

**YRITYSTEN KONKURSSITODENNÄKÖISYYDEN
ENNUSTAMINEN LOGISTISELLA
REGRESSIOMALLILLA**

**Jyväskylän yliopisto
Matemaattis-luonnontieteellinen tiedekunta**

Pro gradu -tutkielma

2018

**Tekijä: Ville Lehmusvirpi
Oppiaine: Tilastotiede
Ohjaaja: Juha Karvanen**



JYVÄSKYLÄN YLIOPISTO

TIIVISTELMÄ

Tekijä Ville Lehmusvirpi	
Työn nimi Yritysten konkurssitodennäköisyyden ennustaminen logistisella regressiomallilla	
Oppiaine Tilastotiede	Työn laji Pro gradu -tutkielma
Aika (pvm.) Tammikuu 2018	Sivumäärä 32
Tiivistelmä - Abstract	
<p>Luotettavan ennustemallin rakentaminen on yksi konkurssitutkimuksen keskeisin tavoite. Tällainen malli voi antaa tietoa mahdollisesta lähestyvistä konkurssitapahtumasta jo paljon ennen laskennallista konkurssihetkeä.</p> <p>Tämän tutkimuksen tarkoituksena on löytää mahdollisimman kompakti mutta silti hyvin kuvaava malli yritysten konkurssitodennäköisyydelle. Selittävien muuttujien määrän minimoimisella pyritään välttämään mahdollinen yliparametrisointi, jottei malli tule riippuvaiseksi mallinnusaineistosta ja ei näin ollen olisi yleistettävissä muille aineistoille.</p> <p>Mallin toimivuuden arviointiin kiinnitetään myös erityistä huomiota validoimalla se mallinnusaineiston ulkopuolisella testiaineistolla. Tämä on tärkeää, koska on havaittu, että suurimmassa osassa yrityksen konkurssitodennäköisyyttä käsittelevissä vertaisarvioituissa tutkimuksissa mallin validointia ei ole suoritettu testiaineistolla.</p> <p>Tulosten mukaan perinteiset tilinpäätöksen tunnusluvut, jotka kuvaavat yrityksen vakavaraisuutta, kannattavuutta ja maksuvalmiutta, toimivat hyvinä selittäjinä ennustettaessa yrityksen konkurssitodennäköisyyttä. Logistisen regressiomallin ennustekyky säilyy myös hyvänä, kun sitä sovelletaan testiaineistoon.</p>	
Asiasanat Logistinen regressioanalyysi, yritysten konkurssitodennäköisyys	
Säilytyspaikka Jyväskylän yliopisto/Matematiikan ja tilastotieteen laitos	

SISÄLLYS

1	JOHDANTO	1
1.1	Tutkimuksen taustaa	1
1.2	Tutkimuksen tavoitteet ja rakenne	3
2	AIKAISEMPIA TUTKIMUKSIA	4
3	KONKURSSITODENNÄKÖISYYDEN ENNUSTAMINEN.....	7
3.1	Binäärinen vaste	7
3.2	Logistinen regressio.....	7
3.3	Kerrointen merkitsevyyden testaaminen ja Waldin testi.....	8
3.4	Askeltava regressio mallin valinnassa	9
3.5	Aineisto.....	10
4	TUTKIMUKSEN TULOKSET	13
4.1	Yleistä.....	13
4.2	Konkurssitodennäköisyyden parametriestimaatit.....	14
4.3	Mallin testaus ulkopuolisella aineistolla	23
5	JOHTOPÄÄTÖKSET	26
	LÄHTEET	28
	LIITTEET	30

1 JOHDANTO

1.1 Tutkimuksen taustaa

Luotettavan ennustemallin rakentaminen on ollut konkurssitutkimuksen keskeisin tavoite. Tällainen malli voi antaa tietoa mahdollisesta lähestyvistä konkurssitapahtumasta jo paljon ennen laskennallista konkurssihetkeä. Varoitussignaali on arvokas tieto muun muassa yrityksen johdolle, omistajille sekä vieraan pääoman rahoittajille, jotta edellä mainitut voivat aloittaa saneerauksen ja yrityksen tervehdyttämisen. Rahoittajat voivat puolestaan miettiä hyvissä ajoin mahdollisia vaatimuksiaan sekä edellytyksiä jatkorahoitukselle. (Laitinen 1995, 63.)

Kuten edellisestä kappaleesta huomattiin, lainanhakijan luottokelpoisuus eli kyky hoitaa luottonsa on luotonantajan kannalta keskeistä. Tämän vuoksi erilaisia kvantitatiivisia ja kvalitatiivisia taustamuuttujia käytetään apuna, kun yritysasiakkaan todennäköisyyttä päätyä konkurssiin arvioidaan. Vaikka puhtaasti matemaattisten menetelmien käyttö on lisääntynyt jatkuvasti luottokelpoisuuden arvioinnissa, niin analyytikon kokemukseen ja arvioon perustuvat luottoluokitukset ovat edelleen melko yleisiä. (Bluhm, Overbeck, Wagner 2003, 18.)

Luottokelpoisuuden tilastomatemaattinen mallintaminen "Credit Scoring" kehittyi Yhdysvalloissa jo 1950- ja 1960-lukujen taitteessa ja levisi sieltä pääasiallisesti Eurooppaan ja Japaniin. Suomessa ensimmäiset asiakkaita koskevat luottoluokittelumallit otettiin käyttöön 1980-luvulla. Nykyään luottoriskimallinnuksessa erotetaan kaksi päätyyppiä eli varsinainen Credit Scoring, joka tarkoittaa luottoriskin arvioimista hakemushetkellä. Toinen päätyyppi on käyttäytymismallinnus "Behavior Scoring", jolla pyritään ennustamaan luoton saaneen asiakkaan käyttäytymistä. (Djerf 1995, 12-13.) Tämän tutkimuksen näkökulma on ensiksi mainittu eli Credit Scoring.

Luottoriskimallinnuksen avulla voidaan saavuttaa monenlaisia hyötyjä. Ensinnäkin luotonhakijan lainahakemuksen käsittely nopeutuu huomattavasti, koska hänen tietonsa voidaan syöttää suoraan tietokoneelle, joka laskee luottokelpoisuuden. Toiseksi lainahakemusten käsittely yhdenmukaistuu, koska luotonkäsittelijän henkilökohtaisen arvioinnin vaikutus vähenee merkittävästi. (Djerf 1995, 26.)

Luotonantajan kannalta etuja on vielä huomattavasti enemmän. Ensinnäkin kuten jo edellä todettiin, luottopäätöksen teko tehostuu, koska kaikki olennaiset taustamuuttujat on jo huomioitu ennustemallissa. Tämä tarkoittaa myös sitä, että luotonmyöntämisestä aiheutuvat kustannukset sekä luottotappiot pienenevät. Myös markkinointi tehostuu, koska ennustemallin avulla lainatarjoukset voidaan kohdistaa suoraan asiakkaisiin, joiden luottoriski on arvioitu pieneksi. Ennustemallit auttavat myös luoton hinnoittelussa, sillä yrityksille, joiden luottoriski on pienempi, voidaan lainaa myöntää edullisemmin. Myös luoton määrää voidaan suhteuttaa mallin antamaan luottoluokitukseen. (Pulkinen 1995, 44–45.)

Kuten lainanhakijan tapauksessa myös luotonmyöntäjä hyötyy siitä, että luotonmyöntöprosessi yhdenmukaistuu organisaatiossa. Tämä on erityisen tärkeää, sillä eri luotonkäsittelijät arvioivat lainahakemuksia oman subjektiivisen näkemyksensä mukaan, jolloin arviot samasta lainahakemuksesta saattavat poiketa käsittelijästä toiseen. Tästä huolimatta malli on kuitenkin apuväline, joka ei poista täysin subjektiivisen harkinnan käyttämistä päätöksenteossa. Harkinta on joissain tapauksissa välttämätöntä, sillä aineisto ei ikinä sisällä kaikkia reaali maailman tapahtumia ja näin ollen yksinomaan mallin avulla voidaan päätyä virheelliseen arvioon käsiteltäessä harvinaislaatuista luotonhakijaa, jonka kaltaisia muuttujatietoja ei ole ollut aineistossa käytettävissä. Tämä korostaakin aineiston huolellista tarkastelua sekä mallin validointia testiaineiston avulla ennen sen käyttöönottoa. (Djerf 1995, 26–27.)

1.2 Tutkimuksen tavoitteet ja rakenne

Tämän tutkimuksen tavoitteena on mallintaa yritysten konkurssitodennäköisyyttä logistisella regressiomallilla. Mallinnus tehdään Suomen Asiakastieto Oy:ltä saadulla yritysaineistolla. Haluan kiittää tässä vaiheessa Suomen Asiakastietoa mahdollisuudesta käyttää heidän aineistoaan.

