relevantti tilinpäätös joko Voitto+tietokannasta tai sitä ei ollut rekisteröity kaupparekisteriin. Nämä yritykset jätettiin aineiston ulkopuolelle. Lisäksi muutaman konkurssiyrityksen kohdalla oli tilinpäätösten rekisteröintipäivämääristä nähtävissä, että tutkielman kannalta relevanttia tilinpäätöstä oli muokattu tai täydennetty jälkikäteen. Tutkielman laatijan ei ollut mahdollista päätellä, mitkä tunnusluvut olivat mahdollisesti muuttuneet jälkikäteen, joten tällaiset yritykset jätettiin myös analyysien ulkopuolelle validiteetin varmistamiseksi.

Tarvittavat tiedot löytyivät yhteensä 143 konkurssiyrityksen tilinpäätöksestä. Kun konkurssiyritykset ja niiden tutkielman kannalta relevantit tilinpäätökset oli identifioitu, oli seuraava vaihe etsiä toimivien yritysten tilinpäätöksiä, jotka toimivat vertailuperusteena konkurssiyritysten tilinpäätöksille. Tähän sovellettiin konkurssitutkimuksissa yleisesti käytettyä vastinparimenettelyä, jonka tavoitteena on minimoida muiden kuin tut-

---

<sup>35</sup> Konkurssi ei tällöinkään saisi olla tiedossa siinä vaiheessa, kun tilinpäätöstä olisi mahdollista analysoida ensimmäisen kerran, joka puolestaan on riippuvainen tilintarkastuksen aloittamisajankohdasta.

kittavien muuttujien vaikutusta tutkimustuloksiin. Käytännössä tämä tarkoitti sitä, että jokaiselle konkurssiyritykselle pyrittiin löytämään mahdollisimman samankaltainen yritys vertailtavaksi sillä erotuksella, että se ei ollut ajautunut konkurssiin tilinpäätöksen julkistamishetken jälkeen ennen seuraavan tilinpäätöksen julkistamishetkeä.

Ensimmäinen kriteeri konkurssiyrityksen vastinpariryityksen etsimiselle oli sama Tilastokeskuksen määrittämä toimialaluokitus (TOL 2008). Tosin joissain tapauksissa toimialaluokitukseen kuuluvat yritykset olivat niin vähälukuisia, että sopivaa vastinpariryitystä ei löytynyt. Tällöin toimialaluokitusta jouduttiin laajentamaan vähemmän tarkaksi, mutta tämä oli tarpeen vain muutamassa tapauksessa. Toinen kriteeri oli taseen loppusumma sinä tilikautena, joka vastasi konkurssiin ajautuneen yrityksen sitä tilikautta, jota käytettiin aineiston informaatiolähteenä. Vastinpariryitykseksi valittiin se yritys, jonka taseen loppusumma oli absoluuttisesti mitattuna lähimpänä konkurssiyrityksen taseen loppusummaa. Tämä koettiin esimerkiksi liikevaihtoon perustuvaan kriteeriin nähden mielekkäämmäksi, sillä taseen loppusumma karkeasti indikoi sitä varallisuutta, joka yritykseen on sidottuna ja täten uhattuna konkurssin toteutuessa.

Analyyseista ei suljettu ulkopuolelle konkurssiyrityksiä, vaikka konkurssiin asettamispäivän ja viimeisen tilinpäätöksen julkaisupäivän välillä olisi pitkäkin aika. Toimivien yritysten kohdalla sellaisia yrityksiä ei valittu vastinpareiksi, joiden tilinpäätöstä ei ollut julkaistu vuoden sisällä tilinpäätöspäivästä. Tämä logiikka myötäilee aiemmin kuvattua, eli tutkielmassa ei haluttu perustaa aineiston dataa sellaiseen tietoon, jota ei ollut oikeasti saatavilla tilanteessa tai tilanteissa, joita tutkielman asetelmassa pyritään simuloimaan. Lisäksi saneerauksessa olevia tai konkurssiin myöhemmin ajautuneita toimivia yrityksiä ei valittu vastinpareiksi, sillä toimivien yritysten haluttiin edustavan nimenomaan toimivia yrityksiä. Muita tarkentavia kriteerejä ei käytetty, jotta otannan satunnaisuus toteutuisi ja tilastollinen yleistettävyyss säilyisi mahdollisimman hyvin.

Vastinparimenetelmän käyttäminen ei tarkoita sitä, että yrityksiä verrattaisiin analyyseissa suoranaisesti pareittain. Tutkielman tavoitteena oli laatia yleispätevä ennustamismalli, joka ei ole riippuvainen esimerkiksi toimialasta tai yrityksen koosta. Menetelmällä pyritään eliminoimaan muiden kuin mitattavien muuttujien vaikutusta.

Hypoteettisena esimerkkinä voidaan esittää tilanne, jossa toimialan A tunnusluvun X mediaani on 2, toimialan B 4 ja toimialan C 6. Jos toimialan A konkurssiyrityksen vastinpariksi valikoituisi toimialan C yritys, niin tämä vääristäisi konkurssiyrityksen ja toimivan yrityksen välistä selittävää vaikutusta tunnuslukuun, sillä osa siitä voisi selittyä toimialojen mediaanieroilla. Tämä puolestaan heikentäisi datan pohjalta laaditun mallin ennustamistarkkuutta tunnusluvun suuruuden suhteessa välimaastossa olevien toimialan B yritysten kohdalla (*ceteris paribus*), johon saattaisi lukeutua enemmistö yrityksistä. Esimerkki on hypoteettinen, mutta sen tarkoituksena on perustella vastinparimenettelyn käyttöä. Lisäksi se voi auttaa havainnollistamaan problematiikkaa, joka liittyy heterogeenisen perusjoukon tutkimiseen tilastollisin menetelmin. GLMM-

menetelmän käytöllä voidaan mahdollisesti lieventää tarvetta vastinparimenettelyn käyttöön, mutta tutkielmassa tätä mahdollisuutta ei tutkittu empiirisesti.

### 3.5 Ennustamismallien laatimisessa käytetyt potentiaaliset tunnusluvut

Tutkielman ennustamismallien laatimisessa käytetyt alustavat potentiaaliset tunnusluvut valikoituivat osittain käytetyn pääasiallisen aineistolähteen perusteella, joka oli Suomen Asiakastieto Oy:n ylläpitämä Voitto+ -tietokanta. Tietokannassa oli laskettu 18 tunnuslukua jokaiselle yritykselle viideltä viimeiseltä tilikaudelta. Näiden tunnuslukujen ensisijainen käyttö mallien laatimisessa oli perusteltua huomioiden tutkielmaan käytettävissä olevat resurssit ja vaadittavan manuaalisen työn määrä muiden kuin valmiiksi lasketujen tunnuslukujen laskemisessa jokaiselle aineiston yritykselle. Kaikkia aineistossa saatavilla olevaa 18 tunnuslukua ei otettu mukaan analyysiin. Esimerkiksi ostovelkojen ja myyntisaamisten kiertoaikoja tai henkilöstömäärään suhteutettua liikevaihtoa ei arvioitu sellaisiksi muuttujiksi, joiden sisällyttäminen malleihin parantaisi ennustamistarkkuutta.

Altmanin (1968) Z-mallin tunnuslukujen laskeminen oli tarpeen, jotta tutkielmassa laadittujen mallien ennustamistarkkuutta voitaisiin verrata tunnetuimman konkurssin ennustamismallin tarkkuuteen samassa aineistossa. Näiden tunnuslukujen arvot piti laskea manuaalisesti aineiston yritysten tuloslaskelmien ja taseiden pohjalta. Taulukossa 2 on esitetty Voitto+ -tietokannan tunnusluvut sekä Altmanin tunnusluvut kaavoineen,<sup>36</sup> jotka otettiin mukaan mallien laatimisvaiheeseen.

Taulukossa nähtävien tunnuslukujen lisäksi tutkielmassa tehtiin alustavia kokeiluja tunnusluvulla, joka vertaa yrityksen *liikevoittoprosenttia* toimialan mediaaniin.<sup>37</sup> Se ei kuitenkaan toiminut yhtä hyvin alustavissa malleissa kuin pelkkä *liikevoittoprosentti*, joten sitä ei sisällytetty lopulliseen mallien laatimisvaiheeseen. Lisäksi tällaisen tunnusluvun mukana olo olisi heikentänyt ennustamismallin helppokäyttöisyyttä, sillä sen käyttäminen vaatisi tietoa tilinpäätöksen ulkopuolelta (toimialan liikevoittoprosentin mediaanin).

---

<sup>36</sup> Altmanin Z-mallin muuttujan X4 kohdalla laskettiin listaamattomille yrityksille tarkoitettu versio tunnusluvusta, jota käsitellään luvussa 4.2.3.

<sup>37</sup> Toimialojen keskiarvoja konkurssin ennustamisessa ovat hyödyntäneet ainakin Platt ja Platt (1990).

Taulukko 2 Alustavien tunnuslukujen laskentakaavat

<b>Tunnusluku</b>	<b>Kaava</b>
Käyttökate-%	$\frac{\text{käyttökate}}{\text{liikevaihto}}$
Liikevoitto-%	$\frac{\text{liiketoiminnan tulos}}{\text{liikevaihto}}$
Quick Ratio	$\frac{\text{rahoitusomaisuus}}{\text{lyhytaikaiset velat - saadut ennakot}}$
Current Ratio	$\frac{\text{rahoitusomaisuus} + \text{vaihto-omaisuus}}{\text{lyhytaikaiset velat}}$
Käyttöpääoma-%	$\frac{\text{vaihto-omaisuus} + \text{myyntisaamiset (lyhytaikaiset)} - \text{ostovelat (lyhytaikaiset)} - \text{saadut ennakot (lyhytaikaiset)}}{\text{liikevaihto}}$
Omavaraisuusaste	$\frac{\text{oma pääoma} + \text{varaukset}}{\text{taseen loppusumma} - \text{saadut ennakot}}$
Sijoitetun pääoman tuotto-%	$\frac{\text{tulos ennen satunnaisia eriä} + \text{vieraan pääoman kulut}}{((\text{tuoreimman taseen loppusumma} - \text{korottomat velat}) + (\text{edellisen taseen loppusumma} - \text{korottomat velat})) / 2}$
Kokonaispääoman tuotto-%	$\frac{\text{tulos ennen satunnaisia eriä} + \text{vieraan pääoman kulut}}{(\text{tuoreimman taseen loppusumma} + \text{edellisen taseen loppusumma}) / 2}$
Suhteellinen velkaantuneisuus-%	$\frac{\text{pitkä- ja lyhytaikaiset velat} + \text{pakolliset varaukset} - \text{saadut ennakot (lyhyt- ja pitkäaikaiset)}}{\text{liikevaihto}}$
<b>Altmanin tunnusluvut</b>	
X1	$\frac{\text{nettokäyttöpääoma}}{\text{koko pääoma}}$
X2	$\frac{\text{kertyneet voittovarot}}{\text{koko pääoma}}$
X3	$\frac{\text{tulos ennen korkoja ja veroja}}{\text{koko pääoma}}$
X4	$\frac{\text{oma pääoma}}{\text{vieras pääoma}}$
X5	$\frac{\text{myynti}}{\text{koko pääoma}}$

*Liikevoittoprosentin ja käyttökateprosentin* tunnusluvuista laskettiin keskiarvot analyysissä käytetyn tilinpäätöksen sekä sitä edeltävän tilinpäätöksen arvoista. Nämä kaksi keskiarvoihin perustuvaa tunnuslukua toimivat itsenäisinä tunnuslukuina alustavien potentiaalisten tunnuslukujen joukossa. Keskiarvojen laskemisella pyrittiin vähentämään kyseisiin tunnuslukuihin sisältyvän luontaisen yrityskohtaisen varianssin määrää. Mikäli yritys oli uusi ja se oli laatinut vasta yhden tilinpäätöksen, olivat näiden tunnus-

lukujen arvot samat kuin yhteen tilikauteen perustuvat *liikevoitto-* ja *käyttökateprosentit*.

Alustavissa kokeiluissa oli mukana myös Voitto+ -tietokannassa olevan *gearing-prosentin*<sup>38</sup> muunnelmia. *Gearing-prosentti* on ominaisuuksiltaan sellainen, että se ei sovellu tutkielmassa laadittuihin malleihin. Siinä sekä osoittaja että nimittäjä voivat saada positiivisia tai negatiivisia arvoja. Osoittajassa negatiivinen arvo on lähtökohtaisesti suotavaa yrityksen kannalta ja nimittäjässä puolestaan positiivinen arvo. Näistä syistä johtuen tunnusluvun (tai sen muunnelmien) arvojen vertailu yritysten välillä konkurssin ennustamisen yhteydessä ei ole mielekäästä, joten *gearing-prosenttia* ei hyödynnetty tutkielman lopullisissa analyyseissa.

Tietyille muuttujille tehtiin *logaritmimuunnoksia*,<sup>39</sup> jolla tarkoitetaan tässä tutkielmassa luonnollisen logaritmin ottamista tunnusluvun arvosta tai sen muunnelmasta. Luonnollisen logaritmin ottaminen ei onnistu negatiivisesta luvusta, joten esimerkiksi *liikevoittoprosentin* kohdalla, joka voi saada negatiivisen arvon, luonnollinen logaritmi otettiin arvosta  $(1 - \text{liikevoitto-\%})$ . Vastaavasti sellaisten tunnuslukujen kohdalla, jotka voivat saada yli 100 %:n arvoja, tunnusluvun arvo vähennettiin riittävän suuresta kokonaisluvusta, jotta kaikkien havaintoyksikköjen kyseisen tunnusluvun arvoista oli mahdollista tehdä logaritmimuunnos. Tällainen yli 100 %:n arvoja saanut tunnusluku oli esimerkiksi *kokonaispääoman tuottoprosentti*.

Logaritmimuunnoksia tehtiin tunnuslukujen tilastollisten ominaisuuksien muuntamiseksi. Esimerkiksi *current ration* ja *quick ration* arvojen jakaumat olivat vahvasti positiivisesti eli oikealle vinoja. Tämä johtui siitä, että niiden pienin mahdollinen arvo on nolla johtuen tunnuslukujen laskentakaavoista. Ottamalla tunnusluvuista luonnolliset logaritmit niiden huipukkuutta ja vinoutta saatiin huomattavasti pienennettyä, joten muunnoksen arvojen jakaumat noudattivat läheisemmin normaalijakaumaa kuin alkuperäiset tunnusluvut. Odotetusti juuri *current ration* ja *quick ration* logaritmimuunnokset suoriutuivat alkuperäisiä tunnuslukuja paremmin ennustamismallien laatimisvaiheessa. Muiden tunnuslukujen kohdalla muunnoksesta ei ollut havaittavaa hyötyä.

Lopullisten muuttujien valinta ennustamismalliin voidaan tehdä teoreettisin perustein, jolloin valinta perustuu konkurssiyrityksen käyttäytymistä kuvaavaan malliin, joka voi olla käsitteellinen tai matemaattinen.<sup>40</sup> Suurimmassa osassa konkurssitutkimuksia on käytetty Laitisen ja Laitisen (2004, 169) mukaan empiirisiin perustein valittuja muuttujia ennustamismallin laatimisessa. Näissä tutkimuksissa selittävät muuttujat valikoitu-

<sup>38</sup> Kaava:  $(\text{korollinen vieras pääoma} - \text{rahat ja rahoitusarvopaperit}) / \text{oma pääoma}$

<sup>39</sup> Logaritmimuunnoksia ovat käyttäneet muun muassa Altman (1977) ja Ohlson (1980).

<sup>40</sup> Teoreettista mallia ovat käyttäneet esimerkiksi Wilcox (1971), Prihti (1975), Santomero ja Vinso (1977) sekä Laitinen ja Laitinen (1998).

vat tutkimuksen empiirisen aineiston perusteella sen mukaan, kuinka hyvin ne toimivat analysoitavassa aineistossa. Empiirisiin perusteisiin laadittavien mallien yhteydessä suoritetaan yleensä ensin esivalinta, jossa rajataan mukaan harkittavien selittävien muuttujien joukko. Esivalinnan kriteerejä voivat olla esimerkiksi tunnuslukujen aikaisempi toimivuus tutkimuksissa, käytön yleisyys liike-elämässä sekä saatavuus. (Laitinen & Laitinen 2004, 169.)

Esivalittujen muuttujien karsintaan ja lopullisen mallin laatimiseen voidaan käyttää tilastotieteellistä lähestymistapaa. Logit-mallin ja lineaarisen erotteluanalyysin tapauksessa on mahdollista käyttää *askeltavaa menetelmää*, joka valitsee malliin tietyin perustein lisää muuttujia tai poistaa niitä siten, että mallin suorituskyky paranee asteittain. (Laitinen & Laitinen 2004, 172.)

Toinen mahdollinen lähestymistapa on hyödyntää tilastollista *faktorianalyysia*. Siinä selvitetään selittävien muuttujien niin sanottuja perusulottuvuuksia tai faktoreita.<sup>41</sup> Menetelmän tarkoituksena on karsia muuttujien määrää suuresta joukosta alustavia potentiaalisia tunnuslukuja, joiden mukaan ottamista harkitaan konkurssin ennustamismalliin. Faktorianalyysin avulla on mahdollista löytää aineiston pohjalta ulottuvuuksia, johon kuuluvat muuttujat korreloivat vahvasti keskenään, mutta eivät toisiin ulottuvuuksiin kuuluvien muuttujien kanssa.<sup>42</sup> Valitsemalla yksi tunnusluku kustakin faktorianalyysin avulla selvitetystä perusulottuvuudesta voidaan välttyä ennustamismallin *multikollinearisuuden* aiheuttamilta ongelmilta. (Laitinen & Laitinen 2004, 172–173.) Luvussa 4.2 on käsitelty tutkielmassa sovellettua tapaa selvittää selittävät muuttujat alustavien muuttujien joukosta.

---

<sup>41</sup> Konkurssin ennustamismallin laatimisen yhteydessä kyseessä on useimmiten eksploratiivinen faktorianalyysi (KvantiMOTV 2016), jossa faktorit selvitetään aineistolähtöisesti ilman, että tutkijalla on vahvoja käsityksiä etukäteen faktoreista tai niiden määrästä. Pinches ym. (1973) ja Pinches ym. (1975) käyttivät faktorianalyysia tunnuslukuanalyysissa ja sittemmin menetelmän lisääntyneen käytön myötä muun muassa Taffler (1982) on käyttänyt sitä yrityksen epäonnistumisen ennustamisessa.

<sup>42</sup> Tämä edellyttää suorakulmarotaatiomenetelmän käyttöä, joka tuottaa keskenään korreloimattomia faktoreita (KvantiMOTV 2016).

### 3.6 Estimointiryhmän ja testiryhmän laatiminen

Tutkielman koko aineisto, joka koostui 143 konkurssiyrityksen ja 143 toimivan yrityksen tilinpäätösaineistoista, jaettiin kahteen ryhmään: 144 yrityksen estimointiryhmään ja 142 yrityksen testiryhmään. Yhdeltä toimivalta yritykseltä valikoitui aineistoon kaksi eri tilikauden tilinpäätöstä vastinparimenettelyn seurauksena. Näitä käsiteltiin erillisinä tapauksina ja havaintoyksikköinä. Toinen näistä tilinpäätöksistä sisältyi estimointiryhmään ja toinen testiryhmään. Ennustamismallien laatiminen tehtiin estimointiryhmän datan perusteella ja ennustamismallien toimivuuden testaaminen toteutettiin soveltamalla laadittuja malleja testiryhmän yrityksiin. Tällä yleisesti konkurssin ennustamistutkimuksissa käytettävällä menetelmällä pyritään saamaan näyttöä sille, kuinka hyvin laadittu malli voisi käytännössä toimia, sillä testiryhmä on riippumaton estimointiryhmästä ja sen perusteella laaditusta ennustamismallista.

Ilman aineiston jakamista estimointi- ja testiryhmään on olemassa mahdollisuus, että malli toimii selvästi paremmin siinä aineistossa, jonka perusteella malli on laadittu kuin aineiston määrittämässä perusjoukossa. Perusjoukon muodostavat tutkielman tapauksessa tilintarkastusvelvollisuuden kokorajat ylittävät suomalaiset yritykset.

Estimointiryhmään kuuluneet 72 konkurssiyritystä valittiin kyseiseen ryhmään sillä perusteella, että niiden konkurssit rekisteröitiin kaupparekisteriin aikaisemmin kuin jäljelle jäävien 71 konkurssiyrityksen. Estimointiryhmän yritysten konkurssit sijoituivat aikavälille 21.6.2011 - 28.5.2013 ja testiryhmän vastaavasti aikavälille 30.5.2013 - 16.12.2014. Tietolähteenä konkurssin rekisteröintipäivämäärän selvittämisessä on käytetty Patentti- ja rekisterihallituksen ylläpitämää Virre-tietopalvelua.

