

Konkursprediksjon

En empirisk analyse

Christoffer Foss Hals og Kenneth Fegri

Veileder: Professor Svein-Arne Persson

Masterutredning innenfor finansiell økonomi

NORGES HANDELSHØYSKOLE

Denne utredningen er gjennomført som et ledd i siviløkonomutdanningen ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at høyskolen innestår for de metoder som er anvendt, de resultater som er fremkommet eller de konklusjoner som er trukket i arbeidet.

Sammendrag

I denne utredningen konstruerer og estimerer vi en enkel modell som kan måle risiko i form av konkurssannsynlighet. Vi ønsker å undersøke om det er mulig å predikere konkurs ett, to og tre år frem i tid på bakgrunn av offentlig regnskapsinformasjon.

Ved å sammenligne beregnede nøkkeltall fra bedrifter som har gått konkurs med bedrifter som fortsatt eksisterer, har vi kommet frem til tre modeller som beregner konkurssannsynlighet for ett, to og tre år frem i tid. Vår konklusjon er at modellen for ett år predikerer godt, modellen for to år er tilfredsstillende, mens modellen for tre år bør man være varsom med. Det er vår konklusjon at konkurssannsynlighet beregnet ut fra modellene helst bør ses på i relativ sammenheng, enten ved at en bedrift sammenlignes over flere år eller at ved sannsynligheten til flere bedrifter sammenlignes. Selskapene som sammenlignes bør være av samme størrelse.

Forord

Denne utredningen er skrevet som en del av masterstudiet ved Norges Handelshøyskole våren 2012, og utgjør ett semester med fulltidsstudier.

Temaet for utredningen er valgt som følge av interesse for finansiell informasjon i form av regnskap, samt kredittratingselskaperens metoder og rolle i samfunnet generelt. Vi ønsket å kombinere disse to interessene og med det utgangspunktet å estimere vår egen modell for konkurssannsynlighet. I tillegg til å tilegne oss ny kunnskap på disse områdene, virket tanken om å forhåpentligvis komme frem til et praktisk anvendbart resultat forlokkende.

Arbeidsprosessen har vært både tidkrevende og utfordrende, men absolutt givende og lærerik. En av utfordringene har vært knyttet til arbeidet med å sette sammen datautvalget analysene er basert på, da det relativt sett er få selskaper som går konkurs i Norge.

Til slutt vil vi benytte anledningen til å takke vår veileder, professor Svein-Arne Persson, for gode diskusjoner og nyttige tilbakemeldinger underveis i prosessen.

Bergen, juni 2012.

Christoffer Foss Hals

Kenneth Fegri

Innhold

1. Innledning	7
1.1 Motivasjon	7
1.2 Avgrensning	8
1.3 Struktur	9
1.4 Problemformulering	9
1.4.1 Hva menes med konkurs?.....	10
1.4.2 Hva er hensikten med å predikere konkurser?.....	10
2. Litteraturgjennomgang	12
2.1 Metoder for estimering av konkursprediksjonsmodeller	12
2.2 Datautvalg	14
2.2.1 Konkursbedrifter.....	15
2.3 Valg av uavhengige variabler	16
2.4 Resultater fra tidligere studier	20
3. Metode	22
3.1 Univariate vs. multivariate metoder	22
3.2 Komponentanalyse	22
3.2.1 Egenverdi og egenvektorer.....	22
3.2.2 Utførelse av komponentanalyse.....	23
3.2.3 Tolkning av komponentene.....	24
3.2.4 Utvelgelse av variabler.....	24
3.3 T-test	25
3.4 Levens test	26
3.5 Regresjon	27
3.5.1 Forutsetninger for regresjon.....	27
3.5.2 Logit.....	28
3.5.3 Maximum Likelihood Estimation.....	30
3.5.4 Signifikanstesting av koeffisientene.....	30
3.5.5 Tolkning av koeffisientene.....	31
3.5.6 Goodness of fit.....	33
3.6 Korrigering av utvalgsskjevhet	36

4. Analyse	38
4.1 Datautvalg	38
4.1.1 Innsamling av data	38
4.1.2 Størrelse på datautvalget	39
4.2 Valg av uavhengige variabler	41
4.2.1 Operasjonalisering av regnskapene	41
4.2.2 Primærmengde av nøkkeltall	43
4.2.3 Ekstremverdier.....	44
4.2.4 Komponentanalyse	45
4.2.5 Valg av nøkkeltall.....	48
4.3 Logit-analyse	51
4.3.1 Resultat av regresjon	52
4.3.2 Resultater fra modellene	61
5. Diskusjon	66
5.1 Oppsummering og konklusjon.....	66
5.2 Kritikk av modellene.....	69
5.3 Forslag til videre forskning.....	71
6. Referanser	72
7. Appendiks	75
7.1 Grenseverdier nøkkeltall.....	75
7.2 Egenverdier og egenvektorer.....	76
7.3 Korrelasjon mellom nøkkeltall og komponent	79
7.4 Innbyrdes korrelasjon blant valgte nøkkeltall	80
7.5 Levens test for lik varians.....	81
7.6 T-test for likt gjennomsnitt.....	81
7.7 Minitab-utskrift logit-analyse for ett års prognoseperiode.....	82
7.8 Minitab-utskrift logit-analyse for to års prognoseperiode	83
7.9 Minitab-utskrift logit-analyse for tre års prognoseperiode	84
7.10 Konkurssannsynlighet for konkursbedrifter ett års prognoseperiode	85
7.11 Konkurssannsynlighet for konkursbedrifter to års prognoseperiode	85
7.12 Konkurssannsynlighet for konkursbedrifter tre års prognoseperiode.....	86
7.13 Fremgangsmåte beregning av konkurssannsynlighet	87

Figurer og tabeller

Figur 1.1: Åpnede konkurser av AS og ASA pr kvartal 2006 – 2011 (1:ssb.no).....	8
Figur 3. 1: S-kurven – sannsynligheter plottet mot logit-verdier (Tuftes 2000).....	29
Figur 4. 1: S-kurve ett års prognoseperiode	54
Figur 4. 2: S-kurve to års prognoseperiode	55
Figur 4. 3: S-kurve tre års prognoseperiode	56
Tabell 4. 1: Antall AS konkurs årlig i forhold til totalt antall AS.....	39
Tabell 4. 2: Oversikt over ikke-konkursbedrifter i datautvalget	40
Tabell 4. 3: Oversikt over konkursbedrifter i datautvalget	40
Tabell 4. 4: Fordeling av konkursbedrifter i datautvalget pr år.....	41
Tabell 4. 5: Resultatregnskap med symboler	42
Tabell 4. 6: Balanseregnskap med symboler.....	43
Tabell 4. 7: Primærmengde nøkkeltall	44
Tabell 4. 8: Komponenter med egenverdi over 1	46
Tabell 4. 9: Valgte nøkkeltall og deres korrelasjon med respektive komponenter	49
Tabell 4. 10: Andel feilklassifiseringer ett års prognoseperiode.....	64
Tabell 4. 11: Andel feilklassifiseringer to års prognoseperiode.....	64
Tabell 4. 12: Andel feilklassifiseringer tre års prognoseperiode.....	65
Tabell 5. 1: Oversikt over hvilke nøkkeltall som inngår i de estimerte modellene.....	68