Yrityksen konkurssitodennäköisyyttä ennustavia malleja on tehty paljon aiemminkin mutta kuten Aziz ja Dar (2006) toteavat laajassa vertailussaan, vain noin 46 prosentissa vertaisarvioituista tutkimuksista on käytetty erillistä aineistoa, jolla varmistetaan mallin toimivuus. Tämän vuoksi tässä tutkimuksessa pyritään huomioimaan mallin toiminta myös testiaineiston avulla.

Tutkimus on rakenteeltaan seuraavanlainen: Luvussa kaksi perehdytään aihetta käsitteleviin aiempiin tutkimuksiin. Seuraavassa luvussa käsitellään logistisen regressioanalyysin teoriaa sekä esitellään askeltavan regression käyttö mallin valinnassa. Luvussa neljä tarkastellaan yritysten konkurssitodennäköisyyden ennustamista mikroaineiston avulla sekä tulkitaan saadut tulokset. Viimeisessä luvussa tehdään yhteenveto tutkimuksen päätuloksista.

2 AIKAISEMPIA TUTKIMUKSIA

Logistinen regressioanalyysi ja erotteluanalyysi (Multiple Discriminant Analysis MDA) ovat olleet käytetyimmät menetelmät konkurssia mallinnettaessa. Aziz ja Dar (2006) kuitenkin huomauttavat, että niiden runsas käyttö ei välttämättä kerro niiden ennustetarkkuudesta. Heidän mukaansa yleisesti onkin saatu hieman tarkempia tuloksia käyttämällä AIES-malleja (AIES Artificial Intelligence Expert Systems). AIES-lähestymistavassa pyritään siihen, että annettujen esimerkkien avulla ohjelma oppii havaitsemaan muuttujien väliset riippuvuussuhteet suoraan aineistosta. Yleisesti ottaen AIES-malleja voidaan pitää tilastollisen mallinnuksen jatkeena, sillä tyypillisesti niissä hyödynnetään usean muuttujan tilastollisissa analyysissä käytettyjä tekniikoita. (Aziz, Dar 2006, 21, 26.)

AIES-lähestymistapa on kuitenkin vielä varsin uusi ja tehtyjen tutkimusten lukumäärä niin pieni, ettei lopullisia johtopäätöksiä sen paremmuudesta voida antaa. Sen sijaan esimerkiksi logistinen regressiomalli on johdonmukaisesti osoittanut useiden tutkimusten perusteella korkeaa ennustetarkkuutta sekä pieniä tyypin yksi ja kaksi virheitä. (Aziz, Dar 2006, 29.) Seuraavaksi tutustutaankin tarkemmin joihinkin logistisella regressiomallilla tehtyihin konkurssitutkimuksiin.

Ohlsonia (1980) edeltävissä tutkimuksissa yllä mainittu erotteluanalyysi (MDA) oli ollut tyypillisin tapa mallintaa konkursseja. Menetelmään liittyi kuitenkin rajoituksia, joista yhtenä esimerkkinä voidaan mainita, että se ei mahdollistanut dummy-muuttujien käyttöä. Käyttämällä logistista regressiomallia Ohlson (1980, 112) pystyi välttämään kaikki erotteluanalyysissä esiintyvät ongelmat.

Ohlsonin aineisto kattoi vuodet 1970–1976 ja otoksena olivat teollisuusyritykset, joiden osakkeilla käytiin kauppaa joko pörssissä tai OTC-markkinoilla. Selittävinä muuttujina tutkimuksessa olivat *yrityksen koko, velkojen ja taseen*

loppusumman suhde, käyttöpääoman ja taseen loppusumman suhde, lyhytaikaisen vieraan pääoman ja vaihtuvien vastaavien suhde, yrityksen velat ylittävät yrityksen varat (arvo=1, muulloin nolla), nettotuloksen ja taseen loppusumman suhde, liiketuloksen ja velkojen suhde, nettotulos negatiivinen kaksi viimeistä vuotta (arvo=1, muulloin nolla) sekä nettotuloksen muutos edellisestä vuodesta. (Ohlson 1980, 114, 118–119.)

Lopulliseen malliin nousi mukaan vain neljä muuttujaa, jotka osoittautuivat tilastollisesti merkitseviksi. Näitä olivat *yrityksen koko, velkojen ja taseen loppusumman suhde, nettotuloksen ja taseen loppusumman suhde* sekä *käyttöpääoman ja taseen loppusumman suhde*. Velkojen ja taseen loppusumman suhteen parametriestimaatti sai positiivisen arvon eli kyseisen muuttujan arvon kasvu lisäsi yrityksen konkurssiriskiä. Muiden muuttujien kertoimet saivat puolestaan negatiiviset arvot. Tulokset vaikuttavat odotetunlaisilta sillä on luonnollista, että esimerkiksi yrityksen likviditeetin (*käyttöpääoman ja taseen loppusumman suhde*) kasvaessa sen konkurssiriski pienenee. (Ohlson 1980, 119–121.)

Myös Zavgren (1985) käytti Ohlsonin tapaan logistista regressioanalyysiä mallintaessaan konkurssitapahtumaa. Hänen aineistonsa oli kuitenkin hyvin pieni kattaen 45 konkurssiyritystä ja 45 ei-konkurssiyritystä. Tutkimuksessa käytettiin selittävinä muuttujina ainoastaan tilinpäätöksestä saatuja tietoja vuosilta 1972–1978. (Zavgren 1985, 25.)

Zavgren mallinsi konkurssitapahtumaa viidessä eri aikaluokassa vuoden aikajännteillä jakaen mallit siten, että ensimmäisessä konkurssitapahtuma realisoiutui vuoden päästä. Mallissa numero kaksi konkurssi tapahtui kahden vuoden päästä ja viimeinen viides malli ennusti konkurssia viisi vuotta ennen sen realisoitumista. Tutkimuksessa mallien ennustekyky pääsääntöisesti kasvoi mitä lähemmäksi konkurssitapahtumaa menttiin mutta huomionarvoista oli, että tilinpäätösmuuttujista saatava informaatio oli merkitsevää jopa viisi vuotta ennen varsinaista konkurssitapahtumaa. (Zavgren 1985, 19, 28–29.)

Kuten Aziz ym. (2006) totesivat aiemmin, selkeää näyttöä eri konkurssimallien paremmuudesta toisiinsa nähden ei ole. Tästä huolimatta Laitinen ja Kankaanpää (1999) saivat mielenkiintoisen tuloksen vertaillaan kuutta suosituinta konkurssin ennustamismallia keskenään. Tutkimuksen mukaan yhden vuoden ennustehorisontilla logistisen regressiomallin ennustetarkkuus oli vertailuryhmän paras. (Laitinen, Kankaanpää 1999, 83.) Myös toinen konkurssimallien vertailuun keskittynyt tutkimus (Laitinen, Lehto 1995) antoi tuloksen, jonka mukaan logistinen regressiomalli toimi testiaineistolla validoitaessa paremmin kuin verrokkinsa. (Laitinen 1995, 88–90.)

Edellä käsitellyt tutkimukset osoittavat, että logistinen regressioanalyysi on toimiva ja yksi käytetyimmistä lähestymistavoista konkurssien ennustamisessa. Monissa aiemmissa konkurssitutkimuksissa ongelmaksi on muodostunut kuitenkin käytettävän aineiston pieni koko. Vertaisarvioituissa tutkimuksissa jopa 42 prosentilla otoskoko on pienempi kuin sata. (Aziz ym. 2006, 23.)

Edellä mainitun lisäksi myös mallinnusaineiston ulkopuolisen testiaineiston käyttö, jolla mallin toimivuus voitaisiin varmistaa, on vähäistä. Ainoastaan 46 prosenttia vertaisarvioituista tutkimuksista käytti testiaineistoa. (Aziz ym. 2006, 23.) Tämä tutkimus pyrkiikin kiinnittämään erityistä huomiota mainittuihin ongelmiin käyttämällä riittävän suurta aineistoa sekä validoimalla malli huolellisesti testiaineiston avulla.

3 KONKURSSITODENNÄKÖISYYDEN ENNUSTAMINEN

3.1 Binäärinen vaste

Tässä tutkimuksessa selitettävä muuttuja on binäärinen saaden arvon yksi ($Y = 1$) todennäköisyydellä π , kun yritys on konkurssissa ja arvon nolla ($Y = 0$) todennäköisyydellä $1 - \pi$, kun yritys ei ole konkurssissa. Yksinkertaisin yhteys odotusarvon $E(Y) = \pi$ ja lineaarisen ennustemallin välille saadaan yhtälöllä $\pi = x^t\beta$. Kyseisellä mallilla on kuitenkin useita puutteita, joista ongelmallisimpana voitaneen pitää sitä, että todennäköisyys π ei välttämättä rajoitu välille $0 < \pi < 1$. (Davison 2003, 487.)