Valittu ryhmien muodostamistapa perustuu siihen ajatukseen, että näin menettelmällä mallinnetaan mahdollisimman realistisesti todellista ennustamistilannetta, jossa tulevaisuudessa mahdollisesti tapahtuvia konkursseja ennustetaan menneisyudessa tapahtuneiden konkurssien ja niihin liittyvän tilinpäätösinformaation perusteella. Konkurssiyritysten vastinparirytykset valikoituivat luvussa 3.4 määritetyllä tavalla, joten toimivien yritysten jako estimointi- ja testiryhmiin määräytyi niiden vastinparirytysten konkurssin rekisteröintipäivän perusteella.

## 4 TUTKIMUSTULOKSET

### 4.1 Aineiston kuvailu sekä konkurssiyritysten ja toimivien yritysten vertailu

Konkurssin ennustamista käsittelevissä tutkimuksissa ja tutkielmissa käytetyn aineiston kuvailu on olennaista, jotta tulosten yleistettävyyttä voidaan arvioida. Tämän tutkielman aineiston yritysten liikevaihdon keskiarvo oli noin 1 231 000 €, alakvartiili 430 000 €, mediaani 728 000 € ja yläkvartiili 1 393 000 €. Liikevaihdon määrä on suhteutettu 12 kuukauden pituiseen tilikauteen yrityksillä, joiden tilikauden pituus oli tästä poikkeava. Taseen loppusumman vastaavat luvut olivat 540 000 €, 173 000 €, 292 000€ ja 691 000 €. Liikevaihdon määrä vaihteli noin 200 000 eurosta 19 384 000 euroon ja taseen loppusumma 100 000 eurosta 4 940 000 euroon. Taulukossa 3 on nähtävissä erikseen toimivien yritysten ja konkurssiyritysten kokoa mittaavat luvut.

Taulukko 3 Aineiston yritysten kokoluokat

Kokomuuttujat (1 000 €)	Keskiarvo	Alakvartiili	Mediaani	Yläkvartiili	Minimi	Maksimi
Kaikki yritykset (N 286)						
<i>Liikevaihto</i>	1231	430	728	1 393	200	19 384
<i>Taseen loppusumma</i>	540	173	292	691	100	4 940
Toimivat yritykset (N 183)						
<i>Liikevaihto</i>	1 214	437	715	1 401	200	16 930
<i>Taseen loppusumma</i>	534	176	294	681	100	4 940
Konkurssiyrietykset (N 183)						
<i>Liikevaihto</i>	1 248	426	746	1 390	212	19 384
<i>Taseen loppusumma</i>	547	163	290	722	100	4 917

Suuria eroja ei ollut odotettavissa konkurssiyritysten ja toimivien yritysten taseiden loppusummien välillä johtuen aiemmin kuvatusta vastinparimenettelystä, mutta myös liikevaihdon mitattuna näiden kahden yritysryhmän koot olivat lähellä toisiaan. Taulukosta on nähtävissä, että aineiston yritykset olivat enimmäkseen mikroyrityksiä ja pienyrityksiä sekä Suomen kirjanpitolain (KPL 1:4 a–b) että Euroopan Unionin komission (2003)<sup>43</sup> määritelmän mukaan.

Aineiston yritykset edustivat yhteensä 91:tä eri Tilastokeskuksen määrittämää toimialaa. Jopa 71 toimialalta oli mukana vain kaksi yritystä, eli konkurssiyritys ja sen vas-

<sup>43</sup> Komission suositus 2003/361/EY mikroyritysten sekä pienten ja keskisuurten yritysten määritelmästä.

tinpari samalta toimialalta. Selvästi suurin määrä yrityksiä, eli yhteensä 50, oli toimialalta 41200 (Asuin- ja muiden rakennusten rakentaminen). Toiseksi eniten yrityksiä oli toimialalta 49410 (Tieliikenteen tavarankuljetus), joita oli yhteensä 12.

GLMM-menetelmällä haluttiin hyödyntää toimialamuuttujaa, mutta 91 eri toimialaa oli liian suuri määrä kategorioita johtuen havaintoyksikköjen rajallisesta määrästä, joten analyyseja varten yritysten toimialaluokituksia tyhistettiin siten, että 91 toimialaa tiivistyi 15 toimialaksi. Taulukossa 4 on nähtävissä nämä 15 laajempaa toimialaa sekä niiden frekvenssit tutkielman aineistossa.

Taulukko 4 Aineiston yritysten toimialaluokitukset

TOL 2008	Toimiala	Frekvenssi	Osuus %
A	Maatalous, metsätalous ja kalatalous	4	1,4 %
B	Kaivostoiminta ja louhinta	2	0,7 %
C	Teollisuus	62	21,7 %
E	Vesihuolto, viemäri- ja jätevesihuolto, jätehuolto ja muu ympäristön puhtaanapito	2	0,7 %
F	Rakentaminen	76	26,6 %
G	Tukku- ja vähittäiskauppa; moottoriajoneuvojen ja moottoripyörien korjaus	62	21,7 %
H	Kuljetus ja varastointi	18	6,3 %
I	Majoitus- ja ravitsemistoiminta	8	2,8 %
J	Informaatio ja viestintä	2	0,7 %
L	Kiinteistöalan toiminta	2	0,7 %
M	Ammatillinen, tieteellinen ja tekninen toiminta	26	9,1 %
N	Hallinto- ja tukipalvelutoiminta	6	2,1 %
Q	Terveys- ja sosiaalipalvelut	4	1,4 %
R	Taiteet, viihde ja virkistys	4	1,4 %
S	Muu palvelutoiminta	8	2,8 %
	<b>Yhteensä</b>	286	100,0 %

Kolme toimialaluokituksen pääluokkaa erottui selvästi muita pääluokkia suuremmalla edustuksella. Näitä olivat: C (Teollisuus), F (Rakentaminen) sekä G (Tukku- ja vähittäiskauppa; moottoriajoneuvojen ja moottoripyörien korjaus). Yhdessä ne muodostivat 69,9 %:n osuuden kaikista aineiston yrityksistä. Estimointiryhmän ja testiryhmän yritysten toimialat määräytyivät satunnaisesti, sillä yrityksen ryhmän määritti vastinparin konkurssiyrityksen konkurssin ajankohta. Näin ollen toimialojen osuudet olivat erilaisia ryhmien välillä. Sen sijaan konkurssiyrityksiä ja toimivia yrityksiä oli kultakin toimialalta yhtä suuri määrä johtuen vastinparimenettelystä.

Tutkielmassa kerättyjen ja laadittujen tunnuslukujen arvoissa oli havaittavissa eroja konkurssiyritysten ja toimivien yritysten välillä. Tämä oli odotettavissa aikaisempien

tutkimusten perusteella lähtien jo Beaverin (1966) tutkimuksesta. Taulukkoon 5 on koottu tunnuslukujen keskiarvot ja standardipoikkeamat toimivien ja konkurssiyritysten osalta erikseen.

Taulukko 5 Estimointiryhmän tunnuslukuvertailu toimivien ja konkurssiyritysten välillä

Tunnuslukuvertailu (N = 144)	Keskiarvo			Standardi- poikkeama		Sig.	Sig. (2-tailed)
	Toimiva	Konkurssi	Erotus	Toimiva	Konkurssi		
<b>Tunnusluku</b>						Varianssit yhtä suuret	Keskiarvot yhtä suuret
Käyttökate-%	11,7	-6,5	18,2	11,7	21,0	0,085	< 0,001
Käyttökate-% keskiarvo kahdelta viimeiseltä tilikaudelta	10,4	-3,8	14,2	10,3	13,6	0,43	< 0,001
Liikevoitto-%	8,5	-8,8	17,3	10,5	23,4	0,037	< 0,001
Liikevoitto-% keskiarvo kahdel- ta viimeiseltä tilikaudelta	7,1	-7,0	14,1	9,0	14,4	0,088	< 0,001
Quick Ratio	1,8	0,6	1,2	1,95	0,54	< 0,001	< 0,001
Quick Ratio logaritimuunnos (ln)	0,18	-0,79	0,97	0,92	0,80	0,248	< 0,001
Current Ratio	2,47	0,94	1,53	2,41	0,72	< 0,001	< 0,001
Current Ratio logaritimuunnos (ln)	0,59	-0,30	0,89	0,80	0,72	0,509	< 0,001
Käyttöpääoma-%	14,6	7,6	7,0	17,5	25,2	0,274	0,054
Omavaraisuusaste	47,4	-39,7	87,1	27,6	68,0	< 0,001	< 0,001
Sijoitetun pääoman tuotto-%	31,6	-16,2	47,8	53,2	40,7	0,566	< 0,001
Kokonaispääoman tuotto-%	20,6	-10,5	31,1	25,7	21,7	0,342	< 0,001
Suhteellinen velkaantuneisuus-%	26,4	64,8	-38,4	20,9	55,8	< 0,001	< 0,001
<b>Altmanin tunnusluvut</b>							
X1 Nettokäyttöpääoma / koko pääoma	0,26	0,13	0,13	0,24	0,35	0,014	0,008
X2 Kertyneet voittovarot / koko pääoma	0,27	-0,29	0,56	0,4	0,53	0,02	< 0,001
X3 Tulos ennen korkoja ja vero- ja / koko pääoma	0,18	-0,19	0,37	0,22	0,35	0,023	< 0,001
X4 Oma pääoma / vieras pääoma	1,99	-0,14	2,13	3,45	0,35	< 0,001	< 0,001
X5 Myynti / koko pääoma	2,55	2,95	-0,40	1,38	1,87	0,075	0,143

Tunnuslukujen keskiarvojen ja varianssien mahdollisia eroavaisuuksia toimivien yritysten ja konkurssiyritysten välillä selvitettiin *kaksisuuntaisella kahden riippumattoman otoksen t-testillä*, joka testaa ryhmien keskiarvojen eroja, sekä *Levenen testillä*, joka testaa varianssien eroja ryhmien välillä. Testien tulokset ovat taulukon 5 kahdella viimeisellä sarakkeella, joissa ovat testien p-arvot. Yli 0,05 p-arvo tarkoittaa varianssien osalta sitä, että tunnuslukujen varianssit eivät ole yhtä suuria toimivilla yrityksillä ja konkurssiyrityksillä. Keskiarvojen eroja mittavaan testin p-arvoja tulkitaan siten, että

pienempi arvo tarkoittaa tilastollisesti merkitsevämpää eroavaisuutta ryhmien tunnuslukujen keskiarvojen välillä.<sup>44</sup>

Keskiarvojen erot ovat kaikkien tunnuslukujen osalta tilastollisesti erittäin merkitseviä lukuun ottamatta tunnuslukuja *käyttöpääomaprosentti*, *myynti/koko pääoma* ja *nettokäyttöpääoma/koko pääoma*. Viimeksi mainitun tunnusluvun kohdalla keskiarvojen erot ovat tilastollisesti merkitseviä ( $p < 0,01$ ), mutta eivät erittäin merkitseviä ( $p > 0,001$ ). Taulukon 5 erotus-sarakkeelta nähdään, että kaikkien tunnuslukujen paitsi Altmanin Z-mallin viidennen tunnusluvun *myynti/koko pääoma* tapauksessa keskiarvojen erotuksen etumerkki + tai - on intuitiivisen oletusarvon mukainen. Toisin sanoen yleisesti hyvään suorituskykyyn tai taloudelliseen asemaan yhdistettävien tunnuslukujen kohdalla keskiarvot ovat suurempia toimivien yritysten ryhmässä ja päinvastoin (ainoa päinvastoin toimiva tunnusluku on *suhteellinen velkaantuneisuusprosentti*). Altmanin (1968, 596) tutkimuksessa *myynti/koko pääoma* -tunnusluvun erotus oli toiseen suuntaan (1,90 toimivat yritykset ja 1,50 konkurssiyritykset), mutta ero ei ollut siinäkään tilastollisesti merkitsevää.

*Omavaraisuusasteen* keskiarvot erottuvat vertailussa hyvin selkeällä 87,1 prosenttiyksikön erolla toimivien yritysten ja konkurssiyritysten välillä. *Sijoitetun pääoman tuottoprosentin* keskiarvojen erotus ryhmien välillä on myös huomattava 47,8 prosenttiyksikön erotuksella. *Käyttökateprosentin* erotus on hieman *liikevoittoprosenttia* suurempi. Sekä *käyttökateprosentin* että *liikevoittoprosentin* kahden viimeisen tilikauden keskiarvojen erotukset ovat yhden tilikauden erotuksia pienempiä ja lisäksi niiden varianssit ovat pienempiä erityisesti konkurssiyritysten tapauksessa.

Tämän tutkielman luvussa 4.2.1 käsiteltävien tulosten kannalta on huomionarvoista, että myös lyhyen aikavälin maksuvalmiutta mittaavien tunnuslukujen *quick ratio* ja *current ratio* keskiarvot ovat tilastollisesti erittäin merkitsevästi suurempia toimivien yritysten ryhmässä kuin konkurssiyritysten ryhmässä. Näiden tunnuslukujen osalta taulukossa 5 on myös esitetty niiden logaritmimuunnosten keskiarvot. Tutkielman empiiriosassa osassa tehtiin logaritmimuunnoksia myös muille tunnusluville, mutta vain näiden kahden tunnusluvun kohdalla tällä toimenpiteellä oli huomattavissa oleva vaikutus muuttujien selityskykyyn laadituissa ennustamismalleissa. Varianssien samanlaisuutta vertailevasta sarakkeesta on nähtävissä, että muunnoksella oli ryhmien variansseja yhtenäistävä vaikutus. Varianssien yhteneväisyys on tilastotieteellisestä näkökulmasta olen-

---

<sup>44</sup> Yleisesti käytetyn p-arvon tilastollisen merkitsevyyden luokittelun mukaan  $< 0,05$  on tilastollisesti melkein merkitsevää;  $< 0,01$  on tilastollisesti merkitsevää; ja  $< 0,001$  on tilastollisesti erittäin merkitsevää (Holopainen & Pulkkinen 2013, 177).

naista ainakin lineaarisen erotteluanalyysin kohdalla, joka edellyttää yhtenäisiä varianssikovarianssimatriiseja ryhmien välillä (konkurssiyrietykset ja toimivat yritykset).<sup>45</sup>

Sekä aineistoon kuuluneet konkurssiyrietykset että toimivat yritykset rekisteröivät tilinpäätöksensä kaupparekisteriin keskimäärin 180 päivää eli noin kuusi kuukautta tilikauden päättymispäivän jälkeen. Sellaisia konkurssiyrietysten tilinpäätöksiä ei huomioitu, jotka oli julkaistu konkurssiin asettamisen jälkeen. Tilikauden päättymispäivän ja konkurssiin asettamispäivämäärän välillä oli keskimäärin 375 päivää eli hieman yli vuosi. Tilinpäätöksen *julkaisupäivämäärän* ja konkurssiin asettamisen välillä puolestaan oli keskimäärin 195 päivää eli hieman yli puoli vuotta. Näiden kahden aikavälin ero on huomattava, jonka johdosta on olennaista käyttää tilinpäätöksen julkaisupäivämäärää kriteerinä analyyseihin otettavan konkurssiyrietyksen tilinpäätöksen valinnassa, mikäli tavoitteena on simuloida esimerkiksi potentiaalisen vieraan pääoman rahoittajan päätöksentekotilannetta, kuten luvuissa 2.2.3 ja 3.4 on tullut ilmi.

## 4.2 Ennustamismallien laatiminen

Tutkielmassa käytetty lähestymistapa ennustamismallien laatimisessa oli empiirinen, sillä ennustamismallit laadittiin sillä perusteella, mikä malli pystyi erottelemaan konkurssiyrietykset ja toimivat yritykset parhaiten yritysten tilinpäätösinformaation perusteella. Lukuisia erilaisia yhdistelmiä taulukon 5 tunnusluvuista kokeiltiin parhaan erottelukyvyyn löytämiseksi estimointiryhmän datan perusteella. Kokeiluissa oli mukana myös muuttujien logaritimuunnoksia sekä *käyttökateprosentin* ja *liikevoittoprocentin* keskiarvotunnusluvut analyyseissa käytetyltä tilikaudelta sekä sitä edeltäneeltä tilikaudelta.

Selittävien muuttujien valinnassa ei käytetty faktorianalyysia, mutta kyseisen menetelmän sisältämä periaate ja tavoite vaikuttivat tässäkin tutkielmassa mallin laatimiseen. Erilaisten yhdistelmien kokeileminen ei ollut kohtuuttoman vaivalloista, kun aineisto oli saatettu analysoitavaan muotoon, joten yhdistelmiä oli mahdollista kokeilla määrällisesti useita. Lisäksi käyttäen intuitiota ja aikaisempien tutkimusten tuloksia hyväksi oli pääteltävissä, että esimerkiksi *sijoitetun pääoman tuottoprocentti* ja *kokonaispääoman tuottoprocentti* sekä *quick ratio* ja *current ratio* ovat keskenään vahvasti korreloituneita tunnuslukupareja, joten ei ollut mielekäästä tehdä kokeiluja, joissa kumpikin muuttuja näistä pareista olisi ollut mukana. Tilanne olisi toinen, mikäli harkittavien tunnusluku-

---

<sup>45</sup> Richardsonin ja Davidsonin (1983) tutkimuksen mukaan Altmanin Z-mallin selittävien tunnuslukujen varianssikovarianssimatriisit eivät olleet yhtäläisiä konkurssiyrietysten ja toimivien yritysten välillä.

jen määrä olisi huomattavan suuri.<sup>46</sup> Tutkielmassa käytettävissä olevan aineiston ja resurssien puitteissa alustavien tunnuslukujen lukumäärä muodostui riittävän hyvin hallittavaksi siten, että varsinaiseen faktorianalyysiin ei koettu olevan tarvetta, sillä sen toimintaperiaate oli sisällytettävissä analyyseihin edellä kuvatulla tavalla. Lisäksi tutkielman päätavoitteet liittyivät uutuusarvoa omaavan tilastollisen menetelmän uudelleen soveltamiseen sekä valituin rajauksin muodostetun yrityspopulaation analysointiin laajamittaisen selittävien muuttujien roolia korostavan lähestymistavan sijaan.

Ennustamismallien laatimisprosessin alusta alkaen kävi selväksi, että *omavaraisuusaste* oli yksittäisenä tunnuslukuna tärkein selittävä muuttuja. Tämä oli alustavasti pääteltävissä jo taulukon 5 tunnuslukujen keskiarvovertailuista toimivien ja konkurssiyritysten välillä. Käyttämällä Beaverin (1966) soveltamaa menetelmää, jossa tunnusluvulle määritetään raja-arvo, jolla kokonaisluokitteluvirhe saadaan mahdollisimman pieneksi, saavutettiin tutkielman aineistossa 85,4 %:n ennustamistarkkuus estimointiryhmässä ja 82,4 %:n tarkkuus testiryhmässä *omavaraisuusasteen* raja-arvolla 23 %. Virhetyyppien jakauma oli kuitenkin varsin epätasainen, sillä virhetyypin I osuus oli estimointiryhmässä 6,9 % ja testiryhmässä 4,2 % virhetyypin II osuuksien ollessa vastaavassa järjestyksessä 22,2 % ja 31 %. Näin ollen toimivia yrityksiä tulisi luokiteltua konkurssiyrityksiksi jopa 31 %:ssa tapauksista testiryhmän tulosten perusteella, joka nostaisi todellisen kokonaisluokitteluvirheen (ei tässä yhteydessä virhetyyppien I ja II keskiarvon) osuutta todellisessa yrityspopulaatiossa, sillä selvästi suurempi osa yrityksistä ei ajaudu konkurssiin yksittäisen satunnaisesti valitun tilikauden aikana.

*Omavaraisuusasteen* lisäksi malliin lisättiin kannattavuutta ja lyhyen aikavälin maksuvalmiutta mittaavia tunnuslukuja vaihtelevin yhdistelmin. Selittävien muuttujien lukumäärää ei ollut ennalta päätetty. Sen sijaan se määräytyi sillä perusteella, kuinka monta muuttujaa parhaiten ennustavassa mallissa lopulta oli. Eri tunnuslukuyhdistelmiä testattiin ensin kaikkien kolmen menetelmän (GLMM, logit-malli ja lineaarinen erotteluanalyysi) osalta estimointiryhmässä. Mikäli ennustamistarkkuus potentiaalisessa mallissa oli vähintään kohtuullisen korkea verrattuna muihin yhdistelmäkokeiluihin, testattiin estimointiryhmän avulla laadittua mallia myös testiryhmässä. GLMM-menetelmän kohdalla luvussa 4.1 esiintyvät 15 toimialaa määritettiin ryhmitteleväksi tekijäksi luvussa 3.3 esitetyn periaatteen mukaisesti. Kaikissa GLMM-kokeiluissa logit-mallin ja lineaarisen erotteluanalyysin tapaan tavoitteena oli selvittää kiinteiden vaikutusten osuus eli kertoimet eri tunnusluvuille eri tunnuslukuyhdistelmissä. Satunnaisvaikutuksia voidaan huomioda GLMM-menetelmällä varsin vapaasti, kuten luvussa 3.3 jo mainittiin. Satunnaisvaikutuksia huomioitiin kokeiluissa siten, että vain vakiotermi sai sisältää satun-

---

<sup>46</sup> Esimerkiksi El Hennawyn ja Morrisin (1983) tutkimuksessa alustavia tunnuslukuja oli jopa 96.

naisia vaikutuksia sekä siten, että vakiotermin lisäksi kaikki kiinteän vaikutuksen muuttajat saivat lisäksi sisältää satunnaisia vaikutuksia (toimialaluokituksen perusteella).