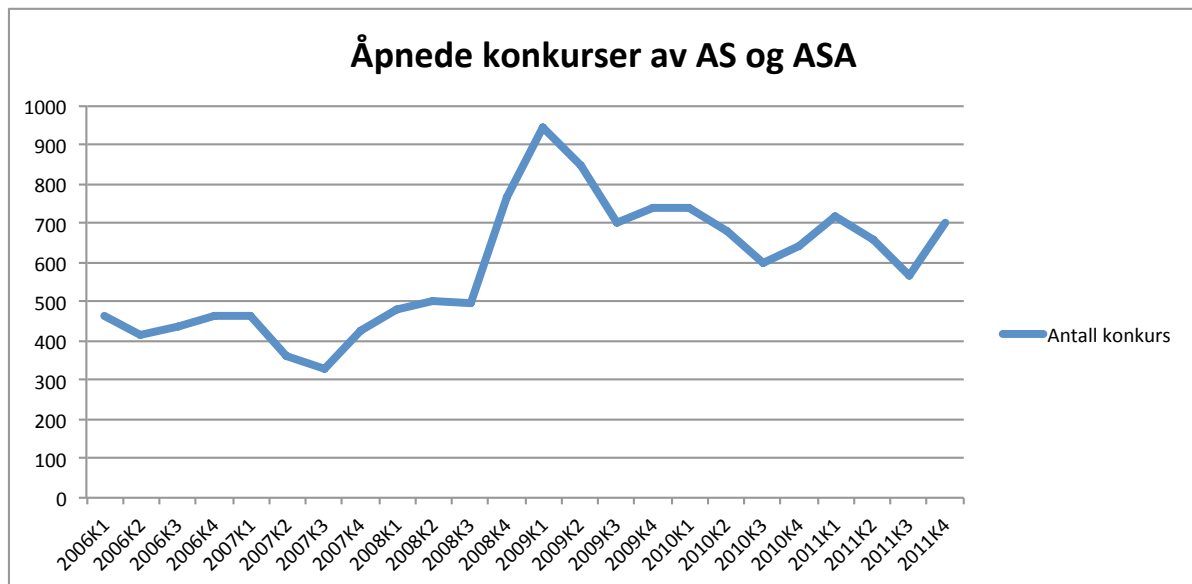
1. Innledning

1.1 Motivasjon

Den globale finanskrisen hadde sitt utspring i USA. Finanskrisen fulgte et klassisk kriseforløp med overoppheting av økonomien og etterfølgende krakk i finansmarkedene. Som følge av sterkt globaliserte finansmarkeder spredde krisen seg raskt til store deler av verden. Krisen utartet seg etter 15. september 2008 da investeringsbanken Lehman Brothers gikk konkurs. Dette førte til en akutt tillitskrise i finansmarkedene; risikopremiene steg og kredittmarkedene stoppet opp (Finanskriseutvalget 2011). Dette fikk konsekvenser, og utover høsten 2008 kom det en økning i antall konkurser også i Norge - hovedsakelig blant små og mellomstore bedrifter. (2: Pressemelding Brønnøysundregistrene 2009).

En lengre periode med lave renter i forkant av kollapsen i finansmarkedene førte til at investorene søkte etter muligheter for god avkastning, noe som igjen førte til nyvinninger i finansbransjen. Boliglån ble verdipapirisert, og de store kredittratingselskapene som Moody's, Standard & Poors og Fitch gav disse verdipapirene kredittrating. Kredittratingselskapene har i ettertid blitt beskyldt for å ha gitt toppkarakterer til verdipapirer som ikke burde hatt toppkarakterer. I den senere tids statsfinansielle krise har de samme kredittratingselskapene på ny blitt utsatt for kritikk, blant annet med anklager om at de reagerer for sent, og således at markedet allerede har absorbert realitetene før ratingselskapene.

Som følge av de siste års finansielle uroligheter har det vært en økning i antall konkurser i Norge (1: ssb.no). Figur 1.1 viser en økning i antall konkurser siste kvartal 2008 og første kvartal 2009. Det ser imidlertid ut til at antallet konkurser har gått noe tilbake og stabiliserte seg mot slutten av 2009, dog på et noe høyere nivå enn før finanskrisen inntraff i 2008.



Figur 1.1: Åpnede konkurser av AS og ASA per kvartal 2006 – 2011 (1:ssb.no)

På bakgrunn av at konkursproblematikken fortsatt er dagsaktuell, jf. den senere tids uroligheter i finansmarkedene, samt at kritikken av ratingselskapene synes å fortsette, mener vi det vil være interessant å konstruere en modell for konkurssannsynlighet. I den senere tid har fokus i hovedsak vært på land, men vi ønsker likevel å se på bedrifter. Når det gjelder forskning på konkursprediksjon, er det gjort en del innenfor dette feltet tidligere som vi vil se i litteraturgjennomgangen, spesielt med amerikanske data. Så vidt oss bekjent er det i Norge ikke forsket noe særlig på konkurssannsynlighet siden starten av 2000-tallet.

1.2 Avgrensning

Utredningen har til hensikt å fokusere på Norge. Med det menes det at vi vil bruke regnskapsinformasjon fra norske selskaper i arbeidet med å konstruere en modell for konkursprediksjon. Ut i fra offentlig regnskapsinformasjon kan man analysere ulike aspekter rundt bedriftens finansielle situasjon og vurdere dens risiko. Dette er informasjon som er tilgjengelig, slik at modellene kan relativt enkelt kan brukes av både bedrifter og privatpersoner.

Når det gjelder valg av data og selskapsform, velger vi å bruke data fra aksjeselskaper (ikke ASA), med en nedre grense for størrelse. Modellen vil ikke rettes inn mot de største

bedriftene. Grunnen til dette fokuset er at store selskaper svært sjelden konkurs i Norge, noe som betyr at det vil være vanskelig å finne data for selskapene. Valg av data og begrunnelse for dette vil kommenteres mer inngående senere i utredningen. Hva metode angår, vil det fokuseres på komponentanalyse og multivariat logit-analyse for å estimere en egen modell for konkurssannsynlighet.

Som vist i litteraturgjennomgangen under, er det benyttet ulike metoder for estimering av prediksjonsmodeller i tidligere studier. Det kunne derfor vært interessant å benytte ulike metoder på samme data, for eksempel univariat analyse, multipl diskriminant analyse, samt den logistiske regresjonsmetoden probit. Dette vil imidlertid være en omfattende oppgave, og er således utenfor denne utredningens rammer.

1.3 Struktur

Utredningen er delt inn i fem hoveddeler. Den første delen inneholder en innledning og problemformulering. Del to er en litteraturgjennomgang av viktige bidrag innenfor konkursprediksjon. Del tre tar for seg teorien bak metodene som blir benyttet i utredningen. Del fire inneholder analysen og resultatene fra de estimerte modellene, samt detaljer rundt datautvalget og innsamlingen av dataene. Del fem inneholder en diskusjon av resultatene fra del fire.

1.4 Problemformulering

Ved å formulere en velbegrunnet problemstilling vil man til enhver tid ha klart for seg problemet utredningen har til hensikt å belyse, noe som vil være til hjelp i arbeidet. I det følgende vil vi først avklare hva som menes med konkurs i denne utredningen, da dette danner et viktig grunnlag. Vi vil deretter gå videre inn på argumenter for hvorfor det er viktig å kunne predikere konkurser. Til slutt under punkt 1.4.2 vil den endelige problemstillingen presenteres.

1.4.1 Hva menes med konkurs?

Konkurser er i Norge regulert gjennom Lov om gjeldsforhandling og konkurs av 1984 (Konkursloven). Det følger av Konkursloven §60 at «er skyldneren insolvent, skal vedkommendes bo tas under konkursbehandling når det begjæres av skyldneren eller en fordringshaver».

Konkursloven §61 regulerer insolvens: «Skyldneren er insolvent når denne ikke kan oppfylle sine forpliktelser etter hvert som de forfaller, med mindre betalingsudyktigheten må antas å være forbigående. Insolvens foreligger likevel ikke når skyldnerens eiendeler og inntekter til sammen antas å kunne gi full dekning for skyldnerens forpliktelser, selv om oppfyllelsen av forpliktelsene vil bli forsinket ved at dekning må søkes ved salg av eiendelene».

Loven sier altså at en person eller bedrift må være insolvent for at det kan begjæres konkurs, og at noen faktisk må begjære vedkommende konkurs for at det kan åpnes konkurs. Kort oppsummert er en bedrift eller person insolvent når den ikke har inntekter til å møte sine forpliktelser over tid, og der salg av vedkommendes eiendeler ikke vil være nok til å dekke kravet.

1.4.2 Hva er hensikten med å predikere konkurser?

Mange vil kanskje henvise til Darwins evolusjonsteori og hevde at de bedriftene som har livets rett vil overleve, jf. tesen om «survival of the fittest». Det kan argumenteres for at konkurs er en mekanisme som på sett og vis rydder opp i samfunnet og regulerer markedene slik at det er de levedyktige foretakene som blir igjen. På en annen side vil det være flere parter som taper penger ved en konkurs - det være seg leverandører, investorer og/eller kunder. Det er derfor ønskelig å forebygge konkurser. Et annet viktig aspekt er det faktum at arbeidsplasser går tapt, slik at det å forhindre konkurser også vil være av samfunnsmessig interesse. En konkurs vil altså som regel føre med seg en rekke konsekvenser utover det at foretaket opphører.