Kuten edellä todettiin, niin lineaarisen todennäköisyysmallin ongelmista merkittävin liittyy sen vääränlaiseen funktiomuotoon. Yrityksen konkurssitodennäköisyyden sekä selittävien muuttujien välisen suhteen tulisi olla pikemminkin S-käyrän muotoinen ja mallin tulisi täyttää seuraavat vaatimukset:

$$(1) \quad \lim_{x^t\beta \rightarrow +\infty} \text{Prob}(y = 1) = 1$$

$$\lim_{x^t\beta \rightarrow -\infty} \text{Prob}(y = 1) = 0$$

Näin ollen sovelluksissa päädytään yleensä käyttämään joko probit- tai logit-mallia. (Greene 1997, 874–875.)

3.2 Logistinen regressio

Binäärisen vasteen tapauksessa tyypillisin valinta funktiomuodoksi on logistinen jakauma ja sitä on päädytty käyttämään myös tässä tutkimuksessa. Tällöin konkurssitodennäköisyydeksi saadaan

$$(2) \quad \Pr(Y = 1) = \pi = \frac{\exp(x^t \beta)}{1 + \exp(x^t \beta)}, \Pr(Y = 0) = 1 - \pi = \frac{1}{1 + \exp(x^t \beta)}.$$

Kaavassa kaksi esitetty logistinen regressiomalli voidaan esittää lineaarisessa muodossa logaritmisoidun vedonlyöntisuhteen (odds) avulla,

$$(3) \quad \frac{\Pr(Y=1)}{\Pr(Y=0)} = \frac{\pi}{1-\pi} = \exp(x^t \beta).$$

Uskottavuusfunktio riippumattomille binäärisille havainnoille y_1, \dots, y_n ja muuttujille x_1, \dots, x_n on muotoa

$$(4) \quad L(\beta) = \prod_{j=1}^n \left\{ \frac{\exp(x_j^T \beta)}{1 + \exp(x_j^T \beta)} \right\}^{y_j} \left\{ \frac{1}{1 + \exp(x_j^T \beta)} \right\}^{1-y_j} = \frac{\exp(\sum_j y_j x_j^T \beta)}{\prod_j \{1 + \exp(x_j^T \beta)\}}.$$

(Davison 2003, 490.)

3.3 Kerrointen merkitsevyyden testaaminen ja Waldin testi

Waldin testisuureen avulla voidaan määrittää mitkä estimoiduista kertoimista ovat merkitseviä. Kyseisen testisuureen arvo saadaan jakamalla suurimman uskottavuuden estimaatti $\hat{\beta}_i$ sen estimoidulla keskivirheellä \widehat{SE} . Waldin testisuure on muotoa

$$W_i = \frac{\hat{\beta}_i}{\widehat{SE}(\hat{\beta}_i)}.$$

Waldin testisuureen kaksisuuntainen p -arvo on $P(|z| > W) = p$, missä z on standardinormaalijakaumasta saatu satunnaismuuttuja. Joissain tapauksissa Waldin testi saattaa ehdottaa kertoimen hylkäämistä, vaikka kerroin olisikin tilastollisesti merkitsevä uskottavuusosamäärätestin perusteella. (Hosmer, Lemeshow ja Sturdivant 2013, 14-15.)

Uskottavuusosamäärätestissä (UO) estimoitujen kertoimien merkitsevyyttä tarkastellaan rajoitetun ja rajoittamattoman uskottavuusfunktion osamäärän avulla. Uskottavuusosamäärän testisuure on muotoa

$$UO = -2\log(l_0/l_1) = -2(L_0 - L_1),$$

jossa L_0 ja L_1 ovat logaritmisten uskottavuusfunktioiden maksimiarvoja. Suuret testisuuren arvot johtavat nollahypoteesin ($H_0 : \beta = \beta_0$) hylkäykseen. (Agresti 2013, 10–11.)

Käytännössä Waldin testi ja uskottavuusosamäärätesti antavat samansuuntaisia tuloksia. Mahdollisia ongelmia saattaa esiintyä lähinnä tilanteissa, joissa arvot ovat lähellä toisiaan toisen testisuureen antaessa p -arvon, joka on alle 0,05 ja toisen ollessa yli 0,05. Tällöin tulisi preferoida uskottavuusosamäärätestin tulosta. (Hosmer, Lemeshow ja Sturdivant 2013, 14–15.)

3.4 Askeltava regressio mallin valinnassa

Askeltava regressio oli joitain vuosia sitten erittäin suosittu metodi mallin valinnassa mutta nykyisin painopiste on siirtynyt enemmän kontrolloidun mallin valinnan -menetelmään (purposeful selection) (Hosmer ym. 2013, 89–125). Tästä huolimatta askeltava regressio on hyödyllinen apuväline sopivan mallin etsimisessä etenkin tämänkaltaisessa tutkimuksessa, jossa käytettäviä muuttujia on kohtuullisen paljon ja teoriasta ei voida suoraan johtaa kaikkia käytettäviä muuttujia. (Hosmer ym. 2013, 125.) Lopullinen mallin valinta tapahtuu kuitenkin aina tutkijan toimesta, jotta mallista tulisi myös teoreettisesti perusteltu (Djerf 1995, 22).

Tässä tutkimuksessa käytetään apuna alustavassa mallin valinnassa eteenpäin askeltavaa regressiota (Stepwise Selection). Askeltavien regressiomenetelmien tapauksissa valinta tyypillisesti tehdään eteenpäin askeltavan tai taaksepäin

askeltavan menetelmän välillä. Myös edellä mainittujen yhdistelmät ovat mahdollisia. Eteenpäin askeltavan algoritmin valintaa voidaan perustella ennustetarkkuudella, sillä sen käyttö tyypillisesti pienentää ennustevarianssia mutta voi samalla hieman lisätä harhaa. Yleisesti ottaen mallin ennustekyky saattaa kuitenkin parantua. (Hastie, Tibshirani ja Friedman 2009, 57–60.) Tämän tutkimuksen tarkoituksena on löytää mahdollisimman vähäparametrinen mutta silti hyvin ennustava malli, joten eteenpäin askeltava regressio antaa hyvän lähtökohdan muuttujien alustavalle valinnalle. Käydään seuraavaksi läpi eteenpäin askeltavan algoritmin heuristiikkaa.

Ensimmäiseksi tutkijan täytyy päättää mikä on se tilastollinen merkitsevyystaso P , jolla muuttujat valikoidaan malliin. Tämä valinta ratkaisee sen, kuinka paljon muuttujia sisällytetään malliin. Hosmer ym. (2013) suosittelevat, että sopiva valinta P :lle voisi olla välillä $[0,15, 0,20]$. Mallin ensimmäiseksi muuttujaksi x_1 valikoituu se, joka toimii parhaana yksittäisenä selittäjänä, kun kriteerinä on edellä mainittu tilastollinen merkittävyys. Seuraavassa askeleessa malliin otetaan mukaan muuttuja x_2 , joka toimii parhaana lisäselittäjänä muuttujan x_1 ollessa jo mallissa. Koska muuttujan x_2 lisääminen malliin saattaa vaikuttaa muuttujan x_1 merkitsevyyteen, niin seuraavassa askeleessa tarkastellaan tulisiko aiemmin lisätty muuttuja poistaa, mikäli sen p -arvo on heikentynyt jonkin kriittiseksi määritellyn raja-arvon yli. Tämän jälkeen kyseistä prosessia jatketaan samalla tavalla, kunnes jäljellä olevien muuttujien tilastollinen merkitsevyystaso ylittää tutkijan alussa määrittelemän tason ja eivät näin ollen voi tulla valituksi malliin. (Hosmer ym. 2013, 126–128.)

3.5 Aineisto

Tutkimuksessa käytettävä aineisto pohjautuu Suomen Asiakastieto Oy:ltä saatuun yritysaineistoon. Poiminta rajattiin siten, että mukaan otettiin yritykset, joiden liikevaihto on yli 100 000 euroa. Näin saatiin sopivan kokoinen aineisto, jossa konkurssiin joutuneiden yritysten suhteellinen osuus on riittävä

tilastolliseen mallinnukseen. Alkuperäinen aineisto oli jaettu kahteen osaan eli selittävien muuttujien aineistoon sekä vastemuuttujan aineistoon, joka sisältää yritysten konkurssitiedot. Selittävien muuttujien aineisto kattaa vuodet 2011–2013 (n=146497) ja siinä on mukana suuri joukko erilaisia tilinpäätösmuuttujia, jotka kuvaavat muun muassa yrityksen kannattavuutta ja maksuvalmiutta. Myös rahoituksen tunnuslukuja sekä yrityksen ikää hyödynnetään mallinnuksessa.

Konkurssiaineisto sisältää tiedot yrityksistä, jotka ovat päätyneet konkurssiin ja se kattaa vuodet 2012–2014 (n=1411). Konkurssiaineiston ajanjakso eroaa selittävien muuttujien aineistosta, koska tarkoituksena on tutkia millä todennäköisyydellä yritys joutuu konkurssiin seuraavan vuoden aikana. Tällöin esimerkiksi vuoden 2011 selittävien muuttujien aineistoon yhdistetään vuoden 2012 konkurssitieto ja sama menettely tehdään koko aineistolle.