#### 4.2.1 *Laaditut mallit ja ennustamistarkkuudet*

Tutkielmassa laadittu parhaiten suoriutuva ennustamismalli voidaan määrittää kahdesta eri näkökulmasta. Kuten edellisessä luvussa mainittiin, eri tunnuslukuyhdistelmiä laadittiin useita ja niiden ennustamistarkkuutta testattiin testiryhmässä, mikäli malli suoriutui ainakin kohtuullisen hyvin estimointiryhmässä verrattuna aikaisempiin kokeiluihin. Näin ollen tutkielman laatijan tietoon tuli mallin laatimisprosessin yhteydessä se, miten hyvin muutamia kymmeniä parhaiten toimivat eri yhdistelmät toimivat testiryhmässä. Tämä mahdollisti eri mallien ennustamistarkkuuden arvioimisen kokonaisuutena sekä estimointiryhmässä että testiryhmässä.

Yhdestä näkökulmasta voidaan katsoa, että testiasetelmassa ei voida huomioida testiryhmän tuloksia ennen kuin on päätetty "lopullinen malli tai lopulliset mallit", sillä tällöin testiryhmä vaikuttaa mallin valintaan. Toisaalta voidaan ajatella, että tutkielman kokonaisuus testiryhmän sisältämä informaatio mukaan lukien määrittää käytetyn aineiston perusteella mahdollisimman optimaalisen mallin, jota voidaan käyttää tutkielman aineiston ulkopuolella odotusarvoisesti toimivimmin vaihtoehtoisista ennustamismalleista.

Tästä jälkimmäisestä lähestymistavasta käytetään tutkielmassa nimitystä lähestymistapa I. Siinä kriteerinä oli mahdollisimman suuri ennustamistarkkuus estimointiryhmässä ja testiryhmässä kokonaisuutena, mutta kuitenkin siten, että kahden tai useamman mallin ollessa tasatilanteessa estimointiryhmän ja testiryhmän keskimääräisessä ennustamistarkkuudessa valittiin se malli, joka ennusti paremmin testiryhmässä.<sup>47</sup> Lähestymistavassa II malli on valittu sillä perusteella, minkä mallin ennustamistarkkuus estimointiryhmässä on ollut korkein. Testiryhmän sisältämä informaatio ei ole siis vaikuttanut edes epäsuorasti tämän mallin laadintaan.<sup>48</sup>

Taulukosta 6 on nähtävissä kummallakin lähestymistavalla eri menetelmien vakio-termit, selittävien muuttujien kertoimet, virhetyypit estimointiryhmässä ja testiryhmässä sekä kokonaisennustamistarkkuudet estimointi- ja testiryhmässä. Linearisesta erottelu-

---

<sup>47</sup> Ilman tätä tasatilannemenettelyä lähestymistavan I parhaiten suoriutuvia malleja olisi ollut kaksi, joista toinen olisi ollut sama kuin lähestymistavan II parhaiten suoriutuva malli.

<sup>48</sup> Tutkielman laatija ei ole havainnut aiemmassa kirjallisuudessa näiden vaihtoehtoisten lähestymistapojen esille tuomista. Kumpikin lähestymistapa on perusteltavissa, joten tutkielmassa päätettiin esitellä kummankin tavan tulokset.

analyysistä käytetään taulukossa lyhennettä LDA (linear discriminant analysis). Kaikilla vaihtoehtoisilla ennustamismenetelmillä on käytetty samoja selittäviä muuttujia lähestymistapojen sisällä, jotka määräytyivät parhaiten ennustavan mallin perusteella riippumatta siitä, millä menetelmällä tämä parhaiten ennustava malli oli laadittu.<sup>49</sup> Kummallakin lähestymistavalla tämä parhaiten suoriutuva menetelmä oli GLMM.

Taulukko 6 Laadittujen ennustamismallien vakiotermit, kertoimet ja tarkkuudet estimointi- ja testiryhmissä

Ennustamismallien vertailu	Vakiotermit ja tunnuslukujen kertoimet					Virhetyypit estimointiryhmässä		Virhetyypit testiryhmässä		Ennustamistarkkuus	
	Vakio-termi	X1	X2	X3	X4	I	II	I	II	Estimointiryhmä	Testiryhmä
<b>Lähestymistapa I</b>											
LDA	-0,190	1,206	0,687	1,325	0,323	18,1 %	9,7 %	16,9 %	15,5 %	86,1 %	83,8 %
Logit-malli	-1,378	6,312	0,393	7,676	-0,155	12,5 %	18,1 %	11,3 %	16,9 %	84,7 %	85,9 %
GLMM	-1,652	7,983	2,466	5,696	-0,342	9,7 %	8,3 %	9,9 %	15,5 %	91,0 %	87,3 %
<b>Lähestymistapa II</b>											
LDA	-0,195	1,288	2,154	-0,646	0,254	15,3 %	9,7 %	15,5 %	21,1 %	87,5 %	81,7 %
Logit-malli	-1,163	6,451	1,004	9,590	-0,435	12,5 %	16,7 %	11,3 %	19,7 %	85,4 %	84,5 %
GLMM	-1,572	7,559	9,112	1,528	-0,664	9,7 %	5,6 %	11,3 %	16,9 %	92,4 %	85,9 %
<p><b>Tunnusluvut lähestymistapa I</b>  X1 = Omavaraisuusaste  X2 = Sijoitetun pääoman tuotto-%  X3 = Käyttökate-% keskiarvo kahdelta viimeiseltä tilikaudelta  X4 = Current ratio logaritimuunnos (ln)</p> <p><b>Tunnusluvut lähestymistapa II</b>  X1 = Omavaraisuusaste  X2 = Kokonaispääoman tuotto-%  X3 = Liikevoitto-% keskiarvo kahdelta viimeiseltä tilikaudelta  X4 = Current ratio logaritimuunnos (ln)</p>											

<sup>49</sup> Tunnuslukuyhdistelmä *omavaraisuusaste, quick ratio ja tulos ennen korkoja ja veroja / koko pääoma* oli mielenkiintoinen sikäli, että tunnusluvut vastaavat läheisesti Laitisen ja Kankaanpään (1999) käyttämiä selittäviä muuttujia aineistossa, joka oli samankaltainen tutkielman aineiston kanssa (ks. luku 2.3). Laitinen ja Kankaanpää saavuttivat 89,5 % ennustamistarkkuuden logit-mallilla käyttäen *Lachenbruchin validointimenetelmää* mallin ennustamiskyvyn arvioimiseen. Tarkkuudet mainitulla tunnuslukuyhdistelmällä tutkielman aineistossa olivat järjestyksessä estimointiryhmä ja testiryhmä: lineaarinen erotteluanalyysi 86,1 % ja 83,8 %; logit-malli 85,4 % ja 83,8 % sekä GLMM-menetelmä 86,8 % ja 85,2 %.

Taulukon 6 vakiotermeistä ja tunnuslukujen kertoimista on johdettavissa logit-mallin ja GLMM-menetelmän antama konkurssin todennäköisyys mille hyvänsä yritykselle, jonka mallissa käytetyt tunnusluvut ovat tiedossa. Esimerkiksi lähestymistapaa I soveltaen GLMM-menetelmällä laaditun mallin soveltamiskelpoinen muoto voidaan esittää seuraavalla tavalla:

$$(kaava\ 6) \quad P_j = \frac{1}{1 + e^{-(-1,652 + 7,983X_1 + 2,466X_2 + 5,696X_3 - 0,342X_4)}}$$

missä

$P_j$  = yrityksen j konkurssin todennäköisyys

$e$  = Neperin luku

$X_1$ – $X_4$  = selittävät (yrityskohtaiset) tunnusluvut (ks. taulukko 6)

Edellä esitetty malli ennusti konkurssia parhaiten kaikista kokeilluista malleista tutkielman testiryhmässä. Sen kokonaisennustamistarkkuus oli 87,3 % virhetyypin I ollessa 9,9 % ja virhetyypin II 15,5 %. Estimointiryhmässä ennustamistarkkuus oli 91 % ja virhetyypit I ja II samassa järjestyksessä 8,3 % ja 9,9 %. Taulukossa 7 on esitetty ennustamismallin luokittelutarkkuus testiryhmässä ristiintaulukoinnin avulla.

Taulukko 7 Testiryhmässä parhaiten suoriutuneen ennustamismallin tarkkuuden ristiintaulukointi

<b>Konaisluokittelutarkkuus = 87,3%</b>			
Havaittu		Ennustettu	
		Toimiva	Konkurssi
Toimiva	Lukumäärä	60	11
	% havaituista	84,5%	15,5%
Konkurssi	Lukumäärä	7	64
	% havaitusta	9,9%	90,1%

Malli ennusti konkurssiyritysten konkurssit 90,1 %:n tarkkuudella ja toimivien yritysten toiminnan jatkumisen 84,5 %:n tarkkuudella. Linearisella erotteluanalyysilla laadittu malli erotteli yritykset 86,1 %:n tarkkuudella estimointiryhmässä ja 83,8 %:n tarkkuudella testiryhmässä, kun selittävien muuttujien määrittämiseen käytettiin lähestymistapaa I. Vastaavat tarkkuudet logistisella regressiolla laaditulla mallilla olivat 84,7 % ja 85,9 %. Luvussa 4.2.2 käsitellään sitä, miten luokitteluvirheiden jakaumaan voi vaikuttaa.

Estimointiryhmän ennustamistarkkuuksien yhteydessä on huomionarvoista se, että GLMM-menetelmällä tarkkuuteen tuo kontribuutiota menetelmän komponentti  $Z_w$  (ks.

luku 3.3), joka ei ole siirrettävissä testattavaksi testiryhmään. Se on aineistokohtainen ja vaikuttaa tilasto-ohjelmiston laskelmissa, jotka eivät ole näkyvissä käytetyn tilasto-ohjelmiston antamissa tulosteissa. Ajatuksena on, että tämä komponentti  $Z_w$  mahdollistaa kiinteiden vaikutusten (tunnuslukujen kertoimien) ikään kuin satunnaisista vaikutuksista suodatetun selvittämisen tavalla, joka luvussa 3.3 käsitellyistä periaatteista kävi ilmi. Näin ollen GLMM-menetelmällä voisi lähtökohtaisesti katsoa olevan aiheeton etu mallin laadinnassa selittäviä muuttujia valittaessa, koska osa estimointiryhmän tarkkuuteen kontribuutiota tuovasta mallintamisesta ei ole siirrettävissä testiryhmään tai muuhun estimointiryhmän ulkopuoliseen aineistoon. Lukuisat eri tunnuslukuyhdistelmät kuitenkin osoittivat, etteivät logit-malli ja lineaarinen erotteluanalyysi saavuttaneet millään kokeillulla yhdistelmällä GLMM-menetelmän korkeinta erottelukykyä parempaa tarkkuutta.<sup>50</sup> Suurimmassa osassa kokeiluja GLMM-menetelmän tarkkuus oli menetelmistä korkein sekä estimointi- että testiryhmässä.

Ennustamistarkkuuksien eroja voidaan arvioida tilastollisesti käyttämällä kaksisuuntaista suhteellisten osuuksien vertailutestiä. Siinä käytetään seuraavaa kaavaa  $z$ -arvon selvittämiseen:

$$(kaava 7) \quad z = \frac{p_1 - p_2}{\sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

missä

$p_1, p_2$  = ennustamistarkkuudet vertailtavilla menetelmillä

$n_1, n_2$  = yritysten lukumäärät vastaavissa ennustamistarkkuuksissa

$p$  = ennustamistarkkuuden painotettu keskiarvo huomioiden kumminkin tarkkuudet sekä yritysten lukumäärät (yritysten lukumäärät ovat tässä tapauksessa samat eli 142 testiryhmässä jokaisella verrattavalla menetelmällä)

Merkitsevyytasolla  $p = 0,05$   $z$ -arvon pitäisi olla vähintään 1,96, jotta ennustamistarkkuuksissa kahden menetelmän välillä voitaisiin todeta olevan eroa käyttäen tätä tilastollista testiä. Edes parhaiten ennustavan GLMM-menetelmän (87,3 %) ja heikoiten ennustavan lineaarisen erotteluanalyysin (83,8 %) tarkkuuksilla testiryhmässä ei ollut tilastollisesti merkitsevää eroa ( $z$ -arvo = 0,84).

---

<sup>50</sup> Logistisen regression ja lineaarisen erotteluanalyysin kohdalla mallin laadinnassa kokeiltiin myös askelettavaa menetelmää muuttujien valinnassa, mutta sillä ei saavutettu parempaa erottelukykyä kuin manuaalisella muuttujien valinnalla, joka oli lukuisten yhdistelmien kokeilujen tulos.

Eron voidaan katsoa kuitenkin olevan merkittävä, sillä mikäli ennustamistarkkuudet vastaavat todellisuutta, on 87,3 % selvästi parempi kuin 83,8 % konteksti huomioiden. Ensinnäkin lähtökohtana voidaan pitää 50 %:n ennustamistarkkuutta, sillä se on saavutettavissa puhtaasti arvaamalla yrityksen tuleva konkurssistatus. Toiseksi todella korkeita ennustamistarkkuuksia ei ole edes realistista tai intuitiivisesti perusteltua odottaa huomioiden tutkielman aineiston määrittämä perusjoukko, eli enimmäkseen pienet tai keskisuuret yritykset, sekä ennustamisessa käytetty informaatiolähde eli tilinpäätökset. Esimerkiksi uusia yrityksen toiminnan jatkuvuuden varmistavia pääomasijoituksia tilinpäätöksen perusteella heikosti suoriutuviin yrityksiin ei ole mitenkään mahdollista ennustaa tilinpäätöksen perusteella, joten täydellinen tai lähes täydellinen ennustamistarkkuus ei ole edes mahdollinen. Mikäli tutkielman aineisto olisi ollut noin 5,4 kertaa suurempi ja edellä mainitut ennustamistarkkuudet GLMM-menetelmällä ja lineaarisella erotteluanalyyysilla olisivat olleet samat, olisi ero ollut tilastollisesti merkitsevä ( $p = 0,05$ ) tasolla. Merkitsevyystasoon ( $p = 0,1$ ) olisi vastaavasti riittänyt noin 3,9 kertaa suurempi aineisto.

Logistisella regressiolla ja GLMM-menetelmällä tilasto-ohjelmiston tulosteesta on nähtävissä kunkin selittävän muuttujan tilastollinen merkitsevyys mallissa. Ainoastaan *omavaraisuusaste* oli tilastollisesti merkitsevä näissä malleissa sekä lähestymistavalla I että II ( $p < 0,001$  kaikissa neljässä näitä menetelmiä käyttävissä malleissa, jotka ovat taulukossa 6). Muista selittävästä muuttujista lähimpänä tilastollista merkitsevyyttä oli *kokonaispääoman tuotto prosentti* GLMM-menetelmällä lähestymistavalla II ( $p = 0,085$ ). GLMM-menetelmällä toimialakohtaisia satunnaisvaikutuksia oli lähestymistavalla I *sijoitetun pääoman tuotto prosentilla*, *omavaraisuusasteella* ja *current ration* logaritmuunnoksella. Lähestymistavalla II tilanne oli sama sillä erotuksella, että *sijoitetun pääoman tuotto prosentilla* sijaan satunnaisvaikutuksia oli *kokonaispääoman tuotto prosentilla*. Näiden lueteltujen muuttujien satunnaisvaikutukset siis vaikuttivat menetelmällä laadittujen mallien kertoimiin, mutta ne eivät olleet tilastollisesti merkitseviä.

Tutkielmassa käytettiin menetelmää, jossa mahdollisimman korkea ennustamistarkkuus oli mallin valitsemisen kriteerinä, joten tilastollista merkitsevyyttä ei mallin laadinnassa otettu rajoittavana kriteerinä huomioon. Sen sijaan laaditun mallin ensisijainen toimivuuden testaus tehtiin kokeilemalla sitä testiryhmässä, kuten on tyyppillistä konkurssin ennustamistutkimuksissa. Suurempi aineisto olisi luonnollisesti mahdollistanut helpommin myös muuttujakohtaiset tilastolliset merkitsevyydet (mukaan lukien satunnaisvaikutukset).

Tarkastelemalla parhaiten ennustavia GLMM-menetelmällä laadittuja malleja voidaan tehdä tiettyjä huomioita niiden rakenteista (ks. taulukko 6). *Omavaraisuusasteen* kertoimet ovat hyvin lähellä toisiaan kummallakin lähestymistavalla. Sen sijaan muuttujien X2 ja X3 kohdalla painotukset ovat päinvastaisia eri lähestymistavoilla. X2 ja X3 ovat eri muuttujia eri lähestymistavoissa, mutta ne ovat tilinpäätösanalyttisesti koh-

tuullisen lähellä toisiaan. Kummallakin lähestymistavalla laaditussa mallissa X2 ja X3 mittaavat kannattavuutta, mutta eri tavalla. Muuttujat X2 mittaavat yrityksen kykyä hyödyntää pääomaa kannattavasti ja muuttujat X3 operatiivista kannattavuutta huomioiden myös edeltävän tilikauden tunnusluvun satunnaisten vaihteluiden tasaamiseksi. Oli odotettavissa, että X2-muuttujan kertoimen kasvaessa X3-muuttujan kerroin pienenee ja päinvastoin, sillä kumpikin kuitenkin mittaa kannattavuutta huolimatta edellä mainituista eroista.

Muuttujien kertoimien etumerkkejä tarkastellessa huomataan varsin mielenkiintoinen asia. Tunnuslukujen X1–X3 osalta etumerkki on odotusten mukainen lukuun ottamatta lineaarisen erotteluanalyysin X3-kertoimen etumerkkiä lähestymistavalla II. Positiivinen etumerkki tarkoittaa sitä, että konkurssin todennäköisyys pienenee tunnusluvun kasvaessa. X1–X3 tunnuslukujen suurempaa arvoa pidetään lähtökohtaisesti parempana kuin pienempää arvoa ajatellen yrityksen taloudellista suorituskykyä. *Current ration* logaritmuunnoksen kohdalla tilanne on kuitenkin päinvastainen sekä logit-mallissa että GLMM-menetelmällä.<sup>51</sup> Tunnusluvun etumerkki on ennustamismalleissa negatiivinen, vaikka korkea arvoa pidetään lähtökohtaisesti lyhyen aikavälin hyvänä maksuvalmiutena.<sup>52</sup> Myös tutkittavassa aineistossa *current ratio* oli keskimäärin suurempi toimivilla yrityksillä kuin konkurssiyrityksillä (ks. taulukko 5).

Käyttämällä kriittistä raja-arvoa *current ration* logaritmuunnokselle yritykset saadaan noin 75 %:n tarkkuudella eroteltua toimiviksi yrityksiksi ja konkurssiyrityksiksi estimointiryhmässä. Tässä pelkästään kyseistä tunnuslukua käyttävässä luokittelussa suurempi arvo tarkoittaa pienempää todennäköisyyttä ajautua konkurssiin. Negatiivinen etumerkki taulukon 6 malleissa on siis seurausta myös muiden tunnuslukujen huomiomisesta ja niiden yhteisvaikutuksesta.

Mallien laatimisen yhteydessä (edellä mainituista seikoista johtuen) heräsi ajatus, että johtuuko negatiivinen etumerkki vain mahdollisesti tässä mielessä poikkeuksellisesta estimointiryhmästä ja mallin sopeutumisesta siihen, jolloin se ei olisi yleistettävissä estimointiryhmän ulkopuolelle. X4-tunnusluvun lisääminen malliin (negatiivisella etumerkillä) kuitenkin paransi ennustamistarkkuutta myös testiryhmässä.

---

<sup>51</sup> Myös *quick ratio* sai usein negatiivisen etumerkin eri malleissa, mutta sen vaikutus oli pääsääntöisesti pienempi kuin *current ration*.

<sup>52</sup> Logaritmuunnos ei lähtökohtaisesti vaikuta muuttujan kertoimen etumerkkiin. Myös logaritmuunnoksessa pienempi arvo tarkoittaa pienempää alkuperäisen tunnusluvun arvoa, vaikka itse tunnusluku voikin muunnoksen jälkeen saada myös negatiivisia arvoja toisin kuin alkuperäisessä muodossaan. Käyttämällä *current ratiota* alkuperäisessä muodossa sen etumerkki oli myös tällöin negatiivinen mallissa, mutta mallin erottelukyky oli kokonaisuutena heikompi.