Ved å predikere sannsynligheten for konkurs, kan man oppdage symptomer på at noe er galt. Det er likevel slik at en konkursprediksjonsmodell ikke nødvendigvis finner hva som er galt eller gir en eksakt sannsynlighet for at en konkurs vil inntreffe. Dersom man klarer å forutse,

eller finne indikasjoner på, at et foretak kan komme i vanskeligheter, kan det settes inn nødvendige ressurser på å finne hva problemet er og på den måten redusere risikoen for konkurs. Det vil være av interesse å kunne forutse disse vanskelighetene så tidlig som mulig. Det at man forholdsvis enkelt kan måle risikoen for sannsynlighet, kan i seg selv virke forebyggende da de ulike aktørene i markedet løpende kan gjøre nødvendige tilpasninger for å redusere risikoen og sannsynlighet for konkurs.

En konkursprediksjonsmodell kan altså brukes til å måle risiko, og ved å benytte slike modeller kan man prise og allokere risiko i kapitalmarkedene, samt finne indikasjoner på at noe er galt. Når en konkursprediksjonsmodell er estimert, kan man på en relativt lite ressurskrevende måte måle risiko.

Formålet med utredningen er altså å konstruere og estimere en generell modell som kan predikere konkurssannsynligheten til en bedrift på bakgrunn av nøkkeltall beregnet ut ifra offentlig regnskapsinformasjon. Med generell så menes det en modell som kan brukes på tvers av alle bransjer.

Utredningen har til hensikt å besvare følgende spørsmål:

- 1) Kan man ved hjelp av nøkkeltall beregnet ut ifra offentlig regnskapsinformasjon estimere en god, generell konkursprediksjonsmodell inntil tre år før konkurs?*

2. Litteraturgjennomgang

Fra 1960-tallet og frem til i dag har det blitt gjort en rekke studier på konkursprediksjon ved hjelp av ulike nøkkeltallsanalyser. Den største delen av litteraturen innenfor konkursprediksjon er empiriske undersøkelser, men det finnes også noen forholdsvis enkle teoretiske modeller. Denne delen vil ta for seg en del viktige bidrag innenfor dette fagfeltet, samt utviklingen av dette, og således underbygge de metoder denne utredningen benytter seg av. Litteraturgjennomgangen er delt opp slik at den tar for seg de ulike stegene i en analyse av konkursprediksjon ved først å se på hvilke metoder tidligere studier har benyttet, for så å se på innsamling og sammensetningen av datautvalget. Det neste steget blir å se på hvordan nøkkeltallene (de uavhengige variablene) er valgt ut, for så til slutt gjennomgå resultatene de ulike studiene/metodene har gitt.

2.1 Metoder for estimering av konkursprediksjonsmodeller

De første studiene av konkursprediksjon basert på nøkkeltall ble gjort ved hjelp av såkalte univariate analyser. En univariat analyse vil si at man analyserer hvert enkelt nøkkeltall individuelt. Tamari (1966) baserte sin analyse på en univariat modell. Denne studien benyttet ikke en statistisk metode, men hvert enkelt nøkkeltall ble vurdert separat ved at de ble gitt poeng fra 0 til 100. Poengberegningen ble gjort på bakgrunn av en subjektiv vurdering av nøkkeltallenes verdi. Vurderingen av en bedrifts finansielle situasjon ble gjort ved at nøkkeltallenes poeng ble summert, og de nøkkeltallene som ble ansett som viktigst ble vektet tyngst. For å kunne klassifisere bedriftene, beregnet Tamari en grense («cut-off») for hvert enkelt nøkkeltall.

Studien til Beaver (1966) *omtales* ofte som den første innenfor nøkkeltallsbasert konkursprediksjon, og ble også utført som en univariat analyse. Studien sammenliknet gjennomsnittsverdien til nøkkeltallene for ulike konkurs- og ikke-konkursbedrifter, hvor resultatene ble analysert grafisk i en profilanalyse. Studien gikk imidlertid lengre, og brukte en dikotom klassifiseringstest for å teste de ulike nøkkeltallenes prediksjonsevne, samt analyse av likelihood ratios (signifikanstest). Fordelen med å bruke en univariat tilnærming er at metoden er enkel. Problemet med metoden er imidlertid at man kan få motstridende

konklusjoner for ulike nøkkeltall. Et nøkkeltall kan predikere konkurs, mens et annet ikke gjør det. I Beavers forslag til videre forskning foreslås det at man kan få høyere prediksjonsevne ved å analysere flere nøkkeltall samtidig. Den første studien som inkluderte flere nøkkeltall samtidig ble publisert av Altman (1968). Studien benyttet seg av multippel diskriminantanalyse, og kom frem til en modell bestående av fem variabler. «The Z-score model», som den ble kalt, predikerte konkurs dersom den aktuelle bedriftens score ble lavere enn en gitt grense. Denne modellen har vært gjenstand for en del kritikk i etterkant, eksempelvis hevdet Joy and Tollefsson (1975) at diskriminantanalysen ble benyttet ukritisk. Blant annet som følge av denne kritikken ble ZETA-modellen introdusert av Altman et.al (1977). Denne modellen bygger på Z-score modellen og er estimert med samme statistiske metode, men studien ser i tillegg på enkelte andre aspekter rundt bruk av multippel diskriminantanalyse. Modellen tok også hensyn til endrede regler for finansiell rapportering, samt det faktum at gjennomsnittlig størrelse på konkursbedriftene var blitt betydelig større enn de som ble lagt til grunn for Z-score modellen (Altman et al. 1977). Eklund og Knutsen (1997) estimerte ulike modeller ved hjelp av lineær diskriminantanalyse, blant annet reestimerte de Altmans Z-score modell basert på norske data. Av modellene de estimerte predikerte Z-score modellen best.

Ohlson (1980) var blant de første som benyttet logit-analyse for å estimere en konkursprediksjonsmodell. Ohlson nevner tre grunner til at han benytter denne metoden fremfor multippel diskriminantanalyse. Det første argumentet går på at man er ikke avhengig av at varians-kovariansmatrisen er den samme for begge grupper (konkurs- og ikke-konkursbedrifter). Videre argumenteres det for at resultatene fra en multippel diskriminantanalyse er lite intuitive, og at de fungerer mer som en ordinal rangering av variablene (inndelt i konkurs/ikke-konkurs). Ohlson mener også det er noe uklart hva man oppnår eller eventuelt taper ved å pare bedriftene slik man gjør i diskriminantanalysen. Skogsvik (1988) bygger sin analyse på probit-metoden. Skogsvik (1988) undersøker også hvorvidt det å bruke inflasjonsjusterte regnskapstall (gjenanskaffelseskost) øker prediksjonsevnen til modellen. Studien konkluderer med at det er liten forskjell i prediksjonsevnen, og at det er et betydelig merarbeid med å bruke inflasjonsjusterte tall, da de fleste regnskap baserer seg på historisk kost. Sjøvoll (1999) tok utgangspunkt i en modell Norges Bank benyttet for å måle kredittrisiko, og estimerte ved hjelp av probit-analyse en modell for konkurssannsynlighet. Modellen baserte seg på både teori om industriell organisering og analyse av nøkkeltall fra bedrifters regnskaper. Bernhardsen (2001)

estimerte en modell for konkurssannsynlighet ved hjelp av logit-analyse basert på norske data. Studien så også på hvorvidt bedrifter som på et tidspunkt er insolvente går konkurs. I tillegg ble effekten av å inkludere makroindikatorer (bankenes utlånstap) undersøkt, noe som gav positiv innvirkning på modellens prediksjonsevne.

2.2 Datautvalg

I studier av konkursprediksjon benyttes det et utvalg av både konkursbedrifter og ikke-konkursbedrifter. Hvordan disse bedriftene selekteres kan ha stor betydning for å få et representativt utvalg. I så måte kan man se for seg to måter å gjøre dette på; enten å trekke ut bedrifter tilfeldig fra hele populasjonen, for så å identifisere bedriftene som har gått konkurs, eller å først identifisere konkursbedriftene, for så å sette sammen utvalget på bakgrunn av. På grunn av at det er relativt sett få konkurser er det hovedsakelig sistnevnte metode som benyttes. De fleste studier har også en større andel konkursbedrifter enn hva som er tilfelle i virkeligheten.