Käytettävä aineisto on paneeliaineisto eli tämä mahdollistaisi poikkileikkausyksikön seuraamisen ajassa. Paneeliaineistomenetelmiä ei kuitenkaan käytetä tässä tutkimuksessa, koska mallinnuksessa käytettävän aineiston pituus on varsin lyhyt (2011–2013). Mallinnuksessa käytettävä aineisto lyheni vielä entisestään sillä se jaettiin siten, että varsinainen mallinnusaineisto käsittää ajanjakson 2011–2012 ja vuosi 2013 on jätetty mallin validointia varten.

Aineiston muodostaminen aloitettiin siten, että ensiksi muodostettiin vastemuuttuja ($y=1$), kun yritys on konkurssissa. Tämän jälkeen kyseinen tieto yhdistettiin selittävien muuttujien aineistoon ja näin saatiin varsinaisen mallinnusaineiston kooksi 99396 havaintoa. Aineistossa oli 1824 havaintoyksikköä, joilla oli puuttuvia havaintoja. Puuttuvat arvot olivat jakautuneet satunnaisesti havaintoyksikköjen kesken ja niitä oli suhteellisen pieni määrä (1,8 %), joten kyseisiä puuttuvia arvoja ei nähty tarpeelliseksi imputoida vaan ne päädyttiin jättämään pois lopullisesta aineistosta. Kyseinen

käsittely ei vaikuttanut juurikaan yritysten konkurssien suhteelliseen osuuteen vaan se pysyi lähes samana saaden arvon 0,97 % (0,98 %). Näin saadun lopullisen aineiston (2011–2012) havaintojen lukumäärä on 97572, josta yritysten konkurssien määrä on 948.

Validointiaineistona käytettiin vuoden 2013 havaintoja, joihin yhdistettiin vuoden 2014 konkurssitieto ylläkuvatun tapaisesti ja aineiston kooksi saatiin 47101 havaintoyksikköä. Myös validointiaineistossa oli puuttuvia arvoja (734), joissa ei ollut havaittavissa mitään systematiikkaa, joten ne poistettiin testiaineistosta. Lopullisen validointiaineiston havaintojen lukumäärä on 46367 ja konkurssien määrä 438. Aineiston käsittelyllä ei ollut tässäkään tapauksessa vaikutusta konkurssien suhteelliseen osuuteen vaan se pysyi samana (0,94 %).

Tutkimuksessa käytettiin aineiston muokkauksessa ja mallinnuksessa SAS-ohjelmistoa (versio 9.4). Mallinnuksessa käytetty proseduuri oli PROC LOGISTIC.

4 TUTKIMUKSEN TULOKSET

4.1 Yleistä

Tavoitteena on löytää vähäparametrinen mutta silti yritysten konkurssitodennäköisyyttä mahdollisimman hyvin kuvaava malli. Selittävien muuttujien määrän minimoimisella pyritään muodostamaan vakaa ja helposti implementoitava malli, sillä mitä enemmän muuttujia siihen lisätään, sitä riippuvaisemmaksi malli tulee kyseisestä aineistosta. Tämä yliparametrisointi saattaa johtaa myös suuriin estimoituihin kertoimiin sekä suuriin keskivirheisiin. (Hosmer ym. 2013, 90.)

Mallinnus aloitettiin käymällä läpi kaikkien muuttujien tilastollisia tunnuslukuja, jotta mahdolliset virheelliset havainnot voitaisiin löytää ja korjata (Liite 1). Aineistosta löytyikin havaintoja, joilla oli puuttuvia arvoja. Näitä oli kuitenkin suhteellisen pieni määrä, joten ne päädyttiin poistamaan aineistosta kuten kappaleessa 3.5 kerrottiin. Samalla tarkasteltiin korjauksen vaikutuksia maksuhäiriöiden suhteellisiin osuuksiin ja havaittiin, että toimenpiteellä ei ollut vaikutusta niihin.

Edellä mainitun lisäksi muuttujien väliset korrelaatiot tutkittiin huolellisesti (Liite 2). Suuret korrelaation arvot saattavat vaikuttaa estimoituihin kertoimiin, joten niiden tarkastelu on tärkeää. Liitteestä kaksi havaitaan, että ainakin muuttujien *Quick ratio* ja *Current ratio* välinen korrelaatio on suuri (0,85), joten todennäköisesti toinen muuttujista täytyy jättää mallista pois, mikäli ne osoittautuvat tilastollisesti merkittäviksi selittäjiksi. Myös *Käyttökateprosentin* ja *Liikevoittoprosentin* välinen korrelaatio on melko suuri (0,80) ja myös tämä täytyy ottaa huomioon mallinnuksessa edellä kuvatulla tavalla.

Aineistossa ei ollut senkaltaisia luokittelevia muuttujia, joiden avulla olisi ollut luontevaa muodostaa muuttujien välisiä interaktioita. Suurin osa aineistossa

olevista selittävistä muuttujista on tilinpäätöksellisiä tunnuslukuja, jotka on muodostettu jo useiden muiden muuttujien avulla kuten kappaleesta 4.2 käy ilmi. Näin ollen tällaisiin muuttujiin yhdysvaikutusten lisääminen ei olisi mielekäästä. Tätä seikkaa tukevat myös teoreettiset perusteet, sillä aiemmissa tutkimuksissa, joissa yritysten konkurssitodennäköisyyttä on mallinnettu tilinpäätöksestä saatavilla tiedoilla, ei interaktioita juurikaan näyttäisi esiintyvän.

4.2 Konkurssitodennäköisyyden parametriestimaatit

Taulukossa yksi on nähtävissä estimoidut regressiokertoimet mallille, jossa on mukana kaikki aineiston muuttujat. Tuloksista nähdään, että muuttujista erittäin merkitseviä ovat *Yhtiön ikä vuosissa*, *Kokonaispääoman tuotto prosentti*, *Oma varaisuusaste prosentti* sekä *Käyttöpääoma prosentti*. Kyseisten muuttujien estimoidut kertoimet saavat negatiiviset arvot eli niiden kasvaessa yrityksen konkurssiriski pienenee.

Oma varaisuusaste prosentti kohdalla negatiivinen kerroin on odotetunlainen, sillä kyseinen muuttuja mittaa yrityksen vakavaraisuutta, tappionsietokykyä sekä kykyä selviytyä sitoumuksistaan pitkällä aikavälillä. Myös *Kokonaispääoman tuotto prosentti* kasvun etumerkki vastaa intuitiota, koska se on yksi kannattavuuden tunnusluvuista, jolla kuvataan liiketoiminnan taloudellista tulosta ja on näin ollen yksi jatkuvan liiketoiminnan perusedellytyksistä. (Yritystutkimus ry 2011, 60–64, 66.)

Taulukko 1 Logistinen regressiomalli kaikille aineiston muuttujille

Muuttuja	Kerroin	Keskivirhe	Waldin testisuure	Tn>khiin neliö
Vakio	-5,3127	1,1327	21,9966	<,0001
Yhtiön ikä vuosissa	-0,0271	0,00349	60,5731	<,0001
Liikevaihto, eur	-1,67E-10	3,89E-09	0,0019	0,9657
Käyttökate-%	-0,00475	0,00218	4,7465	0,0294
Liikevoitto-%	0,00273	0,00203	1,8037	0,1793
Nettotulos-%	0,000381	0,000278	1,8764	0,1707
Sijoitetun pääoman tuotto-%	0,000097	0,000177	0,3018	0,5828
Kokonaispääoman tuotto-%	-0,0247	0,00111	492,4981	<,0001
Quick ratio	0,00219	0,00228	0,9265	0,3358
Current ratio	-0,0092	0,0035	6,9331	0,0085
Omavaraisuusaste-%	-0,00055	0,000061	80,754	<,0001
Net gearing	0,000223	0,000141	2,486	0,1149
Suht. velkaantuneisuus-%	-0,00009	0,000105	0,6729	0,4121
Käyttöpääoma-%	-0,00252	0,000446	31,873	<,0001
Taseen loppusumma, eur	-2,42E-08	9,18E-09	6,9455	0,0084
Yhtiömuoto (vertailuryhmä: muut)				
Osakeyhtiö	1,3085	1,1318	1,3367	0,2476
Julkinen osakeyhtiö	-9,352	216,6	0,0019	0,9656
Osuuskunta	0,1322	1,2911	0,0105	0,9185

Käyttöpääomaprocentti on saatu suhteuttamalla käyttöpääoma liikevaihtoon (12kk) ja estimoidun kertoimen mukaan sen kasvu pienentää yrityksen konkurssiriskiä. Käyttöpääoma mittaa yrityksen juoksevaan liiketoimintaan sitoutuneen lyhytaikaisen omaisuuden määrää ja se täytyy rahoittaa omalla pääomalla ja lyhytaikaisella sekä pitkäaikaisella vieraalla pääomalla. Pääomaa sitoutuu yritystoiminnassa, johtuen rahasiirtojen viiveistä suoritteeseen nähden, muun muassa vaihto-omaisuuteen ja myyntisaamisiin. Yrityksen on puolestaan mahdollista saada ostoveloilleen maksuaikaa sekä suorite käyttöönsä ennen maksutapahtumaa. Näin ollen käyttöpääoma kuvaa yritystoiminnasta aiheutuvaa käyttörahoituksen tarvetta ja myös pääoman käytön tehokkuutta. *Käyttöpääomaprocentissa* käyttöpääoma suhteutetaan liikevaihtoon sillä käyttöpääoman eri erät riippuvat liikevaihdon määrästä. (Tunnuslukuopas 2018.)