Mahdollinen selitys ilmiölle liittyy *current ratioon* tunnuslukuna sekä logistisen regression matemaattiseen rakenteeseen, joka pätee myös GLMM-menetelmässä. *Current ration* arvo kasvaa, kun yrityksen vaihto-omaisuuden ja myyntisaamisten kirjanpidollinen arvo kasvaa. Näihin omaisuuseriin liittyy vahvasti epäkuranttiuden riski, joka mahdollisesti kasvaa entisestään konkurssin uhan lähestyessä. Toiminnan jatkuvuuden suhteen epävarmalla pohjalla oleva yritys saattaa tietoisesti pitää kirjanpidossaan myyntisaamisia, jotka todellisuudessa ovat tulevia luottotappioita. Samoin vaihto-omaisuus saatetaan arvostaa liian suureksi kirjanpidossa. Lisäksi vaihto-omaisuutta saattaa kertyä, mikäli yrityksen tuotteille ei ole enää kysyntää.

Toinen eri näkökulmasta lähtevä mahdollinen tunnusluvun luonteeseen liittyvä selitys negatiiviselle etumerkille saattaa olla se, että suurempi *current ration* arvo – jota kasvattaa myös rahoitusomaisuus – saattaa rohkaista yrityksen velkojia hakemaan velallista yritystä konkurssiin. Tällöin heillä on korkeampi mahdollisuus saada saamisiaan konkurssiyritykseltä, mikäli velallisella on likvidiä varallisuutta.

Tällaisten konkurssin ennustamisen kontekstissa esiintyvien hienovaraisten lineaarisuudesta poikkeavien tunnusluvun arvojen kehityssuuntien mallintaminen on ymmärrettävästi haastavaa ennustamismenetelmälle.<sup>53</sup> Menetelmän laatimaan malliin sisällytetään sen matemaattisessa laatimisvaiheessa kaikkien yritysten tietyn tunnusluvun arvot (*current ration* logaritimuunnos tässä tapauksessa). Näihin yrityksiin lukeutuu laaditun mallin perusteella selvästi toimivia yrityksiä, selvästi konkurssiin ajautuvia yrityksiä sekä siinä rajamailla olevia yrityksiä. Rajamailla olevat yritykset ovat kaikkein kiinnostavimpia siinä mielessä, että niiden oikein luokittelussa vaihtoehtoisten menetelmien tarkkuuksien erot yleensä syntyvät ja ne ovat myös käytännön päätöksentekotilanteessa haastavimpia.

Edellä esitetyt mahdolliset tunnusluvun ominaisuuksiin liittyvät tekijät sen etumerkin negatiivisuudessa kytkeytyvät juuri näihin rajamailla oleviin yrityksiin. Jos konkurssi on vääjäämätön, ei myyntisaamisten tai vaihto-omaisuuden yliarvostamisella kirjanpidossa ole käytännössä merkitystä konkurssiin ajautumisen kannalta. Lisäksi hyvin suurella todennäköisyydellä konkurssiin ajautuvan yrityksen kohdalla velkojien ei ole samalla tavalla tarpeen päättää, hakevatko he yritystä konkurssiin vai eivät.<sup>54</sup> Näillä hyvin suurella todennäköisyydellä konkurssiin ajautuvilla yrityksillä on myös usein alhainen *current ratio*, jolloin sen vaikutus ei ole ennustamismallissa olennainen, sillä esimerkik-

---

<sup>53</sup> Mahdolliset selitykset ilmiölle ovat vain tutkielman laatijan arvioita. Selitysten oikeellisuuden arviointia voisi tutkia esimerkiksi kvalitatiivisella lähestymistavalla liittyen *current ration* arvoon vaikuttaviin tekijöihin konkurssin uhatessa ja sen toteutuessa.

<sup>54</sup> Tietysti myös tällaisessa tilanteessa velkojien aktiivisuus on suotavaa, mikäli yritys jatkaa omaisuuden hävittämistä tahallisesti tai tahattomasti.

si vahvasti negatiivinen oma pääoma ja kannattavuus jo luokittelevat yrityksen konkurssiyritykseksi. Vastaavasti hyvin vahvasti mallin mukaan pärjäävät yritykset luokitellaan jo korkean omavaraisuusasteen ja kannattavuuden perusteella niin hyvin toimiviksi yrityksiksi, että mahdollisesti suuren tai suurehkon *current ration* negatiivinen huomioiminen ei enää käännä yrityksen ennustetta konkurssin puolelle.

Logistisen regression matemaattisessa rakenteessa juuri rajamailla olevat yritykset vaikuttavat enemmän valmiin mallin lopulliseen muotoon kuin selvästi konkurssiyritykseksi tai toimivaksi yritykseksi luokiteltavat yritykset.<sup>55</sup> Näin ollen menetelmällä on mahdollista mallintaa tunnusluvun erilaista suhdetta konkurssiyrityksiin rajamailla olevissa yrityksissä ja muissa yrityksissä. Tunnusluvun arvojen kehityssuuntien lineaarisuuden puute konkurssin ennustamisen kontekstissa tosin helposti heikentää sen tilastollista merkitsevyyttä laaditussa mallissa. Muun muassa tästä johtuen tilastollista merkitsevyyttä ei ole syytä tulkita ehdottoman tiukasti ja samalla tavalla kaikissa tilanteissa, vaikka sen puuttuessa tuloksiin liittyy luonnollisesti epävarmuutta.

#### 4.2.2 *ROC-analyysi*

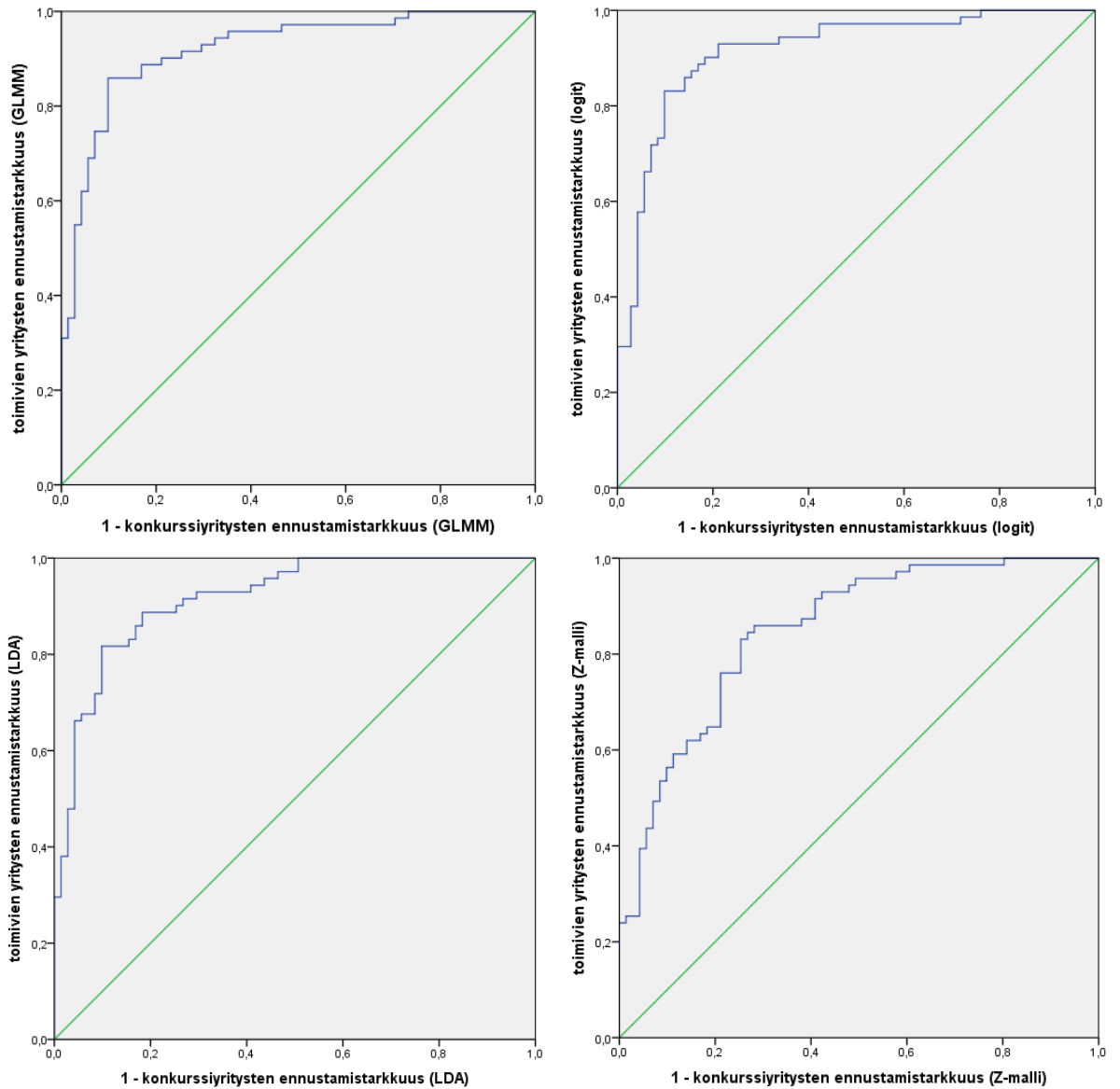
Ennustamismallien suorituskykyä voidaan arvioida pelkän kokonaisluokittelutarkkuuden lisäksi ROC-analyysin (receiver operating characteristic) tuottamien ROC-käyrien avulla, jotka muistuttavat toimintaperiaatteeltaan läheisesti CAP-käyriä (cumulative accuracy profile).<sup>56</sup> ROC-analyysillä voidaan arvioida sitä, kuinka hyvin ennustamismalli toimii silloin, kun kriittisenä raja-arvona käytetään jotain muuta kuin oletusarvoista raja-arvoa, joka esimerkiksi lineaarisessa erotteluanalyysissä on tyypillisesti 0 ja logit-mallissa 0,5.

Kuviossa 4 on esitetty testiryhmän ennustamistarkkuuksiin perustuvat ROC-käyrät lähestymistavalla I laadittujen ennustamismallien sekä Altmanin Z-mallin osalta käyttäen sen alkuperäisiä kertoimia.

---

<sup>55</sup> Lineaarisisessa erotteluanalyysin matemaattisessa rakenteessa ei ole vastaavaa ominaisuutta. Sillä laadituissa malleissa muuttujan  $X_4$  etumerkit olivat positiivisia, mutta niiden ennustamistarkkuudet olivat heikompia kuin logistisella regressiolla ja GLMM-menetelmällä laadituissa malleissa.

<sup>56</sup> Ks. tarkempi kuvaus ROC- ja CAP-käyristä esim. Sobehart ym. (2000) ja Engelmann ym. (2003).



Kuvio 4 Ennustamismallien ROC-käyrät testiryhmässä

Siniset käyrät kuvaavat ennustamismallien luokittelukykyä eri raja-arvoilla. Vihreä kuvaaja puolestaan vastaa täysin satunnaisen mallin luokittelukykyä, joka luokittelee yritykset 50 %:n kokonaistarkkuudella. Raja-arvoilla tarkoitetaan sitä kriittistä arvoa, jota suuremmat mallin antamat arvot luokitellaan toimiviksi yrityksiksi ja jota pienemmät arvot vastaavasti konkurssiyrityksiksi. Raja-arvon numeerista arvoa ei ole nähtävissä kuvioista 4, mutta se on mahdollista selvittää tilasto-ohjelmiston tulosteista kaikissa sinisen kuvaajan kohdissa.

Kriittinen raja-arvo laskee noustaessa pystyakselia ylöspäin pitkin sinistä kuvaajaa.<sup>57</sup> Kun se laskee riittävän alhaiseksi (sininen käyrä saavuttaa pystyakselin huipun), ei enää yksikään toimiva yritys tule luokitelluksi konkurssiyritykseksi, mutta samalla konkurssiyritysten luokittelutarkkuus on laskenut huomattavasti hyvin alhaisen kriittisen raja-arvon seurauksena. Tilanne on päinvastainen, kun sinistä kuvaajaa pitkin kuljetaan pystyakselia ylhäältä alaspäin. Kun saavutetaan vaaka-akselin vasen laita, ei enää yksikään konkurssiyritys tule luokitelluksi toimivaksi yritykseksi, mutta toimivien yritysten luokittelutarkkuus on heikentynyt huomattavasti hyvin korkean kriittisen raja-arvon johdosta. Optimaalinen kokonaisluokittelutarkkuus löytyy näiden ääriarvojen välistä. Se sijaitsee siinä kohdassa, missä kuvaajan tietyn pisteen etäisyydet yhteenlaskettuna sekä vasemmasta reunasta että yläreunasta ovat mahdollisimman pienet. Näin ollen täydellisen luokittelutarkkuuden omaavan mallin ROC-käyrä kulkisi suoraan ylöspäin vasemmasta alakulmasta vasempaan yläkulmaan ja jatkaisi siitä suoraan oikeaan yläkulmaan. Kokonaisennustamistarkkuus kullekin pisteelle sinisellä kuvaajalla saadaan laskemalla keskiarvo pystyakselin eli toimivien yritysten ennustamistarkkuuden ja vaaka-akselin arvosta johdettavissa olevan konkurssiyritysten ennustamistarkkuuden arvosta.<sup>58</sup>

Mallin suorituskykyä mitataan ROC-analyysin yhteydessä AUC-arvolla (area under the curve), joka mittaa ROC-käyrän kattaman alueen suuruutta eli sitä, kuinka suuri osa alueesta jää samanaikaisesti sen oikealle puolelle ja alapuolelle. Kuvion 4 ROC-käyrien AUC-arvot ovat 0,920 GLMM, 0,915 logit-malli, 0,918 lineaarinen erotteluanalyysi ja 0,849 Altmanin Z-malli.

GLMM-menetelmällä laadittu malli suoriutui siis myös ROC-analyysillä arvioiden parhaiten laadituista malleista. Lineaarinen erotteluanalyysi suoriutui tällä analyysitaivalla logit-mallia paremmin, vaikka sen kokonaisluokittelutarkkuus oli logit-mallia heikompi. Tarkastelemalla ROC-käyrien muotoa kuviossa 4 voidaan huomata, että lineaarisella erotteluanalyysillä laadittu malli saavuttaa pystyakselin ylälaidan selvästi aikaisemmin kuin muut mallit, joka kasvattaa AUC-arvoa suurentamalla ROC-käyrän peittoon jäävää pinta-alaa.

Altman ym. (1977, 44–48) toivat esille virhekustannusten huomioimisen konkurssin ennustamismallin käytön yhteydessä, joiden huomioiminen on myös ROC-analyysin ja sen potentiaalisen hyödyllisyyden kannalta olennainen tekijä. Virhekustannuksilla tarkoitetaan tässä yhteydessä sitä laskennallista kustannusta, joka päätöksentekijälle aiheutuu, mikäli hän arvioi yrityksen tulevan konkurssistatukseen väärin. Kuten tutkielmassa

<sup>57</sup> Tutkielman asetelmassa korkeampi mallin antama arvo tarkoittaa pienempää todennäköisyyttä ajautua konkurssiin kuin matalampi arvo.

<sup>58</sup> Esimerkiksi vaaka-akselin arvo 0,2 tarkoittaa 80 % ennustamistarkkuutta, sillä vaaka-akselin arvon kaava on  $1 - \text{konkurssiyritysten ennustamistarkkuus}$ .

on jo tullut esille, tämä arvio voi mennä väärin kahdella eri tavalla: konkurssiyritys arvioidaan toimivaksi yritykseksi (virhetyyppi I) tai toimiva yritys arvioidaan konkurssiyritykseksi (virhetyyppi II). Kaikilla päätöksentekijöillä päätöstilanteen konteksti ei kuitenkaan ole samanlainen, joten toisille päätöksentekijöille saattaa syntyä suurempia kustannuksia esimerkiksi virhetyypin I virhearvioinneista kuin toisille. Esimerkiksi vieraan pääoman ehtoinen sijoittaja saattaa menettää koko sijoituksensa virhetyypin I realisoituessa. Sen sijaan esimerkiksi tilintarkastajan tehdessä virhetyypin I virhearvioinnin (ei huomauta *going concern* -periaatteen uhattuna olemisesta) hän menettää mahdollisesti palkkionsa ja saattaa kärsiä jonkinasteisia mainetappioita, mutta joka tapauksessa virhekustannus on tyypiltään varsin erilainen kuin vieraan pääoman ehtoisella sijoittajalla.

Virhetyypin II virhearvioinnin tekevä tilintarkastaja saattaa puolestaan menettää asiakkaansa, jos *going concern* -periaatteen uhattuna olemisesta mainitaan tilintarkastuskertomuksessa aiheettomasti. Vastaavasti vieraan pääoman ehtoinen sijoittaja tehdessään virhetyypin II virhearvioinnin menettää vain lainaamatta jätetyn pääoman korkotuoton, joka sekin saattaa olla saatavissa vaihtoehtoisesta sijoituskohteesta riippuen tietenkin rahoituksen kysynnästä ja päätöksentekijän mahdollisuudesta tarjota sitä.

Matemaattisesti virhekustannukset voidaan esittää yksinkertaistetusti seuraavasti:

$$(kaava 8) \quad YK = o_1 K_1 V_1 + o_2 K_2 V_2$$

missä

- YK = laskennalliset yhteisvirhekustannukset ennustamismallin käytössä
- $o_1$  = konkurssiyritysten osuus (tilikausitasolla) yrityspopulaatiossa, johon ennustamismallia käytetään
- $K_1$  = laskennallinen virhekustannus virhetyypin I virhearvioinnista
- $V_1$  = virhetyypin I osuus ennustamismallissa tietyllä kriittisellä raja-arvolla
- $o_2, K_2, V_2$  = vastaavat toimivia yrityksiä ja virhetyyppejä II vastaavat arvot

Päätöksentekijän tulisi pyrkiä minimoimaan kaavan 8 arvo määrittämällä optimaalinen kriittinen raja-arvo ennustamismallissa. Raja-arvon muuttuessa kaavan komponenttien  $V_1$  ja  $V_2$  arvot muuttuvat. Esimerkiksi hypoteettisen päätöksentekijän, jonka virhetyypin II virhearvioinnin kustannus on todella suuri, kannattaisi kuvion 4 perusteella käyttää lineaarisella erotteluanalyysillä laadittua mallia, sillä se saavuttaa muita malleja aikaisemmin täydellisen virhetyyppien II välttämisen. Tämä pätee kuitenkin vain hyvin suurella virhetyypin I suhteellisella preferoinnilla, sillä GLMM-menetelmällä laadittu malli ohittaa toimivien yritysten oikein ennustamisessa lineaarisella erotteluanalyysillä laaditun mallin siirryttäessä kohti 60 %:n ja korkeampaa erottelutarkkuutta konkurssiyritysten osalta.

### 4.2.3 Altmanin Z-mallin soveltaminen aineistossa

Tutkielman aineistolla testattiin siinä laadittujen mallien lisäksi kaikkein tunnetuinta konkurssin ennustamismallia eli Altmanin Z-mallia, jota käsiteltiin luvussa 2.2.2. Tällä pyrittiin mahdollistamaan tutkielmassa laadittujen mallien toimivuuden kokonaisvaltaisempi arviointi tarjoamalla niiden suoritustasolle vertailukohta.

Käyttämällä Z-mallia sellaisenaan eli alkuperäisillä kertoimilla ja Altmanin (1968) määrittämällä kriittisellä raja-arvolla ennustamistarkkuudeksi saatiin tutkielman testiryhmässä 78,9 % virhetyypin I ollessa 25,3 % ja virhetyypin II 16,9 %. Altmanin tutkimuksessa kokonaisennustamistarkkuus oli 87 %, johon hänen mukaansa vaikutti heikentävästi keskimääräistä heikommin suoriutuvien toimivien yritysten valinta testiryhmään.

Z-mallin alkuperäisessä muodossa muuttuja  $X_4$  on *oman pääoman markkina-arvo/vieraan pääoman kirjanpitoarvo*, joka ei tutkielman aineiston yritysten kohdalla ollut laskettavissa, sillä ne eivät olleet listattuja yrityksiä, joten tunnusluvun osoittajana käytettiin oman pääoman kirjanpitoarvoa. Altman (2000) teki itse tämän vastaavan muutoksen muokatussa Z-mallissa listaamattomille yrityksille (käyttäen alkuperäistä Z-mallin aineistoa), jolloin muuttujien kertoimet myös muuttuivat. Muokatun mallin kertoimet olivat seuraavat:  $0,717(X_1) + 0,847(X_2) + 3,107(X_3) + 0,420(X_4) + 0,998(X_5)$ , jossa muuttuja  $X_4$  on siis *kirjanpidon oma pääoma/vieraan pääoman kirjanpitoarvo* muiden selittävien muuttujien pysyessä samoina kuin alkuperäisessä Z-mallissa.