Beaver (1966) valgte først ut 79 bedrifter som hadde gått konkurs i perioden 1954 - 1964. Disse konkursbedriftene ble så parett med friske bedrifter etter gitte kriterier. Disse kriteriene gikk ut på at bedriftene skulle tilhøre samme bransje og være av omtrent samme størrelse. Data for de siste fem årene før konkurs ble hentet fra Moody's Industrial Manual. Bedriftene i utvalget representerte 38 ulike industrier, og hadde 6 millioner dollar i balanseførte eiendeler i gjennomsnitt. Altman (1968) benyttet seg av samme metode for valg av data. Han valgte først ut 33 konkursbedrifter som så ble parett med 33 friske bedrifter fra samme bransje og omtrent samme størrelse. I denne studien var gjennomsnittlig størrelse på balanseførte eiendeler hos konkursbedriftene 6,4 millioner dollar, mens tallet var 9,6 millioner dollar for de friske. Altman avfeide at denne forskjellen hadde noen innvirkning, og hevdet at forholdstall er av en slik natur at de tar hensyn til og demper effekten av størrelse. På samme måte som Beaver, hentet Altman data for fem år før konkurs fra Moody's Industrial Manual. I ZETA-modellen (Altman et al. 1977) ble det også benyttet et utvalg av parede bedrifter, nærmere bestemt 53 konkursbedrifter og 58 ikke-konkurs. Bakgrunnen for ulikt antall i de to gruppene var manglende data for fem av konkursbedriftene.

Ohlson (1980) er blant de første studiene som benyttet et tilnærmet likt forhold som i virkeligheten. Denne studien ble basert på 105 konkursbedrifter og 2058 ikke-konkurs for perioden 1970-1976. Regnskapsdataene ble hentet fra publiserte 10-K årsrapporter (etter standard fra U.S. Securities and Exchange Commission), i motsetning til Moody's Industrial Manual som de tidligere studiene benyttet. Resultatene fra logit-metoden som Ohlson benyttet, er i større grad avhengig av fordelingen mellom konkurs- og ikke-konkursbedrifter, noe som er bakgrunnen for det store utvalget av bedrifter. Bedriftene i studien måtte enten være omsatt på børs eller i et OTC-marked (over-the-counter), i tillegg til å være klassifisert som et industriselskap. Ohlson (1980) tar også hensyn til tidspunktet regnskapsinformasjonen offentliggjøres, altså om bedriften gikk konkurs før eller etter at regnskapene ble tilgjengelige.

Skogsvik (1988) benyttet KKR-banken ved Handelshøgskolen i Stockholm (database med regnskapsinformasjon) som utgangspunkt. Databasen inneholder data for svenske aksjeselskaper innenfor industrien. Det ble identifisert 51 konkursbedrifter med tilgjengelig data de siste seks år før konkurs og 328 ikke-konkursbedrifter med data tilgjengelig for perioden 1966 – 1980. I alle bedriftene var det minst 200 ansatte og balanseførte eiendeler på minst 200 millioner svenske kroner. Både Sjøvoll (1999) og Bernhardsen (2001) benyttet data fra SEBRA-databasen. Dette er en database utviklet av Norges Bank med regnskapsinformasjon til alle norske selskaper med rapporteringsplikt. Sjøvoll (1999) inkluderte data for aksjeselskaper for perioden 1993-1998, men utelot bedrifter med ISIC-kode 75000 eller høyere da dette regnes som offentlige tjenester, og selskaper innenfor finansiell tjenesteyting. Bernhardsen (2001) benyttet data for perioden 1988-1999 for ikke-konkursbedrifter, og perioden 1990-1999 for konkursbedrifter. Bedrifter med balanseførte eiendeler på mindre enn 250 000 norske kroner ble ekskludert fra utvalget. Totalt antall observasjoner i utvalget var 398 689, inkludert 8 436 konkurser.

2.2.1 Konkursbedrifter

I konkursprediksjonsmodellene er hvordan man definerer konkurs av stor betydning for hvilke bedrifter som inkluderes i datautvalget. Som dette punktet vil vise er det noe forskjellig definisjon i ulike studier. Noen studier benytter en ren legalistisk tolkning hvor det må være åpnet konkurs i henhold til lovens definisjon. Andre studier inkluderer i utvalget

av konkursbedrifter eksempelvis også bedrifter som har mottatt statlig tilskudd for å overleve.

Beaver (1966) brukte begrepet konkurs («failed») om bedrifter som ikke kan møte sine forpliktelser ved forfall. I denne studien defineres en bedrift som konkurs dersom bedriften er insolvent («bankrupt»), obligasjonsgjeld misligholdes, bankkonti overtrekkes, eller at bedriften ikke betaler utbytte til eiere av preferanseaksjer. Altman (1968) benyttet i sin studie konkursbedrifter som har konkursbegjæring i henhold til Chapter 10 i the National Bankruptcy Act. I ZETA-modellen (Alman et al. 1977) ble det i tillegg inkludert bedrifter i utvalget av konkursbedrifter som har fått et betydelig tilskudd av myndighetene, gjennomgått en tvunget sammenslåing og ett tilfelle hvor banken har tatt over bedriften. Ohlson (1980) inkluderte utelukkende konkursbedrifter hvor det var åpnet konkurs etter den amerikanske lovens definisjon. Skogsvik (1988) inkluderte bedrifter som var rammet av konkurs («finansiell kris») ved at de enten måtte legges ned eller overlevde ved hjelp av annen form for økonomisk omorganisering. Konkursbedriftene i studien til Skogsvik (1988) omfatter insolvente bedrifter som er begjært konkurs, bedrifter hvor styret har bestemt nedleggelse og bedrifter som har fått statlig tilskudd i en eller annen form. Det forutsettes at bedriftene som hadde mottatt tilskudd ikke hadde overlevd uten. Både Sjøvoll (1999) og Bernhardsen (2001) baserer sine analyser på bedrifter som er begjært konkurs etter norsk lov og hvor regnskapsinformasjonen er registrert av Dun & Bradstreet.

2.3 Valg av uavhengige variabler

Det er i tidligere studier benyttet ulike metoder for valg av uavhengige variabler. Enkelte studier baserer sine valg på populariteten i litteraturen, mens andre baserer seg på teoretiske modeller eller statistisk analyse. Det finnes flere argumenter for hvorfor man bør gjøre en grundig jobb i utvelgelsen av de uavhengige variablene, og ikke inkludere flere variabler enn nødvendig. Først og fremst vil en modell med svært mange variabler gjøre den vanskeligere å benytte for brukerne, og kreve større ressurser ved datainnhenting. Et annet argument mot å inkludere for mange variabler, er at det i dette tilfellet kan oppstå multikollinearitet (korrelasjon mellom de uavhengige variablene, Keller 2009) mellom enkelte nøkkeltall, noe man ønsker å unngå.

Beaver (1966) hadde en kontantstrømtilnærming og karakteriserte bedriften som et reservoar av likvide eiendeler. Nøkkeltallene ble valgt ut på bakgrunn av tre kriterier; (i) de skulle være kontantstrømbaserte, (ii) popularitet i litteraturen og (iii) resultater fra tidligere studier. Dersom ett av disse kriteriene var oppfylt, ble nøkkeltallet inkludert i studien. Utgangspunktet ble 30 ulike nøkkeltall, hvorav seks ble med i den endelige modellen. Utvelgelsen ble foretatt ved hjelp av en profilanalyse, hvor gjennomsnittsverdiene til nøkkeltallene ble sammenlignet mellom de to gruppene av bedrifter (konkurs/ikke-konkurs). Det måtte være en tydelig forskjell, subjektivt sett, for at nøkkeltallene ble valgt ut. Beaver fant at kontantstrøm i forhold til total gjeld predikerte konkurs best. Altman (1968) brukte, som Beaver (1966), også popularitet i litteraturen når den initiale mengden nøkkeltallene skulle velges ut. I tillegg ble nøkkeltallene vurdert ut fra potensiell relevans for studien, samt at det ble tatt med et par nøkkeltall som ikke var benyttet i tidligere studier. Dette resulterte i 22 nøkkeltall delt inn i fem kategorier; likviditet, lønnsomhet, finansiering, soliditet og aktivitet. Fem nøkkeltall ble valgt ut på bakgrunn av flere tester. Først ble de testet for statistisk signifikans hvor også de relative bidrag fra hvert nøkkeltall ble evaluert. Deretter ble det foretatt en evaluering av korrelasjonen mellom variablene. Det ble også gjennomført en klassifiseringstest på hvert enkelt nøkkeltall for å finne hvor godt de predikerte konkurs. Til slutt gjorde Altman (1968) en subjektiv vurdering av nøkkeltallene. Det kan være verdt å merke seg at den variabelen Beaver (1966) fant å predikere konkurs best (kontantstrøm/total gjeld), ikke er med i Altmans studie. Altman begrunner dette med manglende informasjon om avskrivninger i regnskapsdataene, slik at det ble vanskelig å finne eksakt kontantstrøm. Altman et al. (1977) tok utgangspunkt i 27 nøkkeltall i ZETA-modellen, de fleste fra tidligere studier, samt at de la til et par egne.