Yrityksen iän vaikutus konkurssiriskiin vaikuttaa myös loogiselta, sillä kansainväliset tilastot ovat osoittaneet, että etenkin yrityksen viisi ensimmäistä vuotta ovat erityisen riskialtista aikaa konkurssille. Tämä johtuu lähinnä siitä, että aloittava yritys tarvitsee alussa paljon vierasta pääomaa ja mikäli yritystä ei saada tarpeeksi nopeasti kannattavaksi kuten usein on, se ei selviydy maksuvelvoitteistaan. Näin ollen vastaperustetut yritykset kuuluvatkin erityiseen riskiryhmään konkurssiuhan suhteen. (Laitinen, Laitinen 2014, 32.)

Muita merkitseviä muuttujia taulukossa yksi ovat alle yhden prosentin merkitsevyydellä *Current ratio* sekä *Taseen loppusumma*. *Current ratio* mittaa yrityksen maksuvalmiutta eli kykyä selviytyä maksuistaan ajallaan ja sen kasvu pienentää tulosten mukaan yrityksen konkurssiriskiä. Myös *Quick ratio* on hyvin samankaltainen maksuvalmiuden tunnusluku kuin *Current ratiokin* mutta taulukosta yksi voidaan havaita, että sen etumerkki on positiivinen (Yritystutkimus ry 2011, 71). Tämä selittyy *Current ration* ja *Quick ration* välisellä erittäin voimakkaalla korrelaatiolla (0,85) (Liite 2), joka kääntää *Quick ration* funktionaalisen riippuvuuden väärinpäin. Kun malli ajetaan ilman tunnuslukua *Current ratio*, niin *Quick ration* etumerkki vaihtuu myös negatiiviseksi. *Taseen loppusumma* kuvaa yrityksen taloudellista asemaa tilinpäätöshetkellä ja myös sen kerroin saa negatiivisen arvon (Yritystutkimus ry 2011, 27).

Viimeisenä merkitsevänä muuttujana alle viiden prosentin merkitsevyydellä havaitaan *Käyttökateprosentti*. Se kertoo yrityksen liiketoiminnan tuloksen ennen poistoja ja rahoituseriä, ja tulosten mukaan sen kasvu pienentää konkurssiriskiä. (Yritystutkimus ry 2011, 61.)

Seuraavaksi ajettiin askeltava regressio taulukon yksi muuttujille. Havaitaan, että edelleen muuttujat *Yhtiön ikä* vuosissa, *Kokonaispääoman tuottoprosentti*, *Omavaraisuusaste* sekä *Käyttöpääomaprocentti* säilyvät tilastollisesti erittäin merkitsevinä muuttujina kuten aiemmassakin mallissa

Taulukko 2 Askeltava regressio (Stepwise)

Muuttuja	Kerroin	Keskivirhe	Waldin testisuure	Tn>khiin neliö
Vakio	-4,0466	0,0598	4577,45	<,0001
Yhtiön ikä vuosissa	-0,0293	0,00346	71,8963	<,0001
Käyttökate-%	-0,00167	0,000609	7,4922	0,0062
Kokonaispääoman tuotto-%	-0,0245	0,00102	579,4781	<,0001
Omavaraisuusaste-%	-0,00056	0,000061	83,4437	<,0001
Suht. velkaantuneisuus-%	-0,00027	0,000096	8,0777	0,0045
Käyttöpääoma-%	-0,00248	0,000405	37,41	<,0001

Myös *Käyttökateprosentti* säilyy mallissa merkitsevänä muuttujana ja sen merkitsevyys kasvaa hieman. Lisäksi *Suhteellinen velkaantuneisuusprosentti* valikoituu malliin mukaan ja se muuttuu tilastollisesti merkitseväksi yhdessä taulukon kaksi muiden muuttujien kanssa.

Korkea suhteellinen velkaantuneisuus edellyttää vastaavasti hyvää liiketulosta. Tämä saattaisi selittää sen miksi muuttujan etumerkki on negatiivinen eli velkaantuneisuus näyttäisi pienentävän konkurssiriskiä. Kyseisen tunnusluvun vertailu voidaan suorittaa kuitenkin vain saman toimialan yritysten kesken. (Yritystutkimus ry 2011, 67.) Tässä tutkimuksessa toimialajaottelua ei ole tehty ja muuttujan vaikutuksen tulkinta on muutoinkin epäselvä, joten kyseinen muuttuja päädytään jättämään pois lopullisesta mallista.

Edelliset mallit tukevat sitä, että lopulliseen malliin voidaan ottaa mukaan ainakin muuttujat, jotka olivat erittäin merkitseviä sekä askeltavassa regressiossa että kaikille muuttujille ajatus regressiossa. Näitä olivat siis

- *Yhtiön ikä vuosissa*
- $Kokonaispääoman tuotto - \% = \frac{Nettotulos+rahoituskulut+verot(12kk)}{Kokonaispääoma} \times 100$
- $Omavaraisuusaste - \% = \frac{Oikaistu oma pääoma}{Oikaistun taseen loppusumma-saadut ennakot} \times 100$
- $Käyttöpääoma - \% = \frac{Käyttöpääoma}{Liikevaihto(12kk)} \times 100$

Yllä esitettyihin mallin pohjana toimiviin neljään muuttujaan kokeiltiin seuraavaksi lisätä erilaisia muuttujayhdistelmiä, jotta löydettäisiin mallispesifikaatio, jonka ennustekyky ja myös teoreettinen perustelu olisi paras mahdollinen. Kokeilujen tuloksena päädyttiin lopullisessa mallin valinnassa taulukon kolme mukaiseen lopputulokseen.

Taulukko 3 Lopullinen mallispesifikaatio

Muuttuja	Kerroin	Keskivirhe	Waldin tesisuure	Tn>khiin neliö
Vakio	-4,0695	0,059	4754,5763	<,0001
Yhtiön ikä vuosissa	-0,0288	0,00345	69,7926	<,0001
Kokonaispääoman tuotto-%	-0,0249	0,001	619,2357	<,0001
Omavaraisuusaste-%	-0,00054	0,000059	82,087	<,0001
Käyttöpääoma-%	-0,00189	0,000396	22,8772	<,0001
Käyttökate-%	-0,00105	0,00053	3,9277	0,0475
Current ratio	-0,00761	0,00315	5,8477	0,0156

Malli poikkeaa stepwise-proseduurin tuottamasta malliversiosta siten, että *Suhteellinen velkaantuneisuusprosentti* päädyttiin jättämään pois, kuten jo aiemmin todettiin ja mukaan nostettiin yrityksen maksuvalmiutta kuvaava tunnusluku *Current ratio*. Taulukosta kolme voidaan havaita, että kaikki muuttujat ovat tilastollisesti merkitseviä vähintään viiden prosentin merkitsevyystasolla ja niiden kertoimet ovat negatiiviset eli muuttujien arvojen kasvu vähentää yrityksen konkurssiriskiä.

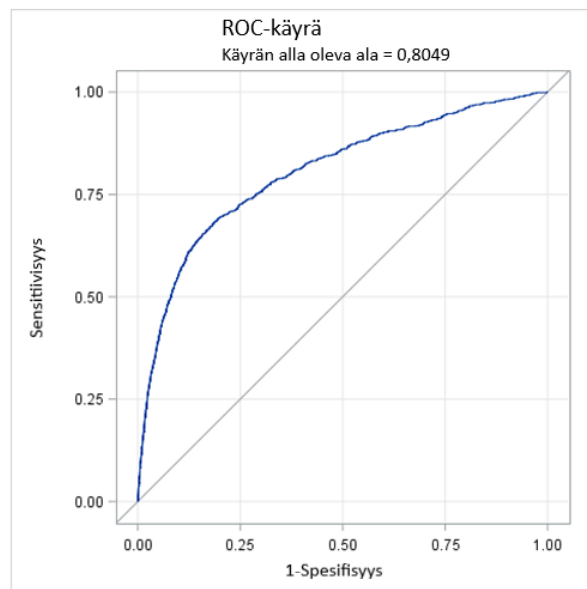
Tarkastellaan seuraavaksi valitun mallin ennustekykyä ROC-käyrän avulla (Reinikainen 2015, 17–20). ROC-käyrä on yleensä konkaavin muotoinen yhdistäen pisteet (0,0) ja (1,1). Käyrää tulkitaan siten, että mitä korkeampi se on eli mitä suurempi ala sen alapuolelle jää, sitä parempi on mallin ennustekyky. Tämän kanssa vastaava mitta on yhteensopivuusindeksi (concordance index). (Agresti 2013, 224.)