Näillä muokatuilla kertoimilla ei päästy tutkielman testiryhmässä alkuperäisen Z-mallin ennustamistarkkuuteen. Altman ei määritellyt tarkkaa kriittistä raja-arvoa listaamattomille yrityksille tarkoitettussa mallissaan. Sen sijaan hän määrittä harmaan alueen 1,23 - 2,90 yrityksen Z-arvolle, jonka välillä oli päällekkäisyyksiä toimivien ja konkurssiyritysten saamista arvoissa. Kokeilemalla eri arvoja tältä väliltä saatiin tutkielman testiryhmässä parhaimmillaan 76,1 %:n ennustamistarkkuus (virhetyypin I 28,2 % ja virhetyypin II 19,7 %). Käyttämällä harmaan alueen keskikohtaa kriittisenä raja-arvona ennustamistarkkuus oli vain 73,9 % (virhetyypin I 40,9 % ja virhetyypin II 11,3 %).

Määrittämällä kertoimet uudestaan lineaarisella erotteluanalyysillä tutkielman estimointiryhmän datan perusteella saatiin ennustamismallin kaavaksi:  $0,173(X_1) + 1,041(X_2) + 2,349(X_3) + 0,126(X_4) - 0,061(X_5)$  kriittisen raja-arvon ollessa -0,045 (vastaa vakiotermiä 0,045). Uudet kertoimet eroavat melko olennaisesti Altmanin Z-mallista. Muuttujan  $X_5$  kohdalla jopa etumerkki vaihtui positiivisesta negatiiviseksi. Ennustamistarkkuudeksi tällä mallilla saatiin estimointiryhmässä 83,3 % virhetyypin I ollessa 20,8 % ja virhetyypin II 12,5 %. Testiryhmässä vastaavat prosentit olivat 83,8 %, 14,1 % ja 18,3 %.

Altmanin Z-mallin ennustamistarkkuus nousi siis 4,9 prosenttiyksikköä tutkielman testiryhmässä, kun sen kertoimet määritettiin uudestaan estimointiryhmän datan perus-

teella. Tosin tämä ero (z-arvo 1,07) ennustamistarkkuuksissa ei ollut vielä tilastollisesti merkitsevä, kun sen arvioimiseen käytettiin luvussa 4.2.1 käsitellyä kaksisuuntaista suhteellisten osuuksien vertailutestiä. Sen sijaan tutkielmassa parhaiten testiryhmässä ennustavan GLMM-menetelmällä laaditun mallin tarkkuuden (87,3 %) ja Altmanin alkuperäisen Z-mallin tarkkuuden (78,9 %) ero on hyvin lähellä tilastollista merkitsevyyttä (z-arvo 1,90). Vain yksi oikein luokiteltu yritys enemmän GLMM-menetelmän mallissa tai yksi väärin luokiteltu yritys enemmän Altmanin Z-mallissa olisi riittänyt tilastolliseen merkitsevyyteen 5 %:n riskitasolla ( $p < 0,05$ ), joka edellyttää z-arvoa 1,96.<sup>59</sup> Vertaattaessa GLMM-menetelmällä laaditun mallin tarkkuutta Altmanin listaamattomille yrityksille tarkoitetun mallin tarkkuuteen tilastollinen merkitsevyys saavutetaan ( $p = 0,004 - 0,014$  riippuen kriittisestä raja-arvosta, jota Altman ei määritellyt tarkkaan). Kohtuullisella varmuudella voidaan todeta, että varsin erilaisen aineiston pohjalta ja eri aikakautena laaditun Altmanin Z-mallin ennustamistarkkuus ei ole tutkielman aineiston määrittämässä perusjoukossa yhtä hyvä kuin tutkielmassa laaditun parhaiten ennustavan mallin.

---

<sup>59</sup> Konkurssin ennustamismallin tarkkuutta arvioitaessa tilastollisen merkitsevyyden raja-arvoa voidaan tutkielman laatijan näkemyksen mukaan arvioida käytetyssä testissä hieman normaalista poiketen, sillä ennustamismallien minimi-tarkkuus käytännössä lähtee 50 % tarkkuudesta, joka on keskimäärin saavutettavissa puhtaasti arvaamalla yritysten konkurssistatus ilman mitään tietoa yrityksestä. Esimerkiksi ( $p < 0,1$ ) edellyttäisi z-arvoa 1,65 arvon 1,96 sijaan.

## 5 JOHTOPÄÄTÖKSET JA YHTEENVETO

Tutkielmassa testattiin kolmea vaihtoehtoista tilastollista konkurssin ennustamismenetelmää samassa aineistossa. Aineisto rajattiin tavalla, jota ei aikaisemmissa tutkimuksissa ole tehty. Aineisto muodostui suomalaisista yrityksistä, jotka kokonsa puolesta henkilöstön lukumäärällä, taseen loppusummalla sekä liikevaihdolla mitattuna ylittivät tilintarkastusvelvollisuuden rajat. Tutkielmassa käytetyt tilastolliset menetelmät olivat lineaarinen erotteluanalyysi, logistinen regressio (logit-malli) sekä yleistetty lineaarinen sekamalli eli GLMM-menetelmä, jossa logistista regressiota on laajennettu sekamallilla. Viimeksi mainittua menetelmää ei ole tietävästi käytetty aikaisemmin konkurssin ennustamistutkimuksessa vastaavalla tavalla kuin tässä tutkielmassa. Kyseistä menetelmää voidaan pitää tilastotieteellisesti edistyneimpänä näistä kolmesta menetelmästä. Selittävien muuttujien valinta tehtiin kokeilemalla erilaisia yhdistelmiä alustavista potentiaalisista tunnusluvuista.

Aikaisemmat tutkimukset ovat osoittaneet, että pienten yritysten konkurssin ennustaminen tilinpäätösinformaation perusteella on haasteellista ja monissa tapauksissa myös mahdotonta (ks. esim. Edmister 1972 ja Pompe & Bilderbeek 2005). Ei ole esimerkiksi mahdollista ennustaa omistajavetoisen pienyrityksen omistajan henkilökohtaisista varoistaan tekemää pääomasijoitusta yritykseen, jolla hän mahdollisesti varmistaa myös työpaikkansa säilymisen omassa yrityksessään. Valtaosa suomalaisistakin yrityksistä on liikevaihdolla mitattuna hyvin pieniä. Lähes puolet yrityksistä vuonna 2014 oli sellaisia, joiden liikevaihto oli alle 40 000 euroa vuodessa. Tutkielmassa käytetty rajaus yritysten koon suhteen oli keskeinen osa tutkielman tutkimusasetelmaa, sillä se rajasi perusjoukon yrityksiin, jotka eivät ole listattuja suuryrityksiä, mutta eivät myöskään kaikkein pienimpiä yrityksiä, joiden tilinpäätösten ennustamiskykyyn liittyy merkittäviä epävarmuustekijöitä.

Etukäteen ei ollut edes likimääräisesti tiedossa, kuinka tarkka ennustamismalli käytetyllä aineiston rajauksella on mahdollista laatia. Rajauksen tuloksena yritykset olivat aineistossa edelleen pieniä verrattuna suurimpaan osaan konkurssin ennustamistutkimuksia. Näin ollen yksi mahdollisuus oli, että rajauksen koon suhteen pitäisi olla vielä korkeampi, jotta edes kohtuullinen ennustamistarkkuus saavutettaisiin.

Tutkielmassa laadittujen ennustamismallien tarkkuudet testiryhmässä olivat lineaarisen erotteluanalyysin (83,8 %) ja logistisen regression (85,9 %) osalta melko lähellä Azizin ja Darin (2006) meta-analyysissä esitettyjä keskimääräisiä tarkkuuksia näille menetelmille, jotka olivat 85,1 % lineaarisella erotteluanalyysillä ja 86,6 % logistisella regressiolla. Näin ollen voidaan päätellä, että käytetty rajaus yritysten koon suhteen oli riittävän korkealla menetelmien keskimääräisen ennustamistarkkuuden saavuttamiseksi. Tämä ei kuitenkaan automaattisesti tarkoita sitä, että laaditut mallit toimisivat vastaavalla tarkkuudella saman tarkkuusluokan saavuttaneiden tutkimusten aineistoissa, jotka

ovat saattaneet muodostua esimerkiksi amerikkalaisista listatuista yrityksistä. Tunnusluvut ja niiden painotukset voivat olla varsin erilaisia vastaavista ennustamistarkkuuksista huolimatta.

Tutkielmassa saavutettuihin ennustamistarkkuuksiin verrattuna meta-analyysin vastaaviin tarkkuuksiin saattoi vaikuttaa myös tilinpäätöspäivän ja tilinpäätöksen julkaisupäivän eriaikaisuuden huomioiminen tutkielmassa, joka vaikutti konkurssiyrityksen ennustamisen kannalta relevantin tilinpäätöksen valintaan. Suuressa osassa tutkimuksia ei ole huomioitu tätä seikkaa, jolla (huomioimatta jättämisellä) on lähtökohtaisesti ennustamistarkkuutta näennäisesti parantava vaikutus. Tällöin on mahdollista, että ennustaminen onkin vain jo tapahtuneen toteamista, sillä tilinpäätös on saatettu julkaista konkurssiin ajautumisen jälkeen. Ohlson (1980) otti tämän seikan huomioon, mutta myöhemmissä tutkimuksissa asia on usein sivuutettu johtuen mahdollisesti lähdeaineiston rajoituksista tai vaadittavan lisätyön määrästä, jonka tilinpäätösten julkaisupäivän selvittäminen tuo mukanaan.

Tutkielman ennustamismalleja laadittaessa esille nousi tiettyjä huomionarvoisia seikkoja liittyen selittäviin muuttujiin eli taloudellisiin tunnuslukuihin. *Omavaraisuusasteen* merkitys oli varsin suuri ja sillä oli mahdollista ennustaa konkurssia kohtuullisen hyvin myös yksittäisenä tunnuslukuna. Kannattavuuden tunnusluvuilla oli ennustamismalleja parantava vaikutus, mutta tilastollisen merkitsevyyden osalta ne eivät olleet *omavaraisuusasteen* tasolla.

Lyhyen aikavälin maksuvalmiutta mittaavan *current ration* ja lievemmin *quick ration* osalta tulokset olivat yllättäviä. Parhaiten ennustavissa malleissa näiden tunnuslukujen logaritimuunnosten etumerkit olivat negatiivisia, vaikka yleisesti näiden tunnuslukujen korkeaa arvoa pidetään suotavana yrityksen toimintaedellytyksiä ajatellen. Mahdollisia syitä tällä ilmiöllä on käsitelty luvussa 4.2.1. Yksi mahdollinen selitys on potentiaalisen konkurssiyrityksen myyntisaamisten ja muiden rahoitusomaisuuteen luettavien saamisten sekä vaihto-omaisuuden kirjanpidollinen yliarvostaminen. Yliarvostettujen erien arvostuksen korjaaminen heikentää suoraan esimerkiksi yrityksen *omavaraisuusastetta*, jolla tämänkin tutkielman tulosten perusteella on vahva yhteys konkurssiin ajautumiseen.

Aikaisemmissa tutkimuksissa ei tutkielman laatijan käsityksen mukaan ole esiintynyt ennustamismallia, jossa *current ration* tai *quick ration* tai näiden logaritimuunnosten etumerkki olisi ollut konkurssin todennäköisyyttä nostava eli yleensä negatiivinen (etumerkin suhde konkurssin todennäköisyyteen riippuu muuttujien koodauksesta aineistossa). Aiemmin mainittu tyypillinen puute konkurssin ennustamistutkimuksissa liittyen tilinpäätöspäivän huomioimatta jättämiseen saattaa olla kytköksissä siihen, että vastaavia tuloksia ei ole saatu *current ration* tai *quick ration* osalta. Näin siksi, että mikäli analysoitavaan aineistoon päätyy tilinpäätöksiä, jotka on laadittu yrityksen ollessa jo konkurssissa, on mahdolliset vaihto-omaisuuteen ja myyntisaamisiin liittyvät yliarvos-

tukset jo todennäköisesti korjattu. Tällöin näiden tunnuslukujen konkurssin uhkaa joissain tapauksissa lisäävä vaikutus jää ennustamismenetelmältä mallintamatta.

Varmuudella ei voida sanoa, selittykö ilmiö tutkielman aineistokohtaisella anomalialla vai esimerkiksi tutkielman aineiston määrittämään perusjoukkoon liittyvänä vaikeasti havaittavana ominaisuutena, jonka logistinen regressio (myös GLMM-laajennuksella) sai matemaattisen rakenteensa ansiosta mallinnettua. Aineistokohtainen anomalia olisi helpommin tehtävä johtopäätös, mikäli estimointiryhmän datan perusteella määräytynyt negatiivinen etumerkki kertoimineen ei olisi parantanut ennustamistarkkuutta myös testiryhmässä. Vähintäänkin on tehtävissä johtopäätös, että korkea *current ratio* tai *quick ratio* eivät ainakaan yksinään estä konkurssia järin tehokkaasti, mikäli vakavaraisuutta ja kannattavuutta mittaavien tunnuslukujen arvot ovat heikkoja.

Tutkielman merkittävin tutkimustulos liittyy tilastollisesti edistyneeseen GLMM-menetelmään, jolla tässä yhteydessä tarkoitetaan logistisen regression käyttöä sekamallilla laajennettuna. GLMM mahdollistaa normaalin logistisen regression kiinteiden vaikutusten eli valmiin ennustamismallin kaavan tunnuslukujen kertoimien ja vakiotermin lisäksi satunnaisvaikutusten huomioimisen, joiden avulla voidaan mallintaa eri hierarkiatasoja aineistossa, kuten eri toimialoja. Eri toimialojen yritysten tilinpäätöksistä johdettujen tunnuslukujen keskiarvot ja varianssit eroavat toisistaan, joka on todettavissa esimerkiksi Tilastokeskuksen tietokantoja selaamalla.

Tutkielmassa käytettiin 15:een eri toimialaluokituksen tyypistettyä luokittelua, joista jokainen toimi GLMM-menetelmällä laaditussa mallissa omana ryhmänään, jonka sisällä sai esiintyä muista ryhmistä poikkeavia satunnaisvaikutuksia selittävien muuttujien ja vakiotermin suhteen. Toisin sanoen toimialakohtaiset poikkeamat muusta aineistosta pyrittiin suodattamaan valmiin mallin kiinteistä vaikutuksista eli malliin valittujen tunnuslukujen kertoimista. Menetelmän tuottamat kiinteät kertoimet muodostavat matemaattiselta rakenteeltaan täysin identtisen ennustamismallin normaalin logistisen regression kanssa, joten näiden menetelmien tuottamien ennustamiskaavojen käytännön soveltaminen ei eroa toisistaan mitenkään.

GLMM-menetelmällä laaditun mallin ennustamistarkkuus testiryhmässä oli korkein testattavista eri menetelmillä laadituista malleista 87,3 %:n kokonaisennustamistarkkuudella. Tämä ennustamismalli suoriutui myös vaihtoehtoisia kriittisiä raja-arvoja huomioivan ROC-analyysin AUC-arvolla (0,920) mitattuna parhaiten. Mallin kaava (kaava 6) on esitetty tutkielman luvussa 4.2.1 muodossa, jossa sitä voidaan soveltaa tutkielman aineiston ulkopuolisiin yrityksiin. Logistisella regressiolla laadittu malli oli toiseksi tarkin (85,9 %) ja lineaarinen erotteluanalyysi vähiten tarkka (83,8 %), kun käytettiin parhaiten ennustavan mallin selittäviä muuttujia mallien laadinnassa. Kahden viimeksi mainitun menetelmän osalta tulokset myötäilivät aikaisempia tutkimuksia, joissa logit-malli on ollut hieman lineaarista erotteluanalyysia tarkempi (esim. Hamer 1983; Laitinen & Kankaanpää 1999 ja Aziz & Dar 2006). Menetelmien tarkkuusjärjes-

tys oli useimmiten sama myös muilla tunnuslukuyhdistelmäkokeiluilla kuin tutkimusraportissa esitetyillä tunnuslukuyhdistelmillä, joten myös tämä seikka tukee osaltaan menetelmien tarkkuusvertailussa saatuja tuloksia.

Metodisessa mielessä GLMM-menetelmän korkeampi ennustamistarkkuus suhteessa logistiseen regressioon ja lineaariseen erotteluanalyysiin on erityisen merkittävä siitä syystä, että kaikki nämä kolme menetelmää ovat varsin vertailukelpoisia keskenään käytännön soveltamisen kannalta. Näillä menetelmillä laaditut mallit ovat suhteellisen yksinkertaisessa ja intuitiivisesti ymmärrettävissä olevassa muodossa ja ennen kaikkea sovellettavissa sellaisenaan mihin tahansa yritykseen, jonka mallissa tarvittavat tunnusluvut on saatavilla tai laskettavissa tilinpäätöksen perusteella. Näin ollen varsin laajasti käytetyn perinteisen logistisen regressioon käyttö ei ole perusteltua, mikäli GLMM-menetelmällä on saatavissa korkeampia ennustamistarkkuuksia.

Tutkimustuloksiin liittyy edellä mainitusta huolimatta tutkimussuuntaukselle usein ominaista tilastollista epävarmuutta. Kaksisuuntaisella suhteellisten osuuksien vertailutestillä mitattuna ennustamistarkkuuksien erot menetelmien välillä eivät olleet tilastollisesti merkitseviä. Parhaiten ennustava GLMM-menetelmällä laadittu malli sen sijaan oli hyvin lähellä parempaa ennustamistarkkuutta 5 %:n riskitasolla verrattuna alkuperäisen Altmanin (1968) Z-mallin tarkkuuteen. Lisäksi yksittäisiin selittäviin muuttujiin ja niiden kontribuutioihin ennustamismalleissa liittyi tilastollista epävarmuutta lukuun ottamatta *omavaraisuusastetta*.

Vaihtoehtoisten ennustamismenetelmien ennustamistarkkuudet liikkuvat pääsääntöisesti muutamien prosenttiyksikköjen vaihteluvälillä, joten jo yhden tai kahden prosenttiyksikön korkeampi ennustamistarkkuus on merkittävä, mikäli se säilyy eri aineistoilla keskimäärin. Näin ollen tutkielman tulokset GLMM-menetelmällä laaditun ennustamismallin tarkkuuden osalta ovat rohkaisevia menetelmän käytettävyyden kannalta, vaikka tilastollisesti merkitseviä johtopäätöksiä voidaan tehdä rajallisesti. Tulokset joka tapauksessa antavat aiheutta tutkia menetelmän käyttöä konkurssin ennustamisen yhteydessä lisää erityisesti suurempia aineistoja hyödyntäen. Jo tämän tutkielman sekä rajoitetummasta menetelmän käyttötavasta huolimatta osittain myös Dakovicin ym. (2010) tulosten perusteella voidaan todeta, että on myös tilastollisesti todennäköisempää, että GLMM-menetelmällä on saavutettavissa parempi ennustamistarkkuus kuin lineaarisella erotteluanalyysillä tai logistisella regressiolla, vaikka tilastollisesti merkitsevä näyttö tästä vielä puuttuu.

Tutkielman aineisto ei ollut käytännön syistä ja käytettävissä olevista resursseista johtuen kovin suuri, vaikka se tosin oli selvästi suurempi kuin monissa erityisesti varhaisemmissa konkurssin ennustamistutkimuksissa. Tästä johtuen monissa toimialaluokituksissa oli aineistossa vain muutama yritys, joka luonnollisesti rajoitti GLMM-menetelmän edellytyksiä suodattaa toimialakohtaisia poikkeamia kiinteistä kertoimista

satunnaisten vaikutusten kautta. Suuremmalla aineistolla GLMM-menetelmän ominaisuudet olisivat mahdollisesti tulleet paremmin hyödynnetyksi.

Erityisen otollinen asetelma GLMM-menetelmälle verrattuna esimerkiksi logistiseen regressioon olisi sellainen, jossa ei käytettäisi vastinparimenettelyä aineiston muodostamisessa, joka puolestaan mahdollistaisi suuremman aineiston käyttämisen. Tällöin on mahdollista ottaa aineistoon toimivia yrityksiä enemmän kuin konkurssiyrityksiä, mikä vastaa myös todellista tilannetta yrityspopulaatiossa. Lisäksi tässä esimerkinomaisessa asetelmassa estimointiryhmän toimialojen jakauma poikkeaisi selvästi testiryhmästä tai siitä ryhmästä, missä sitä käytännössä käytettäisiin. Vastinparimenettely suodattaa jo osaltaan toimialakohtaisia vaikutuksia, kun konkurssiyrityksen vastinpari valitaan samalta toimialalta. Vastaavat toimialojen jakaumat estimointiryhmässä ja testiryhmässä puolestaan tasoittavat toimialakohtaisten poikkeamien osalta ennustamistarkkuutta testiryhmässä myös esimerkiksi logit-mallia käyttämällä, sillä laadittu malli on ikään kuin kompromissi juuri kyseisellä toimialojen jakaumalla. GLMM-menetelmällä sen sijaan toimialakohtaiset poikkeavuudet tulevat huomioiduksi myös ilman vastinparimenettelyä ja vastaavia toimialajakaumia estimointiryhmässä ja testiryhmässä. Tutkielman estimointiryhmän ja testiryhmän toimialojen jakaumat poikkesivat toisistaan, sillä ne määräytyivät satunnaisesti konkurssiyritysten konkurssiin ajautumisen päivämäärän perusteella.