Ohlson (1980) benyttet seg av nøkkeltall fra tidligere studier og gjorde ingen forsøk utover dette på å inkludere nøkkeltall som tidligere ikke hadde vært brukt. Modellen til Ohlson inkluderer totalt ni uavhengige variabler; størrelse, total gjeld/total eier, arbeidskapital/total eier, kortsiktig gjeld/omløpsmidler, årsresultat/total eier, driftsresultat/total eier, og to dummy-variabler som var 1 dersom inntektene hadde vært negative de siste to år og hvis gjelden oversteg eiendelene, ellers 0. Skogsvik (1988) gjorde derimot et omfattende arbeid rundt valg av uavhengige variabler i sin studie, hvor primærmengden bestod av henholdsvis 71 og 79 nøkkeltall basert på historisk-kost og inflasjonsjusterte tall. Nøkkeltallene ble utledet fra sju kategorier; lønnsomhet, kostnadsstruktur, kapitalens omløpshastighet, likviditet, eiendelsstruktur, finansiell struktur

og vekst. Ved hjelp av komponentanalyse ble antallet redusert til henholdsvis 17 og 20. Univariate analyser ble så benyttet på hvert enkelt nøkkeltall for å redusere antallet ytterligere før den endelige probit-analysen hvor syv nøkkeltall ble inkludert. Blant nøkkeltallene som er beregnet ut fra historisk kost, finner Skogsvik at det er særlig to som er viktige; lånekostnader (rentekostnader/gjennomsnittlig gjeld siste to perioder,) og soliditet (egenkapital/totale eiendeler).

Sjøvoll (1999) tok utgangspunkt i Norges Banks modell for risikoklassifisering. Denne modellen baserer seg på tre nøkkeltall; selvfinansiering (driftsresultat etter skatt og avskrivninger/langsiktig gjeld), soliditet (opptjent egenkapital/total kapital) og likviditet ((likvider-kortsiktig gjeld)/driftsinntekter). Sjøvoll utvidet modellen med sju variabler; soliditet (opptjent egenkapital/total gjeld), kontantstrøm (kontantstrøm/total gjeld), gjeldsbyrde (gjeld/kontantoverskudd), alder, og dummy-variabel henholdsvis for om bedriften er i restaurantbransjen eller holder på med utleie av fast eiendom, samt for manglende regnskapsinformasjon. Bernhardsen (2001) tok også utgangspunkt i Norges Banks modell for kredittrisiko, og ved hjelp av litt prøving og feiling kom han frem til totalt 12 variabler som ble inkludert i modellen. Variablene er delt inn i kategoriene; likviditet, lønnsomhet, soliditet, alder, størrelse og industrikarakteristikk.

Scott (1981) går igjennom og utleder noen forholdsvis enkle teoretiske modeller for konkursprediksjon i et forsøk på å forklare variablene som brukes i empiriske konkursprediksjonsmodeller. Modellene klassifiseres i fire hovedgrupper. I den første gruppen antar man at bedriften er operativ i to perioder, mens i de tre siste gruppene antar man en uendelig tidshorisont.

Den første gruppen av modeller er de enkleste, hvor bedriften antas å være operativ i første periode, mens i andre periode likvideres eiendelene. Bedriften går konkurs dersom gjelder overstiger likvidasjonsverdien; $V < D$, hvor V er verdien av bedriftens eiendeler ved slutten av perioden, og D representerer beløpet bedriften skylder sine kreditorer. Modellen utledes så med å ta hensyn til sannsynlighetsfordeling. På standardisert form med forventet verdi μ_v og standardavvik σ_v til eiendelene, gir modellen følgende konkursbetingelse:

$$\frac{V - \mu_v}{\sigma_v} < \frac{D - \mu_v}{\sigma_v}$$

Scott (1981) finner at denne modellen ikke forklarer resultatene fra de empiriske metodene i særlig grad. Black & Scholes opsjonsprisindemodell klassifiseres også i denne kategorien.

Neste gruppe modeller bygger på «the gambler's ruin model» (Wilcox 1976). Her antar man at bedriften har en gitt mengde kapital, hvor K representerer eiendelenes likvidasjonsverdi, og at endringen (betegnet Z) i denne er tilfeldig fordelt. Dersom K blir negativ, går bedriften konkurs. Videre antas det at bedriften ikke har tilgang til kapitalmarkeder og at bedriften må likvidere eiendeler ved tap. Bedriften vil gå konkurs neste periode dersom: $K+Z<0$. Denne modellen kan forklare noen av de empiriske resultatene dersom man lar nøkkeltall fra regnskapene være erstatninger for likvidasjonsverdier. Ved en slik tolkning av modellen, vil K tilsvare bokført verdi av egenkapitalen. Z vil være endringen i opptjent egenkapital. Med forventet verdi μ_z og standardavvik σ_z til Z går bedriften konkurs dersom:

$$\frac{Z - \mu_z}{\sigma_z} < \frac{-(\mu_z + K)}{\sigma_z}$$

I den tredje modellgruppen forutsettes det at bedriften har tilgang på et perfekt kapitalmarked, noe som betyr at aksjonærene ikke trenger å selge eiendeler for å dekke alle tap - de kan også selge verdipapirer. Man finner optimal verdi av bedriften først, for deretter å finne en finansieringsplan som gjør at man oppnår denne verdien. Ut fra dette finner man hva verdien av egenkapitalen vil være ved utløpet av perioden. Hvis den optimale verdien er mindre enn tapet, går bedriften konkurs: $S+X<0$, hvor S er optimal verdi (markedsverdi egenkapital) neste periode, og X neste periodes resultat (tap). Hvis man lar μ_x være forventet resultat, og σ_x standardavviket til neste periodes resultat, går bedriften konkurs dersom:

$$\frac{X - \mu_x}{\sigma_x} < \frac{-(\mu_x + S)}{\sigma_x}$$

I fjerde og siste modellgruppe forutsettes det at bedriften har tilgang til et imperfekt kapitalmarked. Med imperfekt menes det at kapital fra et eksternt kapitalmarked innebærer transaksjonskostnader, betegnet c . Dette betyr at i disse modellene er ikke eksternt og intern finansiering substitutter. I følge modellen går bedriften konkurs dersom det inntreffer et tap som er større eller lik egenkapitalen; $S+(I+c)*(X-\Delta K) \leq 0$. På standardisert form går bedriften konkurs dersom:

$$\frac{X - \mu_x}{\sigma_x} < \frac{-\mu_x - \Delta K - S/(1+c)}{\sigma_x}$$

Dette uttrykket bygger både på «the gambler's ruin model» og modellene som antar perfekte kapitalmarkeder. I tillegg til variablene forklart ovenfor, er ΔK er endringen i bedriftens eiendeler, og kan ses på som optimal mengde eiendeler å selge for en konkurstruet bedrift. $S/(1+c)$ representerer den maksimale mengde egenkapital en konkurstruet bedrift kan utstede, når transaksjonskostnadene (c) er tatt hensyn til.

Konklusjonen til Scott (1981) er at de empiriske modellene til eksempelvis Beaver (1966) og ZETA-modellen (Altman et al. 1977) kan forklares ut ifra teoretiske modeller, dog ikke helt perfekt.

2.4 Resultater fra tidligere studier

Hvor godt modellene predikerer vurderes ofte ut fra total feilklassifisering og andel type 1- og type 2-feil. Type 1-feil betyr at en konkursbedrift predikeres ikke-konkurs, mens type 2-feil betyr at en ikke-konkursbedrift predikeres konkurs. I de fleste tilfeller vil type 1-feil være mer kostbare enn type 2-feil (Altman et al. 1981). Det kan være vanskelig å direkte sammenligne resultatene fra de ulike modellene, da det settes en klassifiseringsgrense for å finne type 1- og type 2-feil. I de fleste tilfeller vil denne grensen settes ut fra en noe subjektiv vurdering. De ulike studiene har også noe forskjellige antagelser og forutsetninger rundt feilklassifiseringskostnader, samt forskjellige datautvalg og bruk av metoder.