Yhteensopivuusindeksille on esitetty seuraavanlaisia yleisiä ohjearvoja (Hosmer ym. 2013, 177):

Jos	{	$c = 0,5$	<i>Mallilla ei ole ennustekykyä</i>
		$0,5 < c \leq 0,7$	<i>Mallin ennustekyky on huono</i>
		$0,7 \leq c < 0,8$	<i>Mallin ennustekyky on tyydyttävä</i>
		$0,8 \leq c < 0,9$	<i>Mallin ennustekyky on kiitettävä</i>
		$c \geq 0,9$	<i>Mallin ennustekyky on erinomainen</i>

Kuviosta yksi nähdään lopulliselle malliversiolle (Taulukko 3) laskettu ROC-käyrä. Liitteeseen kolme on laskettu vertailun vuoksi ROC-käyrä myös askeltavan regression mallille.

Kuvio 1 Lopulliselle mallille laskettu ROC-käyrä



Taulukosta neljä nähdään, että yhteensopivuusindeksi saa arvon 80,5 eli mallin ennustekyky on kiitettävä. Luku on 0,2 prosenttiyksikköä suurempi kuin askeltavasta regressiosta saadulla mallilla (Liite 3). Taulukossa on nähtävissä myös Somersin D (0,61) eli niin sanottu tarkkuuskerroin (Accuracy ratio) (Somers 1962, 799–811), joka on Laitisen ym. (2014, 170) mukaan nykyisin eniten käytetty luokittelukyvyyn mitta. Se mittaa mallin tarkkuutta suhteessa täydelliseen malliin. Kerroin lasketaan ROC-käyrän alle jäävän pinta-alan avulla, jolloin tarkkuuskerroin saadaan seuraavasti $AR = 2 * c - 1$. Mikäli mallilla ei ole ennustekykyä ja se tuottaa täysin satunnaisten ennusteen, tarkkuuskertoimen arvo on nolla. Keskimääräisen mallin tarkkuuskerroin saa puolestaan arvon 0,5

ja täydellisesti ennustavan mallin arvo on yksi. (Laitinen ym. 2014, 170.) Kuten taulukosta neljä havaitaan, estimoidun mallin tarkkuuskerroin 0,61 on selvästi yli keskimääräisen arvon 0,5.

Taulukko 4 Yhteensopivuusindeksi ja tarkkuuskerroin

Ennustettujen todennäköisyyksien ja havaittujen vastausten välinen yhteys			
Yhdenmukaisten parien osuus (%)	80,5	Somersin D	0,61
Ristiriitaisten parien osuus (%)	19,5	Gamma	0,61
Tasasijojen osuus (%)	0	Tau-a	0,012
Parien lukumäärä	91599552	c	0,805

Edellä kuvatun lisäksi mallin arvioinnissa on hyvä tarkastella mallin kalibraatiota eli tutkia sitä, kuinka hyvin ennustetut ja todelliset havaitut konkurssitodennäköisyydet vastaavat toisiaan. Tätä varten SAS-ohjelmassa PROC LOGISTIC komentoon yhdistettiin LACKFIT-valinta, jonka avulla pystyttiin tuottamaan niin sanottu Hosmer-Lemeshow -taulukko (Hosmer ym. 2013, 157–169). Taulukko muodostetaan siten, että estimoidut todennäköisyydet lajitellaan nousevaan järjestykseen, jonka jälkeen aineisto voidaan jakaa luokkiin.

Taulukko 5 Hosmer-Lemeshow -taulukko

Luokka	Yhteensä	Konkurssi=1		Konkurssi=0	
		Havaitut	Estimoidut	Havaitut	Estimoidut
1	9759	17	16,24	9742	9742,76
2	9754	23	33,4	9731	9720,6
3	9761	32	44,65	9729	9716,35
4	9762	23	54,6	9739	9707,4
5	9758	39	64,66	9719	9693,34
6	9754	44	75,29	9710	9678,71
7	9760	56	87,54	9704	9672,46
8	9759	61	102,96	9698	9656,04
9	9758	138	125,15	9620	9632,85
10	9747	515	343,47	9232	9403,53

Taulukosta viisi havaitaan, että etenkin luokissa 2–8 ennustetut konkurssitodennäköisyydet systemaattisesti ylittävät havaitut todennäköisyydet. Puolestaan luokissa 9–10 ennusteet alittavat havaitut

todennäköisyydet. Havainnot saattavat indikoida muuttujien epälineaarisuuksista.

Kyseisen testin ongelma kuitenkin on kuten Allison artikkelissaan *Measures of Fit for Logistic* (2014, 6) huomauttaa, että kyseiset tulokset voivat riippua merkittävästi käytettyjen luokkien määrästä ja ei ole olemassa teoriaa siitä, kuinka suuri tuo luokkien määrä tulisi olla. SAS-ohjelma ei mahdollista luokkien määrän muuttamista kyseisen testin kohdalla. Allison huomauttaa vielä, että vaikka mahdollisia muuttujien muunnoksia lisättäisiin malliin, niin tämä ei monestikaan paranna Hosmer-Lemeshowin testin tuloksia. (*Measures of Fit for Logistic* 2014). Edellä käsitellyn perusteella mallin kalibrointia ei katsottu tarpeelliseksi lähteä modifioimaan.

Tulosten perusteella voidaan olettaa, että valittuun malliin on onnistuttu löytämään käytettävistä muuttujista ne, jotka luotettavasti kuvaavat tutkittua ilmiötä. Valittu muuttujakombinaatio tuottaa muun muassa paremman yhteensopivuusindeksin ja tarkkuuskertoimen arvon kuin askeltavan regression menetelmä. Lisäksi saadut tulokset ja malliin valikoituneet muuttujat ovat juuri sellaisia, jotka ovat toistuneet suurimmassa osassa tutkimuksista.

Tutkimusten mukaan seuraavien tekijöiden on havaittu yleensä esiintyvän maksukyvyttömissä yrityksissä (Laitinen ym. 2014, 109):

- Heikko vakavaraisuus
- Heikko tulorahoitus
- Likvidien varojen vähäisyys suhteessa lyhytaikaiseen vieraaseen pääomaan

Vakavaraisuutta kuvaavana tunnuslukuna tässä tutkimuksessa oli *Omaavaraisuusaste*, jonka kasvu pienensi yrityksen konkurssiriskiä. Tulorahoitukseen liittyy olennaisesti kannattavuus ja sitä kuvaavia tunnuslukuja

olivat *Käyttökateprosentti* sekä *Kokonaispääoman tuottoprosentti*. Yrityksen maksuvalmiuden tunnuslukuna käytettiin *Current ratiota*. Muuttujia, jotka eivät varsinaisesti kuulu yrityksen vakavaraisuuden, kannattavuuden tai maksuvalmiuden piiriin, olivat *Käyttöpääomaprosentti* sekä *Yhtiön ikä* vuosissa.

Käyttöpääomaprosentti riippuu pitkälti yrityksen toimialasta mutta se antaa hyvän kuvan rahoitustarpeesta, jonka yrityksen toiminnan laajentaminen aiheuttaa. Kyseistä tunnuslukua voidaan analysoida tarkemmin laskemalla sen eri osatekijöiden kiertoaikoja. (Yritystutkimus ry 2011, 69.) Tutkimuksen tulosten mukaan *Käyttöpääomaprosentin* kasvu pienensi yrityksen konkurssiriskin todennäköisyyttä. Myös yrityksen ikä on muuttuja, jonka on havaittu olevan keskeinen selittäjä yrityksen konkurssitapahtumassa (Laitinen ym. 2014, 32).

Havainnollistetaan saatuja tuloksia vielä kolmelle aineistosta poimitulle yritykselle, jotka on nimetty Yritys1, Yritys2 ja Yritys3.

Taulukko 6 Mallilla lasketut konkurssitodennäköisyydet esimerkkiyrityksille

	Tilikauden päättymis- päivämäärä	Yhtiön ikä vuosissa	Käyttö- kate %	Kokonaispää- oman tuotto %	Current ratio	Omavarai- suusaste %	Käyttö- pääoma %	Ennus- tettu tn.
Yritys1	31.12.2012	88	51,1	5,3	4,6	86,6	14	0,001
Yritys2	31.12.2012	13	-6,1	-24,7	2,5	78,8	4,4	0,020
Yritys3	31.12.2012	9	-50,7	-167,4	0,7	9,5	10,1	0,467

Taulukon kuusi viimeisessä sarakkeessa on laskettu yrityksen todennäköisyys joutua konkurssiin seuraavan vuoden aikana tarkasteluhetkestä, joka on tässä tapauksessa tilikauden päättymispäivämäärä 31.12.2012. Voidaan havaita, että ensimmäisellä yrityksellä kaikki mallin muuttujat saavat hyvät arvot. Yritys on ollut jo pitkään markkinoilla ja sen vakavaraisuus ja maksuvalmius ovat erittäin korkealla tasolla, kun niitä mitataan *Omavaraisuusaste*prosentilla sekä *Current Ratiolla*. Yrityksen kannattavuutta kuvaavat tunnusluvut *Käyttökateprosentti* ja *Kokonaispääoman tuotto*prosentti ovat myös hyvällä tasolla ja tämän vuoksi yrityksen todennäköisyys joutua konkurssiin seuraavan vuoden aikana tarkasteluhetkestä onkin varsin pieni saaden arvon 0,1 %.