Menetelmien komparatiivisilla analyyseilla ja konkurssin ennustamistutkimusten meta-analyyseilla on mahdollista arvioida vaihtoehtoisten ennustamismenetelmien suorituskykyä, mutta niihinkin liittyy tiettyä problematiikkaa. Menetelmien komparatiivisten analyyseiden tulokset pätevät varmuudella vain kyseisten tutkimusten aineistoilla. Konkurssin ennustamistutkimusten meta-analyyseissa puolestaan eri tutkimusten aineistot vaihtelevat suuresti. Näin ollen luotettavin tapa arvioida vaihtoehtoisten ennustamismenetelmien toimivuutta olisi meta-analyysi ennustamismenetelmien komparatiivisista analyyseista, joissa kussakin on käytetty tutkimuksen sisällä samaa aineistoa menetelmien vertailussa. Tämä puolestaan edellyttäisi lukumääräisesti useita menetelmien komparatiivisia analyyseja, joissa relevantteja menetelmiä verrattaisiin mieluiten suurilla aineistoilla, jotka ovat puolestaan riippuvaisia käytettävissä olevasta tietokannasta. On tietenkin myös mahdollista, että yksittäisessä tutkimuksessa käytetään useita erilaisia aineistoja, joilla vaihtoehtoisia menetelmiä verrataan. Tällöin meta-analyysin tarve ei olisi niin suuri menetelmien ennustamistarkkuuksien luotettavassa arvioinnissa.

Erilaisilla tietokannoilla on merkittävä rooli tilastollisen konkurssin ennustamistutkimuksen mahdollistajana. Näitä tietokantoja ei kuitenkaan laadita ensisijaisesti konkurssin ennustamistutkimuksen tarpeita ajatellen. Esimerkiksi tilinpäätöksen julkaisupäivämäärä ja yrityksen konkurssistatus piti tutkielmassa selvittää erikseen eri tietolähteestä kuin yritysten tilinpäätösdata. Lisäksi tunnusluvut, joita ei ollut tietokannassa valmiiksi laskettu, piti laskea manuaalisesti yksitellen. Tilanne on kuitenkin yleisesti

ottaen selvästi parempi kuin varhaisimpien konkurssin ennustamistutkimusten aikana, mutta edelleen on paljon tehtävissä ainakin konkurssin ennustamistutkimuksen näkökulmasta liittyen datan keräämiseen ja sen saattamiseen tehokkaasti analysoitavaan muotoon. Vain runsas ja relevantti sekä helposti käsiteltävissä oleva data mahdollistaa edistyneiden tilastollisten menetelmien – kuten yleistetyn lineaarisen sekamallin – täysimittaisen hyödyntämisen.

## LÄHTEET

- Altman, E. I. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, Vol. 23 (4), 589–609.
- Altman, E. I. (1977) ZETA analysis: A new model to identify bankruptcy risk of corporations. *Journal of Banking and Finance*, Vol. 1 (1), 29–54.
- Altman, E. I. (2000) Predicting financial distress of companies: Revisiting the Z-score and Zeta models. *Working Paper, Dept of Finance*. New York University, New York, NY.
- Argenti, J. (1976) Corporate planning and corporate collapse. *Long Range Planning*, Vol. 9 (6), 12–17.
- Argenti, J. (1983) Predicting corporate failure. *Accountants Digest*, Vol. 138 (Summer), 1–25.
- Aziz, M. – Dar, H. (2006) Predicting corporate bankruptcy: Where we stand. *Corporate Governance*, Vol. 6 (1), 18–33.
- Beaver, W. H. (1966) Financial ratios as predictors of failure. *Empirical Research in Accounting: Selected Studies*, Supplement to Vol. 4 (3), 71–127.
- Beaver, W. H. – McNichols M. F. – Rhie, J. (2005) Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies*, Vol. 10 (1), 93–122.
- Casey, C. J. Jr (1980) Variations in accounting information load: The effect on loan officers. *The Accounting Review*, Vol. 55 (1), 36–49.
- Chesser, D. L. (1974) Predicting loan noncompliance. *The Journal of Commercial Bank Lending*, Vol. 56 (12), 28–38.
- Dakovic, R. – Czado, C. – Berg, D. (2010) Bankruptcy prediction in Norway: a comparison study. *Applied Economics Letters*, Vol. 17 (17), 1739–1746.
- Demidenko, E. (2013) *Wiley Series in Probability and Statistics: Mixed models: Theory and applications with R*. Wiley, Hoboken, NJ.
- Durand, D. D. (1941) Risk elements in consumer installment financing. Teoksessa: *Studies in Consumer Installment Financing*, 105–142. National Bureau of Economic Research, New York, NY.
- Edmister, R. (1972) An empirical test of financial ratio analysis for small business failure prediction. *Journal of Financial and Quantitative Analysis*, Vol.7 (2), 1477–1493.
- El Hennawy, R. H. A. – Morris, R. C. (1983) The significance of base year in developing failure prediction models. *Journal of Business Finance & Accounting*, Vol. 10 (2), 209–223.

- Engelmann, B. – Hayden, E. – Tasche, D. (2003) Testing rating accuracy. *Risk*, Vol. 16 (1), 82–86.
- Euroopan komissio (2003) Komission suositus 2003/361/EY mikroyritysten sekä pienten ja keskisuurten yritysten määritelmästä.
- Hair, J. F. Jr – Black, W. C. – Babin, B. J. – Anderson, R. E. (2010) *Multivariate data analysis: A global perspective*. Seventh edition. Pearson Education, Inc., Upper Saddle River, NJ.
- Hamer, M. (1983) Failure prediction: Sensitivity of classification accuracy to alternative statistical methods and variable sets. *Journal of Accounting and Public Policy*, Vol. 2 (4), 289–307.
- HE 26/2003 vp: Hallituksen esitys Eduskunnalle konkurssilainsäädännön uudistamiseksi.
- Herranz, N. – Krasa, S. – Villamil, A. (2008) *The impact of owners and policy on small firms*. University of Illinois, Normal, IL.
- Herranz, N. – Krasa, S. – Villamil, A. P. (2009) Small firms in the SSBF. *Annals of Finance*, Vol. 5 (3–4), 341–359.
- Holopainen, M. – Pulkkinen, P. (2013) *Tilastolliset menetelmät*. 5.–8. painos. Sanoma Pro Oy, Helsinki.
- Huyghebaert, N. – Gaeremynck, A. – Roodhooft, F. – Gucht, L. M., Van de. (2000) New firm survival: The effects of start-up characteristics. *Journal of Business Finance & Accounting*, Vol. 27 (5–6), 627–651.
- International Accounting Standards Board (IASB) (2015) *Conceptual framework for financial reporting*. Exposure Draft ED/2015/3. <[http://www.ifrs.org/Current-Projects/IASB-Projects/Conceptual-Framework/Documents/May%202015/ED\\_CF\\_MAY%202015.pdf](http://www.ifrs.org/Current-Projects/IASB-Projects/Conceptual-Framework/Documents/May%202015/ED_CF_MAY%202015.pdf)>, haettu 30.8.2016.
- Jackson, R. – Wood, A. (2013) The performance of insolvency prediction and credit risk models in the UK: A comparative study. *The British Accounting Review*, Vol. 45 (3), 183–202.
- Johnson, R. A. – Wichern, D. W. (2002) *Applied multivariate statistical analysis*. Fifth edition. Pearson Education, Inc., Upper Saddle River, NJ.
- Jones, S. – Hensher, D. A. (2004) Predicting firm financial distress: A mixed logit model. *The Accounting Review*, Vol. 79 (4), 1011–1038.
- Keasey, K. – Watson, R. (1987) Non-financial symptoms and the prediction of small company failure: A test of Argenti's hypotheses. *Journal of Business Finance & Accounting*, Vol. 14 (3), 335–354.
- Kirjanpitolaki 30.12.1997/1336.
- Konkurssilaki 20.2.2004/120.

- Koulu, R. – Havansi, E. – Korkea-Aho, E. – Lindfors, H. – Niemi, J. (2009) *Insolvenssioikeus*. WSOYpro, Helsinki.
- KvantiMOTV - Menetelmäopetuksen tietovaranto (2016) *Faktorianalyysi [verkköjulkaisu]*. Yhteiskuntatieteellinen tietoarkisto, Tampere. <<http://www.fsd.uta.fi/menetelmaopetus/faktori/faktori.html>>, haettu 30.8.2016.
- Lachenbruch, P. A. (1975) *Discriminant analysis*. Hafner Press, cop, New York, NY.
- Laitinen, E. K. (1990) *Konkurssin ennustaminen*. Vaasan Yritysinformaatio Oy, Sundom.
- Laitinen, E. K. (1992) Prediction of failure of a newly founded firm. *Journal of Business Venturing*, Vol. 7 (4), 323–340.
- Laitinen, E. K. – Laitinen, T. (1998) Cash management behavior and failure prediction. *Journal of Business Finance & Accounting*, Vol. 25 (7–8), 893–919.
- Laitinen, E. K. – Laitinen, T. (2004) *Yrityksen rahoituskriisin ennustaminen*. Talentum, Helsinki.
- Laitinen, T. – Kankaanpää, M. (1999) Comparative analysis of failure prediction methods: The Finnish case. *The European Accounting Review*, Vol. 8 (1), 67–92.
- Libby, R. (1975) Accounting ratios and the prediction of failure: Some behavioral evidence. *Journal of Accounting Research*, Vol. 13 (1), 150–161.
- Martin, D. (1977) Early warning of bank failure: A logit regression approach. *Journal of Banking Finance*, Vol. 1 (3), 249–276.
- Merwin, C. L. (1942) Financing small corporations in five manufacturing industries 1926–36. *National Bureau of Economic Research*, New York, NY.
- Ohlson, J. (1980) Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, Vol. 18 (1), 109–131.
- Peng, C. J. – Lee, K. L. – Ingersoll, G. M. (2002) An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, Vol. 96 (1), 3–14.
- Pinches, G. E. – Eubank, A. A. – Mingo, K. A. – Caruthers J. K. (1975) The hierarchical classification of financial ratios. *Journal of business research*, Vol. 3 (4), 295–310.
- Pinches, G. E. – Mingo, K. A. – Caruthers, J. K. (1973) The stability of financial patterns in industrial organizations. *The Journal of Finance*, Vol. 28, (2), 389–396.
- Platt, H. D. – Platt, M. B. (1990) Development of a class of stable predictive variables: The case of bankruptcy prediction. *Journal of Business Finance & Accounting*, Vol. 17 (1), 31–51.

- Pompe, P. P. M. – Bilderbeek, J. (2005) The prediction of bankruptcy of small- and medium-sized industrial firms. *Journal of Business Venturing*, Vol. 20 (6), 847–868.
- Prihti, A. (1975) *Konkurssin ennustaminen taseinformaation avulla*. Painomies Oy, Helsinki.
- Riahi-Belkaoui, A. (2004) *Accounting theory*. Fifth edition. Thomson Learning, London.
- Richardson, F. M. – Davidson, L. F. (1983) An exploration into bankruptcy discriminant model sensitivity. *Journal of Business Finance & Accounting*, Vol. 10 (2), 195–207.
- Santomero A. M. – Vinso, J. D. (1977) Estimating the probability of failure for commercial banks and the banking system. *Journal of Banking and Finance*, Vol. 1 (2–3), 185–205.
- Sobehart, J. R. – Keenan, S. – Stein, R. (2000) Benchmarking quantitative default risk models: A validation methodology. *Moody's Investors Service*. Moody's Investors Service, Inc., New York, NY.
- Suomen Asiakastieto Oy (2016) Voitto+ CD. <<http://www.asiakastieto.fi/web/fi/tuotteet-ja-palvelut/riskienhallinnan-palvelut/yrityspalvelut/voitto>>, haettu 30.8.2016.
- Suomen virallinen tilasto (SVT) (2015) *Yritysten rakenne- ja tilinpäätöstilasto [verkköjulkaisu]*. Tilastokeskus, Helsinki. <[http://www.stat.fi/til/yrti/2014/yrti\\_2014\\_2015-12-18\\_tie\\_001\\_fi.html](http://www.stat.fi/til/yrti/2014/yrti_2014_2015-12-18_tie_001_fi.html)>, haettu 30.8.2016.
- Taffler, R. J. (1982) Forecasting company failure in the UK using discriminant analysis and financial ratio data. *Journal of the Royal Statistical Society*, Vol. 145 (3), 342–358.
- Tilintarkastuslaki 18.9.2015/1141.
- Trabelsi, S. – He, R. – He, L. – Kusy, M. (2015) A comparison of Bayesian, hazard, and mixed logit model of bankruptcy prediction. *Computational Management Science*, Vol. 12 (1), 81–97.
- Yritystutkimus ry (2011) *Yritystutkimuksen tilinpäätösanalyysi*, 9. korjattu laitos. Gaudeamus, Helsinki.
- Wall, A. – Duning, R. W. (1928) *Ratio analysis of financial statements*. Harper and Brothers, New York, NY.
- Walter, J. E. (1959) A discriminant function for earnings price ratios of large industrial corporations. *Review of Economics and Statistics*, Vol. 41 (February), 44–52.
- Wilcox, J. W. (1971) A simple theory of financial ratios as predictors of failure. *Journal of Accounting Research*, Vol. 9 (2), 389–395.

LIITE 1 SPSS-TULOSTEET TAULUKON 6  
ENNUSTAMISMALLEISSA

### Discriminant

**Analysis Case Processing Summary**

Unweighted Cases		N	Percent
Valid		144	100,0
Excluded	Missing or out-of-range group codes	0	,0
	At least one missing discriminating variable	0	,0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	,0
	Total	0	,0
Total		144	100,0

**Group Statistics**

Konk		Valid N (listwise)	
		Unweighted	Weighted
0	ROI	72	72,000
	CRlog	72	72,000
	OMAV%	72	72,000
	KK1-2	72	72,000
1	ROI	72	72,000
	CRlog	72	72,000
	OMAV%	72	72,000
	KK1-2	72	72,000
Total	ROI	144	144,000
	CRlog	144	144,000
	OMAV%	144	144,000
	KK1-2	144	144,000

### Analysis 1

## Summary of Canonical Discriminant Functions

### Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,875 <sup>a</sup>	100,0	100,0	,683

a. First 1 canonical discriminant functions were used in the analysis.

### Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,533	88,009	4	,000

### Standardized Canonical Discriminant Function Coefficients

	Function
	1
ROI	,325
CRlog	,246
OMAV%	,626
KK1-2	,160

### Structure Matrix

	Function
	1
OMAV%	,904
KK1-2	,635
CRlog	,633
ROI	,544

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

**Canonical Discriminant**

**Function Coefficients**

	Function
	1
ROI	,687
CRlog	,323
OMAV%	1,206
KK1-2	1,325
(Constant)	-,190

Unstandardized coefficients

**Functions at Group**

**Centroids**

	Function
	1
Konk	
0	,929
1	-,929

Unstandardized canonical discriminant functions evaluated at group means

**Classification Statistics**

**Classification Processing Summary**

Processed	144
Excluded	0
Missing or out-of-range group codes	
At least one missing discriminating variable	0
Used in Output	144

**Prior Probabilities for Groups**

Konk	Prior	Cases Used in Analysis	
		Unweighted	Weighted
0	,500	72	72,000
1	,500	72	72,000
Total	1,000	144	144,000

**Classification Function Coefficients**

	Konk	
	0	1
ROI	,770	-,506
CRlog	,630	,030
OMAV%	,680	-1,561
KK1-2	3,882	1,421
(Constant)	-1,365	-1,013

Fisher's linear discriminant functions

**Classification Results<sup>a</sup>**

			Predicted Group Membership		Total
			0	1	
Original	Count	Konk 0	65	7	72
		1	13	59	72
	%	Konk 0	90,3	9,7	100,0
		1	18,1	81,9	100,0

a. 86,1% of original grouped cases correctly classified.

**Logistic Regression****Case Processing Summary**

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	144	100,0
	Missing Cases	0	,0
	Total	144	100,0
Unselected Cases		0	,0
Total		144	100,0

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

Original Value	Internal Value
0	0
1	1

## Block 0: Beginning Block

Classification Table<sup>a,b</sup>

Observed		Predicted		
		Konk		Percentage Correct
		0	1	
Step 0	Konk	0	72	,0
		0	72	100,0
Overall Percentage				50,0

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,000	,167	,000	1	1,000	1,000

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	OMAV	60,044	1	,000
		KK12	37,581	1	,000
		CRlog	37,394	1	,000
		ROI	29,616	1	,000
Overall Statistics			67,202	4	,000

## Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	115,811	4	,000
	Block	115,811	4	,000
	Model	115,811	4	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	83,816 <sup>a</sup>	,553	,737

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than ,001.

Classification Table<sup>a</sup>

Observed		Predicted			
		Konk		Percentage Correct	
		0	1		
Step 1	Konk	0	59	13	81,9
		1	9	63	87,5
Overall Percentage					84,7

a. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	OMAV	-6,312	1,538	16,843	1	,000	,002
	KK12	-7,676	4,827	2,529	1	,112	,000
	CRlog	,155	,542	,082	1	,775	1,168
	ROI	-,393	,579	,460	1	,498	,675
	Constant	1,378	,387	12,674	1	,000	3,965

a. Variable(s) entered on step 1: OMAV, KK12, CRlog, ROI.

## Generalized Linear Mixed Models

Case Processing Summary

	N	Percent
Included	144	100,0%
Excluded	0	0,0%
Total	144	100,0%

Model Summary

Target	Konk	
Probability Distribution	Binomial	
Link Function	Logit	
Information Criterion	Akaike Corrected	1040,493
	Bayesian	1054,714

Information criteria are based on the -2 log likelihood (1030,042) and are used to compare models. Models with smaller information criterion values fit better.

**Data Structure<sup>a</sup>**

	Subjects	Target
	TA	Konk
Data for First Subject	A	1
	A	0
	A	1
	A	0
Total Number of Levels	11	

a. Target: Konk

**Classification****Overall Percent Correct = 91,0%<sup>a</sup>**

Observed		Predicted	
		0	1
0	Count	66	6
	% within Observed	91,7%	8,3%
1	Count	7	65
	% within Observed	9,7%	90,3%

a. Target: Konk

**Fixed Effects<sup>a</sup>**

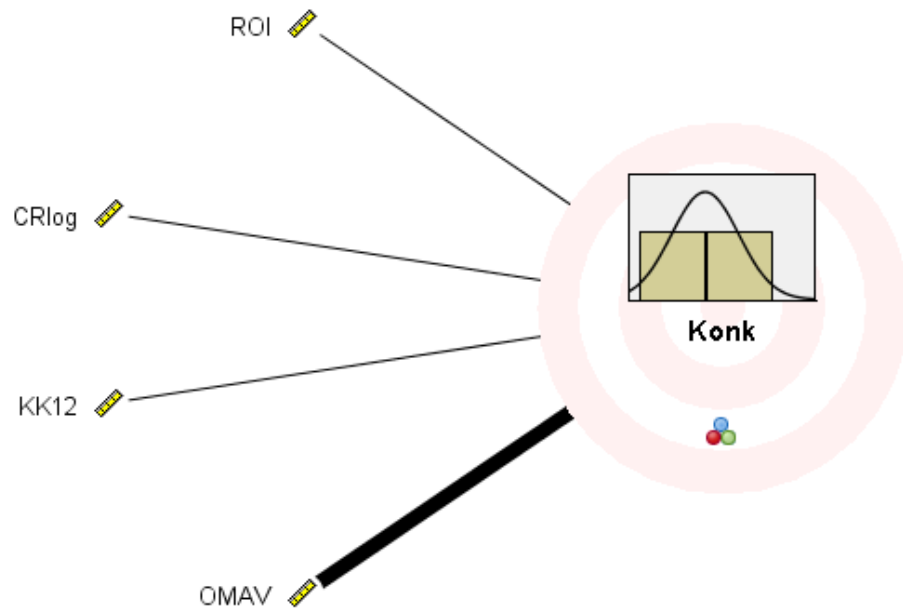
Source	F	df1	df2	Sig.
Corrected Model	5,672	4	139	,000
ROI	1,116	1	139	,293
CRlog	,205	1	139	,651
KK12	,796	1	139	,374
OMAV	15,296	1	139	,000

Probability distribution: Binomial

Link function: Logit<sup>a</sup>

a. Target: Konk

**Fixed Effects**



**Fixed Coefficients<sup>a</sup>**

Model Term	Coefficient	Std. Error	t	Sig.	95% Confidence Interval	
					Lower	Upper
Intercept	-1,652	,4605	-3,587	,000	-2,562	-,741
ROI	2,466	2,3341	1,057	,293	-2,149	7,081
CRlog	-,342	,7550	-,453	,651	-1,835	1,151
KK12	5,696	6,3852	,892	,374	-6,929	18,321
OMAV	7,983	2,0411	3,911	,000	3,947	12,018