Beaver (1966) fant at type 1-feil utgjorde 22 % ett år før konkurs, mens type 2-feilene utgjorde 5 %. Studien konkluderer med at modellen predikerer konkurs inntil fem år forut konkurs. Altman (1968) oppnår bedre prediksjon med sin multivariate analyse ett år før konkurs. Type 1-feil ett år før utgjorde 6,1 %, mens type 2-feil utgjorde 3 %. Konklusjonen ble at Z-score modellen predikerer nøyaktig inntil to år forut for konkurs. Blant annet som følge av ønske om bedre prediksjonsevne lengre forut for en konkurs ble ZETA-modellen estimert (Altman et al. 1977). ZETA-modellen viste seg å være en klar forbedring av Z-score modellen. Type 1-feil ett år før konkurs utgjorde i denne modellen 3,8 %, og 30,2 % fem år

før mot Z-score modellens 64 % fem år før. Type 2-feil økte imidlertid noe i ZETA-modellen, til 10,3 % ett år før konkurs, mot 3 % i Z-score modellen.

Ohlson (1980) finner grensen som minimerer summen av feilklassifiseringer til å være 0,038. Med denne grensen gir modellen 12,4 % type 1-feil, og 17,4 % type 2-feil ett år før konkurs. Dette er høyere enn de tidligere modellene, men ifølge Ohlson (1980) er det flere mulige grunner til at det er vanskelig å sammenligne hans resultater direkte med for eksempel Z-score modellen og ZETA-modellen. En mulig grunn som nevnes er at studien til Ohlson tar hensyn til når bedriftene går konkurs i forhold til når regnskapene offentliggjøres - dette er ikke gjort i tidligere studier. Det nevnes også at forskjellen kan ha noe med at studiene er gjort med data fra forskjellige tidsperioder, samt at utvalgene av konkurs- og ikke-konkursbedrifter er forskjellige.

Skogsvik (1988) tester ulike tilnærminger til feilklassifiseringskostnader og minimering av feilandeler. Studien benytter både et vektet og aritmetisk snitt av type 1- og type 2-feil når det testes for total feilandel. Ett år før konkurs gav modellen total feilandel på 9,8 % og 15,2 % for historisk kostbaserte tall, for henholdsvis et vektet og aritmetisk snitt av type 1- og type 2-feil. Skogsvik (1988) predikerte inntil seks år før konkurs, og de ulike modellene gir tilsvarende resultater som første året. Dette gjelder både analyse basert på historisk kost og inflasjonsjusterte tall, dog predikerer modellen best de to første årene. Sjøvoll (1999) predikerer konkurs for årene 1989-1996, og feilandelen for disse årene er fra 3,6 % til 4,9 %, men som det påpekes i studien er dette sterkt avhengig av hvilken klassifiseringsgrense som settes. Modellen til Bernhardsen (2001) klassifiserer 83 % korrekt når modellen estimeres på datautvalget. Bernhardsen inkluderte også en makroindikator, reell boligprisindeks, for å forklare bankenes utlånstap. Konklusjonen er at det forbedrer modellens statistiske signifikans.

3. Metode

Denne delen vil ta for seg relevant teori bak de metodene som benyttes. Først vil metodene komponentanalyse, t-test og Levens test gjennomgås. Disse tre metodene benyttes i utvelgelsen av de uavhengige variablene (nøkkeltallene). Deretter vil ulike forutsetninger for regresjon og teorien bak logit-analysen gjennomgås. Logit-analysen blir benyttet til å estimere de endelige modellene for ett, to og tre års prognoseperiode. Til slutt vil ulike metoder for å teste de estimerte modellene forklare.

3.1 Univariate vs. multivariate metoder

Som nevnt tidligere, tar univariat-analyse kun hensyn til én variabel av gangen, mens ved multivariate analyser vurderes flere variabler simultant (Manly 2005). Selv om univariat-analysen viser at en variabel er signifikant, er det ikke sikkert den er signifikant i kombinasjon med andre variabler. I analyser med mange variabler er derfor multivariate metoder å foretrekke.

3.2 Komponentanalyse

Formålet med prinsippal komponentanalyse (PCA) er å finne et utvalg av variabler som kan forklare mesteparten av variansen i datasettet. Komponentanalyse gjøres ved å finne et mindre antall ukorrelerte komponenter som er lineære kombinasjoner av de originale variablene (Manly 2005). Komponentene må være ukorrelerte slik at de dekker forskjellige «dimensjoner» av dataene. Best resultat oppnås hvis mange av variablene er høyt korrelerte, positivt eller negativt. Da er det enklere å finne passende komponenter, samt representative variabler for hver komponent.

3.2.1 Egenverdi og egenvektorer

For å forstå hvordan en komponentanalyse utføres er det viktig å ha noe kunnskap om egenverdier og egenvektorer, og hvorfor de er viktige i tolkningen av resultatet av

komponentanalysen. En vektor er en størrelse som kan tilordnes en retning. I et todimensjonalt rom vil en vektor kunne representeres ved en pil med en viss lengde og retning. Å finne egenvektorene til kovariansmatrisen er viktig fordi det sier noe om egenskapene til de aktuelle komponentene. Egenverdiene viser hvor stor andel av den totale variansen til dataene den enkelte komponent forklarer (Manly 2005).

For A som er en $n \times n$ matrise, kalles λ en *egenverdi* for A dersom det fins en vektor x slik at:

$$Ax = \lambda x$$

I så fall kalles x en *egenvektor* for A (Sydsæter og Øksendal 2006). På matriseform blir dette:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}$$

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

3.2.2 Utførelse av komponentanalyse

I en komponentanalyse er det vanlig å standardisere alle variablene slik at de har null i snitt og én i varians for å unngå at noen variabler påvirker komponentene for mye (Manly 2005). Deretter beregnes korrelasjonsmatrisen (kovariansmatrisen hvis man ikke standardiserer) og dens egenverdier. Summen av egenverdiene i korrelasjonsmatrisen er lik summen av variansen, som igjen er lik summen av antall variabler (standardiserte variabler med 1 i varians).

Den første komponenten finnes ved å finne en lineær sammenheng mellom variablene, som varierer mest mulig, gitt at summen av de kvadrerte egenvektorene skal være lik én. De neste komponentene finnes på samme måte, men komponentene kan ikke korrelere med hverandre. På den måten kan man si at den første komponenten står for størst andel av

variansen i datasettet, mens den andre komponenten står for nest størst andel og så videre. Til slutt velges kun de komponentene som står for størsteparten av variansen i datasettet, målt ved egenverdi. Komponentene som velges er basert på skjønn og er avhengig av hvilket formål analysen skal brukes til. En vanlig metode er å velge alle komponentene med egenverdi høyere enn én, som vil si at komponenten representerer en større andel av variasjonen enn hver enkelt variabel i snitt. Ved å benytte denne metoden står man vanligvis igjen med komponenter som står for 80-90 % av den totale variasjonen (Manly 2005).

3.2.3 Tolkning av komponentene

Egenvektorene tolkes slik at de variablene med høyest egenvektorer står for mesteparten av variasjonen i komponenten og følgelig har høyest korrelasjon med komponenten. Hvis de mest betydningsfulle egenvektorene har samme fortegn, er som regel tolkningen av komponenten grei (Jolliffe 1986). For en komponent med kun høye egenvektorer for ulike mål på rentabilitet er følgelig komponenten et generelt mål på rentabilitet. Gitt at denne er den første komponenten (altså høyest egenverdi) vil størsteparten av variasjonen blant bedriftene i datasettet være på grunn av rentabilitet. Har komponenten både positive og negative egenvektorer som er omtrent like betydningsfulle (har relativt lik egenvektor i absoluttverdi), tolkes dette som en kontrast mellom de aktuelle variablene. En komponent med høy egenverdi for likvide midler og omtrent samme verdi på kortsiktig gjeld, bare med negativt fortegn, tolkes følgelig som et mål på likviditet. Gitt at dette er komponent nummer 2, vil størsteparten av variasjonen blant bedriftene komme av likviditet, sett bort i fra rentabilitet (komponent nummer 1).