Yritys kaksi on ollut toiminnassa huomattavasti vähemmän aikaa kuin ensimmäinen yritys ja muuttujan arvot saavat muutoinkin kauttaaltaan hieman pienempiä arvoja kuin Yrityksessä yksi. Erot eivät kuitenkaan ole kovin suuria lukuun ottamatta eroja kannattavuuden tunnusluvussa (*Käyttökateprosentti* ja *Kokonaispääoman tuotto-%*), jotka saavat nyt negatiiviset arvot.

Yritys kolmen tunnusluvut ovat kauttaaltaan heikot (Taulukko 6) mutta erityisesti kannattavuus on matalalla tasolla. Tämän vuoksi yrityksen todennäköisyys joutua konkurssiin seuraavan vuoden aikana on korkea saaden arvon 46.7 %. Tarkastellaan esimerkinomaisesti mitä tapahtuisi kyseisen yrityksen konkurssitodennäköisyydelle, jos sen kannattavuuden tunnusluvut (*Käyttökateprosentti* ja *Kokonaispääoman tuotto-%*) olisivat samalla tasolla kuin Yrityksellä kaksi.

Taulukko 7 Yritys3

	Tilikauden päättymis- päivämäärä	Yhtiön ikä vuosissa	Käyttö- kate-%	Kokonaispää- oman tuotto -%	Current ratio	Omavarai- suusaste- %	Käyttö- pääoma- %	Ennus- tettu tn.
Yritys3	31.12.2012	9	-6,1	-24,7	0,7	9,5	10,1	0,023

Taulukosta seitsemän havaitaan, että konkurssitodennäköisyys pienenee oleellisesti, kun yrityksen kannattavuus paranee. Muutosten jälkeen Yrityksen kolme todennäköisyys joutua konkurssiin vuoden aikana tilikauden päättämispäivämäärästä on nyt 2,3 %.

4.3 Mallin testaus ulkopuolisella aineistolla

Lopullisen mallin valinnan jälkeen on tärkeä validoida sen toiminta testiaineiston avulla. Validoinnin avulla varmistetaan, että malli on sovellettavissa kohdejoukkoon ja ettei yliparametrisointia esiinny. (Siddiqi 2006, 127.) Tätä tarkoitusta varten vuoden 2013 aineisto jätettiin pois varsinaisesta mallinnusaineistosta ja sen avulla suoritettiin mallin validointi.

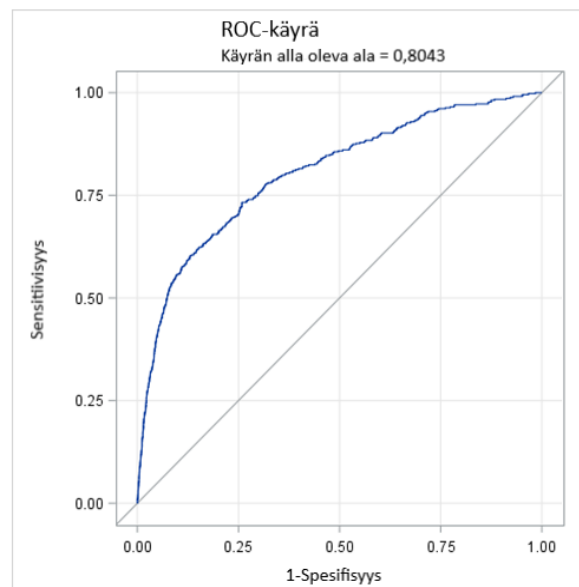
Kun lopullinen malli sovitettiin validointiaineistoon, niin yhteensopivuusindeksin arvoksi saatiin 80,43 (Taulukko 5), joka on lähestulkoon sama arvo kuin mallinnusaineistossa (80,49). Myös tarkkuuskertoimen arvo on likimäärin sama kuin taulukossa neljä. Tämän perusteella estimoitu malli näyttäisi soveltuvan erittäin hyvin ulkopuoliseen dataan.

Taulukko 5 Validointiaineiston yhteensopivuusindeksi ja tarkkuuskerroin

ROC-käyrän laskentaan liittyvät tunnusluvut							
ROC	Mann-Whitney				Somersin D	Gamma	Tau-a
	Ala	Keskivirhe	95% Waldin luottamusväli				
ROC	0.8043	0.0116	0.7816	0.8270	0.6086	0.6086	0.0114

Taulukossa viisi on nähtävissä myös 95 prosentin luottamusväli [0,7816, 0,8270] yhteensopivuusindeksin arvolle sekä sen keskivirhe 0,0116. Kuviosta kaksi havaitaan, että validointiaineistolle laskettu ROC-käyrä on hyvin samanmuotoinen kuin alkuperäiselle mallinnusdatalle laskettu käyrä.

Kuvio 2 Validointiaineistolle laskettu ROC-käyrä



Yllä esitetty logistisen regressiomallin soveltaminen validointiaineistoon osoittaa, että sen ennustekyky säilyy lähes samana kuin mallinnusaineistossakin.

Tulos on linjassa muun muassa Laitisen (1995) artikkelin kanssa hänen vertaillessaan yleisesti käytettyjen mallien suorituskykyä testiaineistossa. Tuolloin tehdyn tarkastelun perusteella saatiin tulos, jonka mukaan logistinen regressiomalli toimi tehokkaammin kuin verrokkinsa.

5 JOHTOPÄÄTÖKSET

Tässä tutkimuksessa pyrittiin löytämään mahdollisimman kompakti mutta silti hyvin kuvaava malli yrityksen konkurssiriskille. Aineistona käytettiin Suomen Asiakastieto Oy:n yritysaineistoa ja tilastolliseksi mallinnusmenetelmäksi valikoitui logistinen regressio.

Parhaiten yrityksen konkurssiriskiä ennustivat seuraavat muuttujat: *Yhtiön ikä vuosissa*, *Käyttökateprosentti*, *Kokonaispääoman tuottoprosentti*, *Omavaraisuusaste-prosentti*, *Käyttöpääomaprocentti* sekä *Current ratio*. Kaikkien muuttujien kertoimet saivat negatiivisen arvon eli muuttujien arvojen kasvu laski yrityksen konkurssitodennäköisyyttä.

Saadut tulokset ovat linjassa aiempien aihetta käsittelevien tutkimusten kanssa, joissa yrityksen iän lisäksi erilaiset yrityksen maksuvalmiutta, vakavaraisuutta ja kannattavuutta mittaavat tunnusluvut ovat osoittautuneet empiirisissä tutkimuksissa keskeisimmiksi konkurssia ennustaviksi tekijöiksi. (Laitinen ym. 2014, 32, 109.)

Tutkimuksessa estimoitu malli validoitiin myös ulkopuolisella testiaineistolla. Tulosten mukaan mallin ennustekyky säilyi lähes samalla tasolla kuin mallinnusaineistossakin. Samankaltaiseen tulokseen päätyi muun muassa Laitinen (1995) vertaillaan logistista regressiomallia yleisimpiin verrokkimalleihinsa.

Mallinnuksessa käytettiin hyväksi lähinnä tilinpäätöksestä laskettuja tunnuslukuja, jotka toimivatkin varsin hyvin konkurssia ennustavina tekijöinä. Tästä huolimatta mallia kannattaisi laajentaa vielä muilla potentiaalisilla tekijöillä kuten esimerkiksi toimialamuuttujilla, yrityksen maksuhäiriöillä sekä yrityksen johdon toimintaa ja ominaisuuksia kuvaavilla muuttujilla. Tällaisen

lisäinformaation hyödyntämisellä mallin ennustekykyä voitaisiin todennäköisesti parantaa vielä merkittävästi.