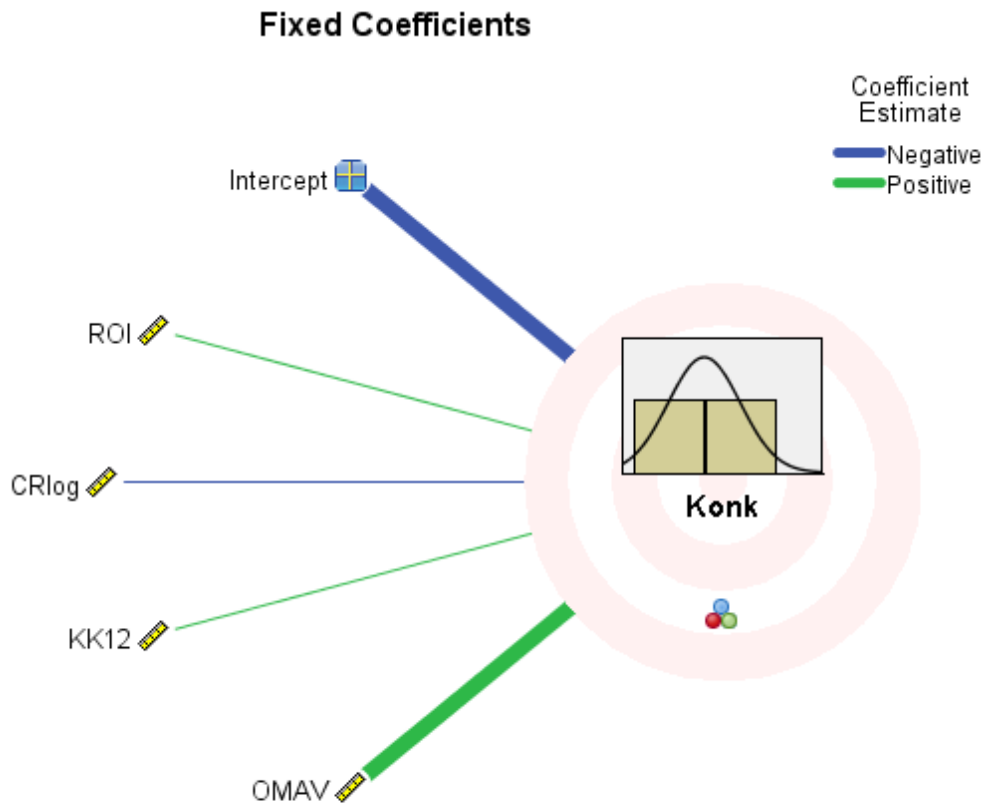
**Fixed Coefficients<sup>a</sup>**

Model Term	Exp(Coefficient)	95% Confidence Interval for Exp(Coefficient)	
		Lower	Upper
Intercept	,192	,077	,476
ROI	11,775	,117	1188,989
CRlog	,710	,160	3,161
KK12	297,642	,001	90474123,244
OMAV	2930,141	51,787	165789,272

Probability distribution: Binomial

Link function: Logit<sup>a</sup>

a. Target: Konk



### Random Effect Covariances

**Random Effect Block 1**

Random Effect Block	Intercept	ROI	OMAV	KK12	CRlog
Intercept	,000	,000	,000	,000	,000
ROI	,000	13,668	,000	,000	,000
OMAV	,000	,000	1,125	,000	,000
KK12	,000	,000	,000	,000	,000
CRlog	,000	,000	,000	,000	1,324

Covariance Structure: Variance components

Subject Specification: TA

### Covariance Parameters

**Covariance Parameters Summary**

Covariance Parameters	Residual Effect	0
	Random Effects	5
Design Matrix Columns	Fixed Effects	5
	Random Effects	5 <sup>a</sup>
Common Subjects		11

Common subjects are based on the subject specifications for the residual and random effects and are used to chunk the data for better performance.

a. This is the number of columns per common subject.

**Residual Effect**

Residual Effect	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower	Upper
Variance	1,000	.	.	.	.	.

Covariance Structure: Scaled Identity

Subject Specification: (None)

**Random Effect**

Random Effect Covariance	Estimate	Std. Error	Z	Sig.	95% Confidence Interval
					Lower
Var(Intercept)	2,291E-7 <sup>a</sup>	.	.	.	.
Var(ROI)	13,668	16,272	,840	,401	1,325
Var(OMAV)	1,125	11,332	,099	,921	3,006E-9
Var(KK12)	1,047E-11 <sup>a</sup>	.	.	.	.
Var(CRlog)	1,324	2,250	,589	,556	,047

**Random Effect**

Random Effect Covariance	95% Confidence Interval
	Upper
Var(Intercept)	.
Var(ROI)	140,965
Var(OMAV)	421129863,042
Var(KK12)	.
Var(CRlog)	36,989

Covariance Structure: Variance components

Subject Specification: TA

a. This parameter is redundant.

## Discriminant

### Analysis Case Processing Summary

Unweighted Cases		N	Percent
Valid		144	100,0
Excluded	Missing or out-of-range group codes	0	,0
	At least one missing discriminating variable	0	,0
	Both missing or out-of-range group codes and at least one missing discriminating variable	0	,0
Total		0	,0
Total		144	100,0

### Group Statistics

Konk		Valid N (listwise)	
		Unweighted	Weighted
0	CRlog	72	72,000
	OMAV%	72	72,000
	ROA	72	72,000
	LVP1-2	72	72,000
1	CRlog	72	72,000
	OMAV%	72	72,000
	ROA	72	72,000
	LVP1-2	72	72,000
Total	CRlog	144	144,000
	OMAV%	144	144,000
	ROA	144	144,000
	LVP1-2	144	144,000

## Analysis 1

## Summary of Canonical Discriminant Functions

### Eigenvalues

Function	Eigenvalue	% of Variance	Cumulative %	Canonical Correlation
1	,945 <sup>a</sup>	100,0	100,0	,697

a. First 1 canonical discriminant functions were used in the analysis.

### Wilks' Lambda

Test of Function(s)	Wilks' Lambda	Chi-square	df	Sig.
1	,514	93,125	4	,000

### Standardized Canonical Discriminant

#### Function Coefficients

	Function
	1
CRlog	,193
OMAV%	,668
ROA	,512
LVP1-2	-,077

#### Structure Matrix

	Function
	1
OMAV%	,870
ROA	,679
CRlog	,609
LVP1-2	,607

Pooled within-groups correlations between discriminating variables and standardized canonical discriminant functions

Variables ordered by absolute size of correlation within function.

**Canonical Discriminant****Function Coefficients**

	Function
	1
CRlog	,254
OMAV%	1,288
ROA	2,154
LVP1-2	-,646
(Constant)	-,195

Unstandardized coefficients

**Functions at Group****Centroids**

	Function
Konk	1
0	,965
1	-,965

Unstandardized canonical discriminant functions evaluated at group means

**Classification Statistics****Classification Processing Summary**

Processed	144
Excluded	
Missing or out-of-range group codes	0
At least one missing discriminating variable	0
Used in Output	144

**Prior Probabilities for Groups**

Konk	Prior	Cases Used in Analysis	
		Unweighted	Weighted
0	,500	72	72,000
1	,500	72	72,000
Total	1,000	144	144,000

**Classification Function Coefficients**

	Konk	
	0	1
CRlog	,506	,017
OMAV%	1,247	-1,240
ROA	3,402	-,755
LVP1-2	-2,574	-1,327
(Constant)	-1,399	-1,023

Fisher's linear discriminant functions

**Classification Results<sup>a</sup>**

			Predicted Group Membership		Total
			0	1	
Original	Count	Konk 0	65	7	72
		1	11	61	72
	%	Konk 0	90,3	9,7	100,0
		1	15,3	84,7	100,0

a. 87,5% of original grouped cases correctly classified.

**Logistic Regression****Case Processing Summary**

Unweighted Cases <sup>a</sup>		N	Percent
Selected Cases	Included in Analysis	144	100,0
	Missing Cases	0	,0
	Total	144	100,0
Unselected Cases		0	,0
Total		144	100,0

a. If weight is in effect, see classification table for the total number of cases.

**Dependent Variable Encoding**

Original Value	Internal Value
0	0
1	1

**Block 0: Beginning Block**

Classification Table<sup>a,b</sup>

		Predicted		
		Konk		Percentage Correct
Observed		0	1	
Step 0	Konk 0	0	72	,0
	1	0	72	100,0
Overall Percentage				50,0

a. Constant is included in the model.

b. The cut value is ,500

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	,000	,167	,000	1	1,000	1,000

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	OMAV	60,044	1	,000
		CRlog	37,394	1	,000
		LVP12	37,215	1	,000
		ROA	43,704	1	,000
Overall Statistics			69,958	4	,000

## Block 1: Method = Enter

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	117,005	4	,000
	Block	117,005	4	,000
	Model	117,005	4	,000

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	82,621 <sup>a</sup>	,556	,742

a. Estimation terminated at iteration number 8 because parameter estimates changed by less than ,001.

**Classification Table<sup>a</sup>**

Observed		Predicted			
		Konk		Percentage Correct	
		0	1		
Step 1	Konk	0	60	12	83,3
		1	9	63	87,5
	Overall Percentage				85,4

a. The cut value is ,500

**Variables in the Equation**

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 <sup>a</sup>	OMAV	-6,451	1,536	17,642	1	,000	,002
	CRlog	,435	,497	,764	1	,382	1,544
	LVP12	-9,590	7,157	1,796	1	,180	,000
	ROA	-1,004	2,239	,201	1	,654	,366
	Constant	1,163	,369	9,918	1	,002	3,199

a. Variable(s) entered on step 1: OMAV, CRlog, LVP12, ROA.

## Generalized Linear Mixed Models

**Case Processing Summary**

	N	Percent
Included	144	100,0%
Excluded	0	0,0%
Total	144	100,0%

**Model Summary**

Target	Konk	
Probability Distribution	Binomial	
Link Function	Logit	
Information Criterion	Akaike Corrected	1082,489
	Bayesian	1096,710

Information criteria are based on the -2 log likelihood (1072,038) and are used to compare models. Models with smaller information criterion values fit better.

**Data Structure<sup>a</sup>**

	Subjects	Target
	TA	Konk
Data for First Subject	A	1
	A	0
	A	1
	A	0
Total Number of Levels	11	

a. Target: Konk

**Classification****Overall Percent Correct = 92,4%<sup>a</sup>**

Observed		Predicted	
		0	1
0	Count	68	4
	% within Observed	94,4%	5,6%
1	Count	7	65
	% within Observed	9,7%	90,3%

a. Target: Konk

**Fixed Effects<sup>a</sup>**

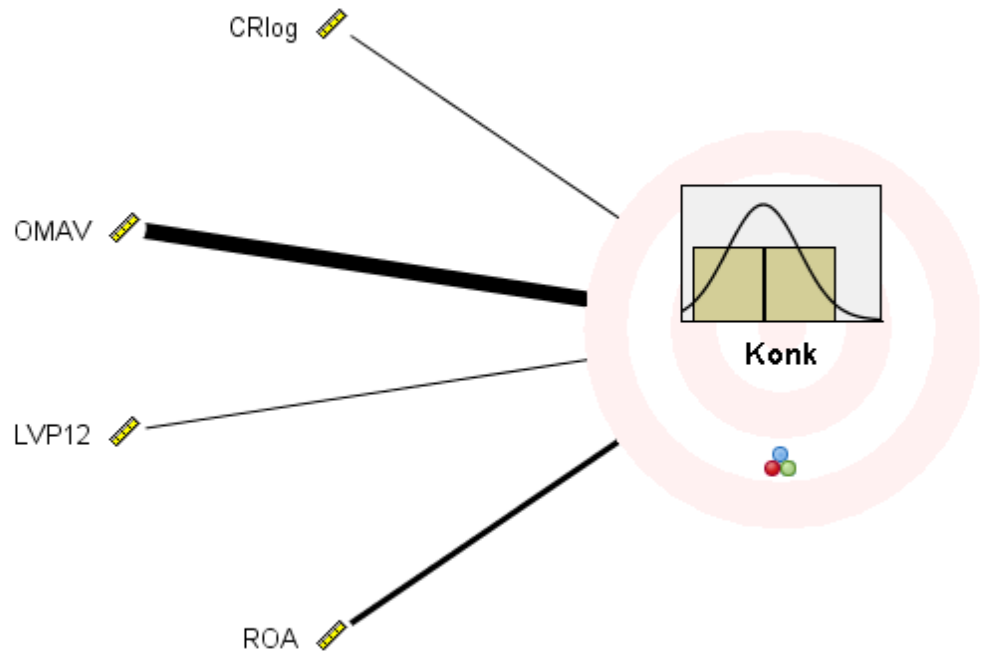
Source	F	df1	df2	Sig.
Corrected Model	4,772	4	139	,001
CRlog	,934	1	139	,335
OMAV	13,167	1	139	,000
LVP12	,047	1	139	,829
ROA	3,016	1	139	,085

Probability distribution: Binomial

Link function: Logit<sup>a</sup>

a. Target: Konk

**Fixed Effects**



**Fixed Coefficients<sup>a</sup>**

Model Term	Coefficient	Std. Error	t	Sig.	95% Confidence Interval	
					Lower	Upper
Intercept	-1,572	,4496	-3,496	,001	-2,461	-,683
CRlog	-,664	,6874	-,967	,335	-2,024	,695
OMAV	7,559	2,0833	3,629	,000	3,441	11,678
LVP12	1,528	7,0546	,217	,829	-12,421	15,476
ROA	9,112	5,2470	1,737	,085	-1,262	19,487

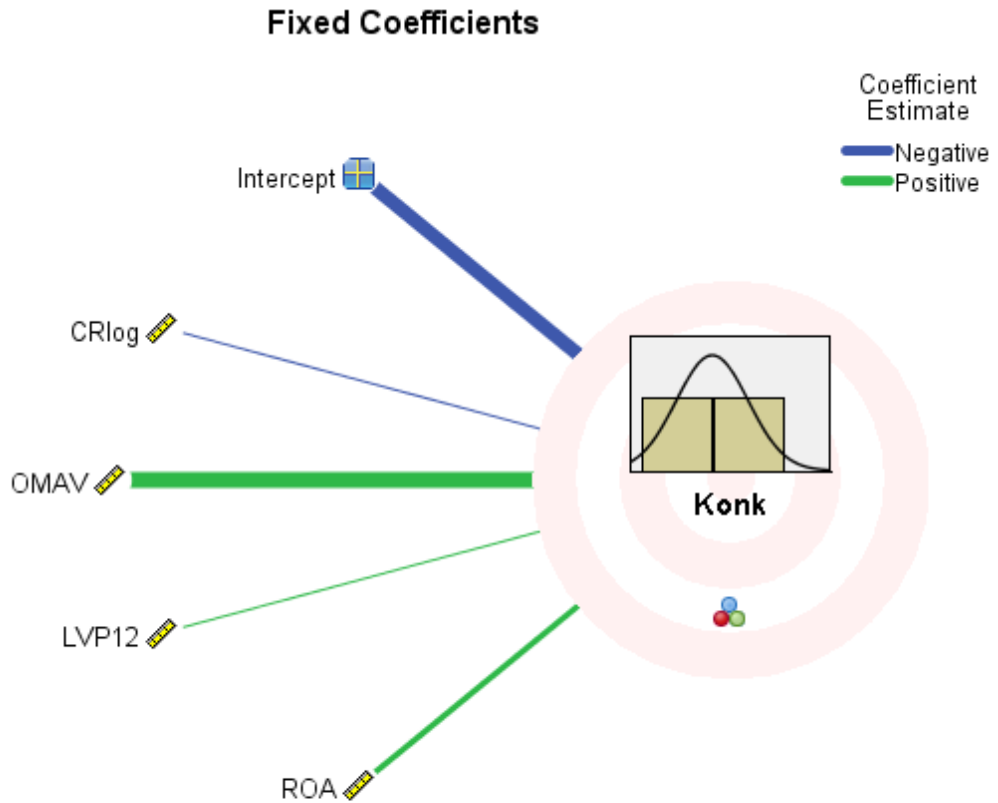
**Fixed Coefficients<sup>a</sup>**

Model Term	Exp(Coefficient)	95% Confidence Interval for Exp(Coefficient)	
		Lower	Upper
Intercept	,208	,085	,505
CRlog	,515	,132	2,003
OMAV	1918,869	31,204	118000,894
LVP12	4,607	4,034E-6	5261133,751
ROA	9066,944	,283	290351929,893

Probability distribution: Binomial

Link function: Logit<sup>a</sup>

a. Target: Konk



### Random Effect Covariances

Random Effect Block 1

Random Effect Block	Intercept	OMAV	CRlog	LVP12	ROA
Intercept	,000	,000	,000	,000	,000
OMAV	,000	3,024	,000	,000	,000
CRlog	,000	,000	,961	,000	,000
LVP12	,000	,000	,000	,000	,000
ROA	,000	,000	,000	,000	79,282

Covariance Structure: Variance components

Subject Specification: TA

## Covariance Parameters

**Covariance Parameters Summary**

Covariance Parameters	Residual Effect	0
	Random Effects	5
Design Matrix Columns	Fixed Effects	5
	Random Effects	5 <sup>a</sup>
Common Subjects		11

Common subjects are based on the subject specifications for the residual and random effects and are used to chunk the data for better performance.

a. This is the number of columns per common subject.

**Residual Effect**

Residual Effect	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower	Upper
Variance	1,000	.	.	.	.	.

Covariance Structure: Scaled Identity

Subject Specification: (None)

**Random Effect**

Random Effect Covariance	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower	Upper
Var(Intercept)	1,985E-8 <sup>a</sup>	.	.	.	.	.
Var(OMAV)	3,024	12,147	,249	,803	,001	7944,861
Var(CRlog)	,961	2,054	,468	,640	,015	63,313
Var(LVP12)	2,471E-13 <sup>a</sup>	.	.	.	.	.
Var(ROA)	79,282	82,390	,962	,336	10,342	607,774

Covariance Structure: Variance components

Subject Specification: TA

a. This parameter is redundant.