3.2.4 Utvelgelse av variabler

En relativt enkel måte å velge ut variablene på er ved å se på korrelasjonen mellom hver enkelt variabel og den/de komponenten(e) den/de inngår i. Det er vanlig å velge én eller noen få variabler fra hver komponent som representerer komponenten godt (Jolliffe 1986). Skogsvik (1988) benytter absoluttverdi over 0,8 som grense for at et nøkkeltall velges. Hvis flere variabler korrelerer sterkt med komponenten er det lite hensiktsmessig å velge mer enn én variabel per komponent da disse vil korrelere med hverandre. Som nevnt er hensikten

med PCA å finne et utvalg av variabler som representer forskjellige «dimensjoner» av dataene.

Til slutt bør det nevnes at en komponent som forklarer mye av variasjonen i de *uavhengige* variablene ikke nødvendigvis forklarer like mye av variasjonen for den *avhengige* variabelen ved regresjon (Jolliffe 1986). Tvert imot kan en komponent med en relativt lav egenverdi ha relativ stor betydning for den avhengige variabelen i en regresjon. Dette er en utfordring når komponentene skal velges ut for regresjon da det er vanskelig å se på forhånd hvilke variabler som vil påvirke den avhengige variabelen mest. Det bør også nevnes at det finnes andre måter å velge ut variabler som tar bedre hensyn til problemene nevnt her. Disse metodene er imidlertid såpass kompliserte at vi velger å ikke gå nærmere inn på de. Et annet argument for dette valget er at i vår komponentanalyse var det forholdsvis mange komponenter med egenverdi over 1, slik at sannsynligheten for ikke å velge ut en viktig variabel var relativt liten.

3.3 T-test

En t-test er en statistisk metode for hypotesetesting og er basert på Student t-fordeling. Metoden kan brukes for å teste om gjennomsnittsverdien i to datautvalg er signifikant forskjellig. T-test kan også brukes for å teste om koeffisientene i en regresjonslikning er signifikant forskjellig fra null (Keller 2009). Normalfordelte data er en forutsetning bak t-testen, men hvis utvalget er stort kan testen benyttes uansett, hvor n (antall observasjoner) ≥ 40 ofte regnes som tilstrekkelig stort (Moore 2007).

For å kontrollere hvorvidt gjennomsnittsverdien til de ulike nøkkeltallene er signifikant forskjellige i gruppen av henholdsvis konkurs- og ikke-konkursbedrifter, kan to-sidige t-tester for to uavhengige utvalg benyttes. Nullhypotesen er at gjennomsnittet for de ulike nøkkeltallene er like i de to gruppene (her representert ved μ_0 og μ_1), mens alternativhypotesen er at gjennomsnittet er ulikt.

$$H_0: \mu_0 = \mu_1$$

$$H_A: \mu_0 \neq \mu_1$$

For å kunne utføre en t-test må gjennomsnittsverdien og variansen til hver populasjon være kjent. I tilfeller hvor den totale populasjonens gjennomsnitt og varians er ukjent (som oftest er de ukjent) brukes de to utvalgenes verdier til å estimere disse. Hvordan t-testen utføres avhenger av om variansen i de to populasjonene hvor datautvalgene er hentet fra er lik eller ulik i de to gruppene. Testobservatoren til t-testen for tilfeller hvor variansen i de to populasjonene er lik ($\sigma_0^2 = \sigma_1^2$) er som følger (Keller 2009):

$$t = \frac{(x_0 - x_1) - (\mu_0 - \mu_1)}{\sqrt{s_p^2 \left(\frac{1}{n_0} + \frac{1}{n_1}\right)}}, \text{ med frihetsgrader} = n_0 + n_1 - 2, \text{ og hvor}$$

$$s_p^2 = \frac{(n_0 - 1)s_0^2 + (n_1 - 1)s_1^2}{n_0 + n_1 - 2}$$

x_0 og x_1 er gjennomsnittsverdien i de to utvalgene, mens μ_0 og μ_1 er gjennomsnittsverdien i de to totale populasjonene. s_p^2 er et estimat på den samlede variansen, det vil si et vektet gjennomsnitt av de to utvalgenes varians med hensyn på antall frihetsgrader.

I tilfeller med ulik varians i de to populasjonene ($\sigma_0^2 \neq \sigma_1^2$) er testobservatoren som følger (Keller 2009):

$$t = \frac{(x_0 - x_1) - (\mu_0 - \mu_1)}{\sqrt{\left(\frac{1}{n_0} + \frac{1}{n_1}\right)}}, \text{ med frihetsgrader } \nu = \frac{\left(\frac{s_0^2}{n_0} + \frac{s_1^2}{n_1}\right)^2}{\frac{(s_0^2/n_0)^2}{n_0-1} + \frac{(s_1^2/n_1)^2}{n_1-1}}$$

hvor x_0 , x_1 , μ_0 og μ_1 betegner det samme som under antakelse om lik varians i de to utvalgene, mens s_0^2 og s_1^2 betegner de to utvalgenes varians.

3.4 Levens test

Som vist i forrige avsnitt er det av interesse å bestemme hvorvidt det er grunnlag for å si at variansen er lik eller ulik i de to populasjonene før en t-test kan gjennomføres. Til dette kan man bruke en tradisjonell F-test. I tilfeller hvor den underliggende fordelingen ikke nødvendigvis er normalfordelt har imidlertid en modifisert versjon av Levens test vist seg å

være mer robust (Brown og Forsythe 1974). Testen tar utgangspunkt i at variansen i de to gruppene er lik og er definert ved følgende hypoteser:

$$H_0: \sigma_0^2 = \sigma_1^2 \quad H_A: \sigma_0^2 \neq \sigma_1^2$$

Denne testen vurderer forskjellen fra de enkelte observasjonene innad i utvalgene mot utvalgets median. Testobservatoren blir som følger (Brown og Forsythe 1974):

$$W_0 = \frac{\sum_i n_i (z_i - z_{..})^2 / (g - 1)}{\sum_i \sum_j (z_{ij} - z_i)^2 / \sum_i (n_i - 1)}$$

hvor $z_i = \sum z_{ij} / n_i$ og $z_{..} = \sum \sum z_{ij} / \sum n_i$. $z_{ij} = |x_{ij} - x'_1|$. x_{ij} er observert verdi, mens x'_1 er medianen. $\sum_i (n_i - 1)$ og $(g - 1)$ er antall frihetsgrader, hvor n_i er antall observasjoner i gruppe i , og g er antall grupper.

3.5 Regresjon

3.5.1 Forutsetninger for regresjon

En vesentlig forutsetning for ordinær regresjon er at den avhengige variabelen må være metrisk eller kvantitativ (Tuft 2000). Den må også strengt tatt være kontinuerlig. Dette innebærer at kvalitative og dikotome variabler ikke tilfredsstillt kravene til regresjon ved minste kvadraters metode. For variabler på nominalnivå er det et problem at verdiene ikke kan rangeres på noen måte. For nominale variabler er det kun mulig å sortere variablene i kategorier som «menn og kvinner», «ja og nei» etc. (Tuft 2000). For variabler på ordinalnivå er problemet å måle avstanden mellom de ulike kategoriene. En ordinal variabel er en kategorivariabel som står i forhold til en annen kategorivariabel (Tuft 2000). Problemet med lineær regresjon på en dikotom (avhengig) variabel er at forutsetningen om at feilleddene skal være normalfordelt ikke er tilfredsstillt, samt at modellen kan gi dårlige prediksjoner. Det at modellen er lineær er også et problem i seg selv (Tuft 2000). Rentabilitet på -15 % eller 15 % utgjør en vesentlig forskjell på sannsynligheten for konkurs, men om bedriften har 30 % eller 60 % rentabilitet har, relativt sett, marginalt å si for konkurssannsynligheten.

En metode for å behandle dikotom avhengig variabel er logistisk regresjon (logit – eller probitmetoden) som blir beregnet ved «Maximum Likelihood Estimation», også kalt MLE. Den viktigste forutsetningen for logistisk regresjon er at den avhengige variabelen følger S-kurven (sannsynlighet plottet mot logitverdien, figur 3.1). For øvrig nevner Aldrich og Nelson (1984) fire forutsetninger for logistisk regresjon:

- 1) Den avhengige variabelen kan kun ha to verdier (dikotom variabel), f. eks. 0 eller 1.
- 2) Sannsynligheten for at den avhengige variabelen er lik 1, er avhengig av de uavhengige variablene:

$$P(Y=1) = \frac{1}{1+e^{(b_0+b_1x_1+\dots+b_nx_n)}}$$

- 3) Observasjonene av de uavhengige variablene antas å være uavhengige av hverandre.
- 4) Det må ikke være sterk lineær sammenheng mellom to eller flere av de uavhengige variablene (multikollinearitet).