LÄHTEET

- Agresti, A. 2013. *Categorical Data Analysis*, 3. laitos. Wiley, New York.
- Aziz M.A. & Dar. H.A. 2006. Predicting corporate bankruptcy: where we stand? *Corporate Governance*, Vol. 6:1, 18–33.
- Bluhm, C., Overbeck, L. & Wagner, C. 2003. *An Introduction to Credit Risk Modelling*. Boca Raton, Florida: Chapman & Hall/CRC Financial Mathematics Series
- Davison, A. C. 2003. *Statistical Models*. Cambridge University Press.
- Djerf, K. 1995. Credit Scoring yksityishenkilöiden riskienhallinnassa. Teoksessa *Luottoriskin hallinnan tilastolliset menetelmät*. Helsinki: Suomen Asiakastieto Oy, 11–30.
- Greene, W. H. 1997. *Econometric Analysis*. Third Edition. Prentice-Hall International Inc.
- Hastie, T., Tibshirani, R. ja Friedman, J. 2009. *The Elements of Statistical Learning*. 2. laitos. Springer, New York.
- Hosmer, D.W., Lemeshow, Stanley ja Sturdivant, Rodney X. 2013. *Applied Logistic Regression*. 3. laitos. Wiley, New York.
- Laitinen, E. K. 1995. Tilastolliset menetelmät konkurssin ennustamisessa. Teoksessa *Luottoriskin hallinnan tilastolliset menetelmät*. Helsinki: Suomen Asiakastieto Oy, 63–92.
- Laitinen, T. & Kankaanpää, M. 1999. Comparative analysis of failure prediction models: the Finnish case. *The European Accounting Review*, 8:1, 67–92.
- Laitinen, T. & Laitinen, E.K. 2014. *Yrityksen maksukyky – Arviointi ja ennakointi*. Helsinki: KHT-Media Oy.
- Ohlson, J. A. 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, Vol. 18, No. 1, 109–131.
- Pulkkinen, K. 1995. Monimuuttujamallin sovellus yritysten riskiarvioinnissa. Teoksessa *Luottoriskin hallinnan tilastolliset menetelmät*. Helsinki: Suomen Asiakastieto Oy, 31–46.
- Reinikainen, J. 2015. *Efficient design and modeling strategies for follow-up studies with time-varying covariates*. Väitöskirja. Jyväskylän yliopisto.
- Measures of Fit for Logistic Regression 2014. <https://support.sas.com/en/support-home.html> 16.01.2018.
- Siddiqi, N. 2006. *Credit Risk Scorecards - Developing and Implementing Intelligent Credit Scoring*. Hoboken, New Jersey: John Wiley & Sons, Inc.
- Somers, R.H. 1962. A New Asymmetric Measure of Association for Ordinal Variables. *American Sociological Review*. Vol. 27:6, 799–811.
- Tunnuslukuopas 2018. Balance Consulting. <http://www.balanceconsulting.fi/tunnusluvut/kayttopaaoma> 23.01.2018.

- Yritystutkimus ry 2011. Yritystutkimuksen tilinpäätösanalyysi. Helsinki:
Gaudeamus Helsinki University Press 9. korj. laitos
- Zavgren C.V. 1985. Assessing the vulnerability of failure of American industrial
firms: A logistic analysis. *Journal of Business Finance &
Accounting*, 12, 19–25.

LIITTEET

Liite 1 Käytettyjen muuttujien tunnuslukuja

Muuttuja	Lkm.	Keskiarvo	Keskihajonta	Minimi	Maksimi
Yhtiön ikä vuosissa	97572	17.5827799	13.7235275	0	119.0000000
Liikevaihto, eur	97572	5686878.44	87916421.24	100006.00	11992000000
Käyttökate-%	97572	11.6706237	31.8559258	-3043.50	275.7000000
Liikevoitto-%	97572	5.9817714	38.8528509	-3043.50	304.4000000
Nettotulos-%	97572	2.6082780	118.4681169	-9999.90	9999.90
Sijoitetun pääoman tuotto-%	97572	17.2758609	116.4501333	-9999.90	9999.90
Kokonaispääoman tuotto-%	97572	13.1963596	43.1296715	-930.0000000	9999.90
Quick ratio	97572	4.7478672	83.5246437	-1463.70	9999.90
Current ratio	97572	7.1134803	112.3496560	-1337.80	9999.90
Omavaraisuusaste-%	97572	30.2609181	152.1931822	-9999.90	100.0000000
Net gearing	97572	5.0382251	136.5397029	-9035.90	9999.90
Suht. velkaantuneisuus-%	97572	116.9525673	528.6326650	-26.0000000	9999.90
Käyttöpääoma-%	97572	12.9833210	59.5463768	-3022.20	3168.10
Taseen loppusumma, eur	97572	7146706.77	140214297	1000.00	19379000000

Liite 2 Käytettyjen muuttujien väliset korrelaatiot

	Yhtiön ikä vuosissa	Liikevaihto, eur	Käyttökate- %	Liikevoitto- %	Nettotulos- %	Sijoitetun pääoman tuotto- %	Kokonaispääoman tuotto- %	Quick ratio	Current ratio	Omavaraisuusaste- %	Net gearing	Suht. velkaantuneisuus- %	Käyttöpääoma- %	Taseen loppusumma, eur
Yhtiön ikä vuosissa	1.00000	0.06020	-0.01639	-0.01566	0.01229	-0.03018	-0.06359	0.00837	0.01136	0.06677	-0.01813	-0.01295	0.04507	0.03626
Liikevaihto, eur	0.06020	1.00000	-0.00752	-0.00192	0.00067	-0.00019	-0.00564	0.0089	0.0004	<0.001	<0.001	<0.001	<0.001	<0.001
Käyttökate- %	<0.001	-0.00752	1.00000	0.5494	0.8344	0.9517	0.0781	0.4985	0.5140	0.4518	0.7608	0.0533	0.4326	<0.001
Liikevoitto- %	-0.01639	0.0188	1.00000	0.80342	0.28861	0.15266	0.24569	0.01104	0.00218	0.15491	0.02187	-0.10160	-0.00850	-0.00999
Nettotulos- %	<0.001	-0.00192	0.80342	1.00000	<0.001	<0.001	0.23509	0.0006	0.4966	<0.001	<0.001	<0.001	0.0079	0.0018
Sijoitetun pääoman tuotto- %	0.01229	0.00067	0.28861	0.33068	1.00000	0.04887	0.08547	0.0353	0.4640	<0.001	<0.001	<0.001	0.01483	-0.01660
Kokonaispääoman tuotto- %	-0.03018	0.00019	0.15266	0.14266	0.04887	1.00000	0.32006	0.00049	0.01258	0.06772	-0.00228	-0.24947	0.04985	-0.00249
Quick ratio	<0.001	0.0781	<0.001	<0.001	<0.001	<0.001	<0.001	0.8779	0.4193	<0.001	0.4763	<0.001	<0.001	0.4371
Current ratio	0.00837	-0.00217	0.01104	0.00674	0.01426	0.00049	1.00000	0.00235	-0.00279	0.09474	-0.00609	-0.02389	-0.00109	0.00334
Omavaraisuusaste- %	0.0089	0.4985	0.0006	0.0353	<0.001	<0.001	<0.001	1.00000	0.85151	0.01759	-0.00920	-0.05215	0.7342	0.2964
Net gearing	0.0004	0.5140	0.4966	0.4640	0.01258	0.00259	0.00279	<0.001	<0.001	<0.001	0.5857	0.1776	0.1169	-0.00810
Suht. velkaantuneisuus- %	<0.001	-0.00097	0.02187	-0.01846	-0.00228	-0.00609	-0.00920	0.85151	1.00000	0.01867	-0.00275	-0.00396	0.17709	0.00066
Käyttöpääoma- %	0.04507	0.7608	<0.001	<0.001	0.4763	0.0569	0.0041	<0.001	<0.001	1.00000	1.00000	1.00000	<0.001	0.8358
Taseen loppusumma, eur	0.03626	-0.00619	-0.10160	-0.28985	-0.24947	-0.02389	-0.05215	0.01759	0.1867	-0.0630	0.0492	<0.001	0.3239	0.00414
	<0.001	0.0533	<0.001	<0.001	<0.001	<0.001	<0.001	0.1776	<0.001	0.0492	<0.001	0.12885	0.00091	0.1958
	0.04326	-0.00251	-0.00850	0.01483	0.04985	0.07342	0.1169	<0.001	<0.001	<0.001	<0.001	-0.00243	1.00000	0.00403
	0.45038	0.4326	-0.00999	-0.01660	-0.00249	-0.00334	-0.00810	0.00048	0.00066	0.00414	0.0403	0.13604	-0.00190	0.2080
	<0.001	<0.001	0.0079	<0.001	<0.001	0.1169	0.1169	<0.001	<0.001	<0.001	0.7757	0.4485	0.00091	0.13604
	0.03626	0.45038	-0.00999	-0.01660	-0.00249	-0.00334	-0.00810	0.00048	0.00066	0.00414	0.0403	0.13604	-0.00190	0.5535
	<0.001	<0.001	0.0018	<0.001	0.4371	0.2964	0.0114	0.8807	0.8358	0.1958	0.2080	<0.001	0.5535	1.00000

Liite 3 ROC-käyrä askeltavalle regressiolle

Ennustettujen todennäköisyyksien ja havaittujen vastausten välinen yhteys			
Yhdenmukaisten parien osuus (%)	80,3	Somersin D	0.606
Ristiriitaisten parien osuus (%)	19,7	Gamma	0.606
Tasasijojen osuus (%)	0	Tau-a	0.012
Parien lukumäärä	91599552	c	0.803