**LIITE 2 TUTKIELMAN AINEISTON  
KONKURSSIYRITYKSET**

Y-tunnus	Yritys	Konkurssin päivämäärä	Tilinpäätöksen päivämäärä	Toimiala
2383801-7	4Vision Oy	12.12.2013	30.6.2013	47719
1729249-2	AJM-Rakennus Oy	15.7.2013	31.10.2012	41200
2161218-8	Alhon Kala Oy	30.7.2013	31.3.2012	10200
0720488-2	Arelmek Oy	23.5.2012	31.12.2010	71125
1793907-3	Arin Ajolinjat Oy	15.12.2014	31.3.2014	49410
1865710-7	ATK-tarvike Solar Comp Oy	29.4.2013	30.6.2012	46699
2050094-2	Aurelia Trading Oy	1.8.2012	31.12.2011	46742
0660441-1	Aurinkosuoja Oy Auri	12.9.2012	31.12.2011	13921
0936085-4	Auterit Oy	16.10.2013	30.11.2012	31090
0656630-4	Autobest Oy	17.4.2012	30.6.2010	45112
2252438-2	Betoprof Oy	29.7.2013	31.12.2012	41200
0815852-3	Binderman Oy	26.3.2013	31.12.2011	18140
2236007-4	Bothnia Carpenters Ab	8.10.2013	31.12.2012	41200
0179003-2	Bröderna Brandt Läderfabrik Ab	23.1.2013	29.2.2012	15110
1044613-7	Citypro Oy	20.5.2014	31.12.2011	46521
1962945-8	CR Suunnittelu Oy	8.11.2013	30.4.2013	71123
2320709-8	Cube Elements Oy	7.12.2012	31.12.2011	41200
2423435-4	Design R & W Oy	31.10.2013	31.12.2012	46650
2226442-9	DigiPatrol Oy	2.11.2012	31.12.2011	43210
1765448-1	Dorripal Oy	27.11.2012	31.10.2011	47764
1809280-4	Dox-Bella Oy	14.8.2013	31.12.2012	41200
2306624-8	Dox-Neljä Oy	21.8.2013	31.12.2012	73111
2189846-2	DuaBros Oy	9.7.2012	31.12.2010	71122
2221767-3	E4 Corporation Europe Oy	12.7.2013	31.12.2011	45201
0878204-3	Ehycon Oy	20.4.2012	30.6.2011	43342
2257919-9	EJJ-Metsätyö Oy	24.4.2012	30.4.2011	2200
2441372-3	Enonkosken Huoltokeskus Oy	21.10.2013	31.12.2012	45201
1048347-2	Ensemble Communications Oy	12.3.2013	30.6.2012	70210
1522650-4	Euro-Pet Trading Oy	30.8.2012	30.6.2011	47764
1014422-7	FinSelena Oy	27.12.2013	31.12.2012	46909
0988644-2	Formia Lakeus Oy	22.11.2012	31.12.2011	25620
2233291-1	Gemini Viihde Oy	8.8.2013	31.12.2012	56101
2144233-4	Gnet Oy	12.3.2013	31.12.2011	47540
2366711-5	Golden Vision Oy	24.10.2012	31.12.2011	41200
2096254-9	H & H Sisustus Oy	7.6.2012	31.12.2011	47523
0634365-3	Hakamet Oy	30.4.2013	31.12.2011	25990
0805579-4	Hakkarin Tasoite ja Maalaus Oy	24.9.2013	28.2.2013	43341
2002440-5	HANNET OY	3.4.2014	30.6.2013	42220
0858349-0	Hannu Pirttilä Oy	7.11.2012	31.12.2011	46741
1068874-2	Har-Kon Konepaja Oy	22.11.2012	30.6.2012	28490
0496578-2	Healthcos Oy	30.12.2013	31.12.2012	96022
2034363-3	Hekometal Oy	19.12.2013	28.2.2013	25110

Y-tunnus	Yritys	Konkurssin päivämäärä	Tilinpäätöksen päivämäärä	Toimiala
2267686-1	Helsingin telineasennus Oy	10.2.2014	31.12.2012	43999
1765497-5	Hiusateljè Tuominen Oy	6.6.2013	30.4.2012	96021
1739737-4	HLR Energia Oy	15.4.2013	31.12.2011	25290
2152634-3	Hotelli-Ravintola Pitkä-Jussi Oy	10.1.2014	31.3.2012	56101
2330431-4	Huolto Ykkönen Oy	2.6.2014	31.12.2011	81100
2276425-5	Hyvinkään SMS-Rakennuspalvelu Oy	7.5.2013	30.9.2012	41200
0158863-5	Imatran LV-Asennus Oy	1.3.2013	30.6.2012	43220
0306369-9	J. Yli-Peltola Oy	2.1.2012	30.4.2011	8920
2190635-4	JMV YHTIÖT OY	9.7.2012	31.1.2012	41200
2270850-1	Jooperi Oy	9.5.2014	31.5.2013	55000
2156175-5	JoutCon Oy	27.12.2013	31.12.2012	17120
0678208-2	Kaarinan Tasopaino Oy	17.8.2012	31.12.2011	18120
0414971-8	Kakkostien Rakennus ja Saneeraus Oy	16.1.2014	31.10.2012	41200
0197513-3	Karin Metall Oy	9.4.2013	31.10.2012	27510
1637085-8	Keski-Suomen TNT-Rakennus Oy	28.2.2013	31.12.2011	41200
2285950-2	Keuruun Puukomponentti Oy	8.5.2012	31.12.2010	16212
2190348-7	KiinteistöMasiina Oy	10.5.2013	30.4.2012	81100
0877697-0	Kirnauskis 2.0 Oy	21.10.2013	29.2.2012	73111
1071134-0	Kitee Ski Oy	30.5.2013	31.7.2012	32300
1013294-7	Kiteen Kotipesä Oy	13.3.2012	28.2.2011	47113
1509784-5	Koivutapio Oy	6.2.2013	31.12.2011	46732
2178600-6	Konehuolto Kriikkula Oy	25.7.2012	30.4.2011	33122
2340811-3	Kuljetus A&S Ojala Oy	31.10.2013	31.12.2011	49410
2033387-3	Kuljetus Ojala & Toivoniemi Oy	17.12.2012	31.12.2011	49410
1072646-2	Kuopion Asemankukka Oy	1.10.2014	31.12.2013	47761
0474052-4	Kuopion Kuormaus- ja Nosto Oy	31.7.2013	31.12.2011	43991
0998327-7	Kuormaus ja kuljetus Ruuskanen Oy	26.6.2014	30.9.2012	49410
2360006-8	Kympin Rakentajat Oy	29.10.2012	31.12.2011	16231
2309367-6	LDR Studio Oy	22.5.2012	31.12.2010	73111
1570519-3	Likinix Oy	11.1.2013	31.12.2011	49410
0902794-7	Lomataika Oy	6.11.2012	31.12.2011	68100
1900299-7	Luovaliike Oy	24.11.2014	31.12.2011	62020
0429855-3	LVI-Nykänen Oy	12.9.2013	31.12.2012	43220
1465730-2	LVI-Savo Mara Oy	14.2.2013	31.5.2012	43220
0458072-7	Macpine Oy	17.4.2013	31.12.2011	16239
0857102-7	Mainostoimisto Linnunrata Oy	30.10.2012	31.12.2011	73111
0953756-8	Mantereen Kenkätehdas Oy	25.10.2012	31.12.2011	15200
2096452-8	Masami Oy	27.6.2013	31.12.2011	71124
2394297-3	Megamuksu Oy	31.10.2013	31.3.2012	47784
1848724-0	Mopsin Kiinteistöpalvelu Oy	26.3.2014	31.12.2012	41200
2072788-0	MP-Ravintolat Oy	22.1.2014	31.12.2011	56101
0683805-8	Muovirinne Oy	22.8.2013	31.1.2012	20160
2415680-0	Nelax Oy	15.8.2014	30.6.2013	41200
1980684-8	Nordiska Tjänster Ab	24.2.2014	31.12.2012	70220
2221551-2	NRDXline OY	28.12.2012	31.12.2011	43999

Y-tunnus	Yritys	Konkurssin päivämäärä	Tilinpäätöksen päivämäärä	Toimiala
0505507-8	Off Road Center Oy	16.12.2014	30.6.2012	45112
1043824-1	Outokummun Puu Oy	17.1.2012	30.6.2011	46732
0832450-1	Oy Formtec Ab	3.9.2013	31.3.2012	18120
2368550-8	Oy HB Consult & Rakennus Ab	26.3.2014	31.12.2012	41200
1841478-9	Oy Substpharma Ltd	7.5.2013	31.12.2011	46461
0988207-0	Oy Team Cards Ltd.	14.6.2013	31.12.2012	47990
2111788-1	Oy Terter Ab	4.9.2012	31.12.2011	86220
2096824-6	Peruskalliokodit Oy	4.4.2013	31.12.2011	87901
0621538-7	Pesutekno Oy	9.9.2013	30.9.2012	28990
2333479-1	PlanData Oy	30.10.2012	31.12.2011	46510
0787091-4	P-O Trading Oy	18.8.2014	28.2.2013	47750
2366438-7	PolmaRent oy	7.12.2012	31.12.2011	78200
2011129-3	POPOT Finland Oy	22.11.2013	30.6.2013	47721
0964819-6	Porvoon hitsauspalvelu Oy	28.10.2013	31.12.2012	25610
1705506-7	Prebend Oy	6.3.2013	31.5.2012	47641
2163989-0	Prof. Helsinki Oy	12.11.2012	31.12.2011	96021
2304459-1	PS. Kakkonen Oy	19.2.2014	31.12.2012	31090
0765393-6	Pukero Engineering Oy	13.6.2012	31.12.2011	71127
2248699-6	PunkaLive Oy	26.10.2012	31.12.2011	16239
2128175-3	Päijät-Hämeen Viemäri-Sako Oy	14.10.2014	31.5.2013	37000
0775073-6	Rainon Rakennus Oy	25.3.2013	31.12.2011	41200
2233522-3	Rakennus ESKI Oy	8.7.2013	31.12.2012	41200
2131148-1	Rakennusliike Kotisirakentaja Oy	10.1.2014	31.12.2011	41200
2240774-1	Rakennusliike Pasi Ikonen Oy	10.10.2011	31.12.2010	41200
0564814-8	Rakennusliike Seppo Simolin Oy	13.1.2014	28.2.2013	41200
2111486-7	Rakennuspalvelut Mauno Kuosa Oy	14.11.2013	31.3.2013	41200
2151864-2	Raksite Oy	20.9.2012	31.12.2011	41200
0924940-0	Rauta-Ekos Oy	3.4.2014	31.3.2013	47521
1703389-6	RKL Invest Ltd Oy	30.10.2013	31.3.2013	41200
2194873-0	Rommaloota Oy	25.7.2013	30.4.2012	47785
1033453-0	Rooftek Oy	28.5.2013	30.4.2012	46189
2271089-8	Safety Glass Experts International Oy Ltd	6.6.2014	31.12.2013	74909
1527900-8	SC Lapinkorpi Oy	13.11.2012	31.3.2012	43999
2293667-8	SL Intertrading Oy	15.4.2013	31.12.2011	41200
0740967-7	Soklex Oy	16.8.2013	31.12.2011	22230
2424774-5	SP Timantti Oy	30.7.2014	31.12.2013	41200
1074179-3	St Michel Serving Partners Oy	5.12.2012	31.12.2011	56101
1062790-7	Steamtec Oy	15.4.2013	31.12.2011	71127
2187043-5	Suomen hyvinvointikeskukset Oy	19.8.2013	31.12.2012	93190
1938076-0	Suomen Metsäpalvelu Oy	20.12.2012	31.12.2011	2400
2416279-2	Suomen Pikaruokalot Oy	11.2.2014	30.9.2012	10510
2175954-2	Suomen Ykkösrakenne Oy	27.2.2013	31.5.2012	41200
0612586-7	Toivalan Konepaja Oy	10.2.2014	30.6.2013	25110
2230047-7	TTP-Yhtiöt Oy	2.5.2012	31.12.2010	25110
1008350-8	Tulimeri Meritus Oy	13.12.2013	31.12.2012	46412

Y-tunnus	Yritys	Konkurssin päivämäärä	Tilinpäätöksen päivämäärä	Toimiala
0401005-4	Turengin Arvotyö Oy	5.3.2013	31.12.2011	41200
1963079-0	Turun Onnelat Oy	7.11.2012	31.12.2011	56101
0782256-8	Tuusniemen Asemahuolto Oy	16.5.2014	31.12.2012	47301
1972504-4	Tähtipolku Oy	30.6.2014	30.11.2013	93210
0983390-6	Vake Oy	10.9.2013	31.12.2012	27400
2343633-3	Vehnasmäen TTB Oy	5.6.2012	31.12.2010	96011
2497128-2	Veljekset Puurula Oy	16.10.2014	31.12.2013	52219
0693919-0	Vihdin Kaivutyö Oy	3.7.2014	31.1.2014	43120
2183561-3	VK-Electric Oy	21.3.2013	31.12.2011	43210
0966017-9	Vuoksitrans Oy	21.6.2011	30.4.2010	49410
1843820-3	Vuoleri Oy	8.1.2013	31.12.2011	23700

### LIITE 3 TUTKIELMAN AINEISTON TOIMIVAT YRITYKSET

Y-tunnus	Yritys	Tilinpäätöksen päivämäärä	Toimiala
0983394-9	Aaro Laasala Oy	31.5.2011	46732
0181902-6	Ab Holmängs Trävaruhandel Oy	31.12.2011	16231
0865636-2	Adconsys Oy	31.3.2011	43210
2192580-3	Ainoa Helsinki Oy	31.1.2011	73111
0747563-1	Aito Hand Held Oy	31.12.2011	46510
0751925-1	Alpensi Oy	31.12.2011	25290
2482804-8	Annival Interior Oy	31.12.2012	47785
0835796-0	APX-Metalli Oy	28.2.2013	25110
1997991-7	Asmex Oy	31.12.2011	41200
2284369-0	Aste Kirjat Oy	31.12.2011	18140
0913819-6	Bellingham Oy	31.12.2013	93210
2240024-0	Bilot Concept Oy	31.12.2011	62020
0107069-2	Blomqvist Oy	28.2.2011	49410
0749793-7	Chemret Oy	31.12.2012	20160
2077990-4	Decem Pharma Oy	30.11.2011	46461
1746085-5	Decokit Oy	31.1.2012	31090
0978588-7	Defisan Hygienia Oy	31.3.2012	27510
0705098-3	Egerstan Oy	31.10.2010	25110
2010183-7	Eläintarvike Murren Murkina Oy	31.12.2011	47764
1858176-3	E-P:n Hinauspalvelu Oy	31.12.2013	52219
1075580-2	Ep:n Rakennustyö Oy	30.6.2011	41200
2065210-2	Etelän Autohuolto Oy	31.12.2011	45201
0992554-4	Finnserviett Oy	30.6.2012	17120
1463431-1	Firmonte Rakenne Oy	31.12.2012	41200
0723907-3	Focusplan Oy	31.12.2011	71127
0144729-3	Föglöbutiken Ab	31.12.2011	47113
1805294-1	Gigacom Oy	31.3.2011	46521
2217534-6	Granström-Yhtiöt Oy Ab	31.12.2012	46189
2305212-9	H & P Rakentajat Oy	31.5.2013	41200
1612566-8	Hair House Seinäjoki Oy	31.12.2012	96021
1680033-6	Hairfix Helsinki Oy	31.7.2011	96021
1703263-1	Harmonia-Kaluste Oy	31.7.2012	31090
2408685-7	Hyvän Olon Puoti Vuokatti Oy	31.12.2012	96022
2129658-4	Indust Oy	31.3.2011	43991
1987188-5	Insinööri-toimisto Laaturakenne Oy	30.9.2013	71123
1104059-3	Itä-Suomen Ilmamestarit Oy	31.1.2012	43220
1085372-8	Jakeluasema Titta Home Oy	31.12.2012	47301
2097929-5	Jalan Oy	28.2.2013	47721
2345879-4	JJ Timber Oy	31.12.2011	2400
1726716-6	JMH Group Oy	31.12.2011	68100
2384829-0	Joensuun Erikoissaneeraus Oy	31.12.2013	41200
1035803-0	JR-Office Oy	31.10.2012	46650

Y-tunnus	Yritys	Tilinpäätöksen päivämäärä	Toimiala
2256141-5	Kaihdi-Lehtonen Oy	31.12.2011	13921
2228483-3	Kakk Rakennus Oy	31.3.2012	41200
0138189-6	Kallion Konepaja Oy	30.6.2012	28490
1704569-4	Karkkilan Lava- ja Teräsrakenne Oy	31.12.2013	25110
1646131-1	Karpalomedia Oy	30.6.2010	73111
2018132-0	Kenttäviiva Oy Automaatio	31.5.2011	71127
0110050-5	Kitkaliimaus Oy	31.12.2012	46699
0943019-7	Kivilinja Oy	31.3.2011	23700
0712176-1	Kolarimaalaus Oy Ollinen	30.6.2012	45201
2059946-0	Koneurakointi A. J. Hyvönen Oy	31.1.2011	8920
0618781-4	Kovalainen Oy	31.1.2011	46741
2281244-8	KRP Oy Koiviston Rakennuspalvelu	31.8.2012	43999
0347622-3	Kuivaniemen Kala Oy	31.12.2012	10200
1895827-6	Kukkamyynänti Piilola Oy	30.4.2013	47761
2189488-3	Kuljetus Filipp Oy	31.12.2013	43120
1855206-5	Kuljetus Jukka Matokangas Oy	31.5.2011	49410
2471567-6	Kuljetusliike Jokijako Oy	31.10.2014	49410
0915703-9	Kuljetusliike Jukka Ranta Oy	31.12.2010	49410
1919389-3	Kuljetuspalvelu Vatiala Oy	31.12.2012	49410
0512663-0	Kurkimäen Huolto Oy	31.8.2012	81100
1540060-5	Lahtihela Oy	31.12.2011	47523
1943205-7	Lakeuden Motohuolto Oy	31.12.2011	33122
1759757-9	Lehtohiltu Oy	31.12.2013	37000
0727057-2	Lemmikkieläinliike Halikatti Oy	31.12.2011	47764
0133533-9	Lenkki Oy	31.12.2011	15200
0930076-3	Levin Lomamökkien Huoltopalvelut Oy	31.5.2011	81100
0892013-5	Limpus-Rakennus Oy	31.5.2013	41200
0896362-9	Lohjan Laatalot Oy	31.12.2010	41200
2127060-8	Loventa Oy	31.12.2011	87901
2130077-4	Lumi Accessories Oy	31.12.2013	47719
0351915-8	LVI-Asennus P. Launo Oy	30.11.2012	43220
0111087-7	Lämpötekkinen toimisto Calor Oy	31.8.2011	71124
1037472-7	Mainemylly Suhina Oy	30.4.2012	70210
0715053-8	Manhattan Steak House Oy	31.12.2012	56101
2164222-1	Master Dressage Oy	31.12.2012	93190
2101408-1	Mericon Oy	30.4.2010	71122
2348115-0	Meridian Trans Oy	31.12.2011	49410
0988306-5	Namires Oy	31.12.2012	18120
0161550-9	Neonteos Oy	31.12.2012	27400
1592977-9	Nomade Collection Ab	31.12.2012	32300
2140521-8	Nora-hus Ab	30.9.2011	41200
2140521-8	Nora-hus Ab (2. kerta, mutta eri tilinpäätös)	30.9.2012	41200
0549741-5	Nummentien Autopiste Arvola Oy	29.2.2012	45112
0630513-9	OP-Teräs Oy	30.6.2011	25990
0633010-7	Oy Geson Ab	31.12.2012	15110

Y-tunnus	Yritys	Tilinpäätöksen päivämäärä	Toimiala
0850269-7	P.K.T. Pintakäsittelytekniikka Oy	31.12.2012	41200
0731609-1	Paino-Arra Oy	30.4.2011	18120
0913965-8	Pelismo Oy	31.12.2012	73111
0892215-7	Pelti- ja Rautatyö E. Salminen Oy	30.6.2012	43999
0758639-0	Pesutimangi Oy	30.9.2010	96011
0935830-8	Picote Oy Ltd	31.12.2011	41200
1845669-8	Pika-Lasi Oy	30.4.2011	43342
0861493-9	Pinnoitus Helin Oy	31.12.2012	25610
1503601-4	PJ-Center Oy	31.3.2012	41200
0831577-9	Propipe Oy	31.12.2011	22230
2274306-7	Pub Amuri Oy	31.7.2011	56101
2216913-0	Punavuoren peikko Oy	31.12.2012	47784
1450069-8	Putkikuningas Oy	31.12.2012	43220
0922794-9	Puumalan Talotarvike Oy	30.6.2011	41200
0760555-5	Puusepäntiike Pentti Muukka Oy	31.12.2011	16239
1856902-7	Rakennus JO-RI Oy	30.9.2011	41200
0810137-5	Rakennus Kurre Oy	31.3.2011	41200
1024805-0	Rakennusliike E Kylä-Kaila Oy	31.1.2012	41200
1983805-7	Rakennusliike M. Ranta Oy	30.9.2011	41200
0291016-3	Rakennustarvikeliike Laukkonen Oy	30.9.2013	47521
0918497-7	Rakennustoimisto Sohlman Oy	31.12.2013	41200
1956935-2	Ranic Oy	31.12.2012	41200
2396203-4	Rantakallion Juustola Oy	31.12.2012	10510
0138648-6	Rauman Erikoislääkärikeskus Oy	31.12.2011	86220
1045296-7	Rauman Juku Oy	31.12.2013	42220
1031956-8	Rauno Rauvala Oy	30.6.2011	2200
0522591-1	Ravintola Peking Oy	30.6.2011	56101
1071026-8	Ravintola Santamäki Turku Oy	31.12.2012	56101
0956576-1	Restahenkilöstö Oy	30.9.2011	78200
2497724-6	Roomstar Hotels Oy	31.12.2013	55000
1532967-7	RST Fasad Oy	31.3.2013	41200
0649066-1	R-Tech Suspension Oy	30.9.2012	47641
2029340-2	RTT-Team Oy	31.12.2010	71125
0133504-8	Satakunnan Autotalo Oy	31.12.2010	45112
0460326-3	Savo-Set Oy	31.3.2012	46909
1080045-8	Seinäjoki Power Welding Oy Ltd	30.4.2011	25620
0949319-8	Sisustus Romppanen Oy	31.12.2012	41200
0549989-1	Starlec Oy	31.12.2011	47540
0807194-6	Suomen Asbestiteknikka Oy	31.8.2011	43999
0824278-4	Suomen Ilmastointitukku Oy	31.12.2011	46742
0623776-2	Suomen Kerroslevy Oy	31.12.2010	16211
0987880-5	Suomen Tervaleppä Oy	31.12.2011	46732
2167067-9	Systeri Oy	31.3.2012	47990
0715955-9	Sähköasennus M. Yli-Mannila Oy	28.2.2011	43210
2331581-4	Thalgo Finland Oy	31.12.2013	47750

Y-tunnus	Yritys	Tilinpäätöksen päivämäärä	Toimiala
2284825-6	Toimitila-Rakentajat EJV Oy	31.12.2012	41200
2228022-6	Touchpoint Oy	31.3.2012	46412
1510935-9	Two Star Oy	31.12.2011	56101
0183265-0	Vaasan Maalaustyö Reinisalo Oy	31.12.2013	43341
1002209-4	VA-Palvelu Oy	31.1.2012	41200
2138364-5	WELLSEC OY	28.2.2013	74909
2126100-7	Verona International Oy	30.6.2012	70220
1973872-6	Vesanrak Oy	31.5.2012	41200
0181009-9	Wik & Drifell Invest Ab	31.12.2011	16239
0920413-0	Virjosvuo Yhtiöt Oy	31.1.2012	28990
2335003-8	Wonder Production Oy	31.12.2012	73111