3.5.2 Logit

Forskjellen på logit og probit ligger i antakelsen om fordelingen til feilleddene. Aldrich og Nelson (1984) skriver at probit-metoden antar normalfordelte feilledd, mens logit-modellen antar at feilleddene er logistisk fordelt, som er omtrent som normalfordelingen, men med flatere haler. Logit og probit vil gi omtrent de samme resultatene. På grunn av dette, samt populariteten i nyere litteratur, er logit-metoden benyttet i denne utredningen.. Probit- og logit-metoden er forøvrig ikke-lineære regresjoner og matematikken bak dette er komplisert. Vi prøver derfor å forklare dette på en enkel måte uten å gjennomgå hele utledningen.

Regresjonsanalyse behandler den avhengige variabelen som om den er kontinuerlig og har ubegrenset utfallsrom. Den dikotome variabelen må derfor regnes om til å bli kontinuerlig og ubegrenset. Den dikotome variabelen kan derfor ses på som en representasjon av underliggende kontinuerlig variabel (Kleinbaum 1994).

Sammenhengen mellom sannsynligheten for et fenomen, Y , og en variabel, X , kan ses på som en S-formet kurve mellom 0 og 1 (se figur 3.1). Sannsynligheten når aldri helt 0 eller 1.

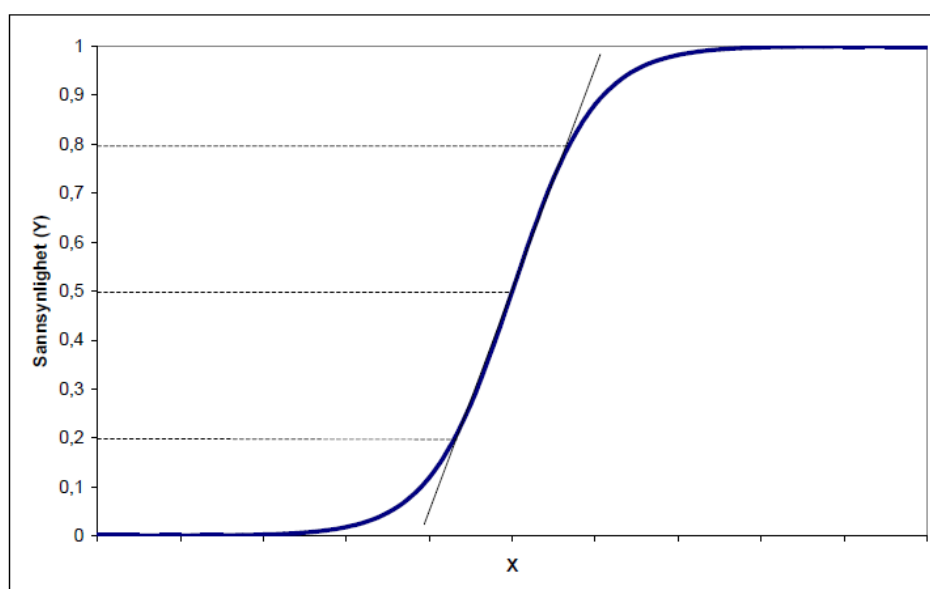
Den avhengige variabelen Y er begrenset til intervallet (0-1). For å fjerne den øvre grensen til Y , gjør man om variabelen til odds. Odds er forholdstallet mellom sannsynligheten for at noe skal inntreffe og sannsynligheten for at det ikke skal inntreffe:

$$\text{Odds} = \frac{p}{1-p}$$

I prinsippet kan oddsen oppnå uendelige verdier, men aldri null eller lavere. For å fjerne den nedre grensen til Y , tar man den naturlige logaritmen til oddsen. Dette uttrykket kalles logiten, L :

$$L = \ln\left(\frac{p}{1-p}\right) = b_0 + b_1X_1 + \dots + b_nX_n + e$$

Logiten viser en lineær sammenheng mellom de uavhengige variablene. Selv om det kan være vanskelig å tolke resultatene direkte fra regresjonen vil det være mulig å se visse sammenhenger uten å regne om til odds-ratioer eller sannsynligheter. Positive fortegn foran koeffisientene viser at det er en positiv sammenheng, mens negativt fortegn viser det motsatte (Tuft 2000). Selv om man tar logaritmen til oddsen endrer ikke det på rekkefølgen på tallene og derfor ikke på om det er positive eller negative sammenhenger. Et større log-odds tall betyr høyere odds tall og følgelig høyere sannsynlighet, gitt at de andre variablene holdes konstant.



Figur 3. 1: S-kurven – sannsynligheter plottet mot logit-verdier (Tuft 2000)

3.5.3 Maximum Likelihood Estimation

Ved ordinær regresjon estimeres regresjonslikningen ved hjelp av minste kvadraters metode. Det er imidlertid teknisk umulig å benytte minste kvadraters metode til å beregne logaritmen av odds (Tuft 2000). Maximum Likelihood Estimation, også kalt MLE, brukes derfor til å estimere regresjonslikningen ved probit- og logit-metoden, eller ved andre ikke-lineære regresjonsmetoder (kan også brukes ved lineær regresjon).

Målet for MLE er å finne en likning basert på ulike variabler som gir den høyeste sannsynligheten for at den avhengige variabelen, Y , faktisk er lik det som er observert. Denne likningen kalles, som nevnt over, logiten. Ved å kjøre flere iterasjoner kan man finne de koeffisientene som gir høyest sannsynlighet for at modellen passer med virkeligheten (Kleinbaum 1994). Modellen, ved hjelp av algoritmer, «prøver» seg frem til den finner en likning som passer best til de faktiske dataene observert. Dette gjøres ved å maksimere den naturlige logaritmen til likelihood-funksjonen, $L(p)$ (beskrevet som «log-likelihood» i utskriften). Det matematiske bak utregningen av dette er komplisert, så vi velger å ikke gå videre inn på dette i denne utredningen.

3.5.4 Signifikanstesting av koeffisientene

For å kunne tolke koeffisientene og benytte seg av tilhørende variabler bør man teste hvorvidt de aktuelle koeffisientene er statistisk signifikante fra null. Dette kan gjøres ved hjelp av ulike metoder, hvorav noen aktuelle gjennomgås nedenfor.

Z-test

Z-verdien fra z-testen viser hvor mange standardavvik den observerte verdien befinner seg i forhold til gjennomsnittet. En Z-test bruker normalfordelingen til å finne ut om den observerte verdien er signifikant forskjellig fra nullhypotesen (Tuft 2000). Nullhypotesen i denne testen er at koeffisienten er lik null. På likningsform blir hypotesene som følger:

$$H_0: b_i = 0$$

$$H_A: b_i \neq 0$$

Med $\alpha = 0,05$, tilsvarer dette 1,96 standardavvik. Hvis Z-verdien er større enn 1,96 forkastes nullhypotesen. Hvis man arbeider med et utvalg hvor fordelingene er ulike, kan det settes spørsmålsteget om z-verdien er et gyldig mål på statistisk signifikans. Likelihoodratio-test kan derfor være bedre for signifikanstesting (Tuftes 2000).

Likelihoodratio-test

Målet ved logistisk regresjon er å maksimere log-likelihood for de gitte variablene. Det som kan være interessant å undersøke er hvorvidt log-likelihood endrer seg når man varierer hvilke variabler som tas med i analysen. Med utgangspunkt i L_0 kan man se om log-likelihood endrer seg signifikant ved å legge til én eller flere variabler, L_1 . Dette kan testes ved å bruke følgende testobservator (Tuftes 2000):

$$G^2 = -2(L_0 - L_1)$$

Testobservatoren er tilnærmet kjikvadratfordelt med frihetsgrader lik forskjellen i antall variabler mellom L_0 og L_1 . Nullhypotesen sier at det er ingen forskjell i den *avhengige* variabelen med hensyn på den inkluderte *uavhengige* variabelen. Hvis endringen i log-likelihood er signifikant, kan man forkaste nullhypotesen om at koeffisienten er lik null (Tuftes 2000). På likningsform blir hypotesene som følger:

$$H_0: b_i = 0$$

$$H_A: b_i \neq 0$$

3.5.5 Tolkning av koeffisientene

Tolkning av koeffisientene ved logistisk regresjon er ikke like rett frem som ved lineær regresjon, nettopp på grunn av at regresjonen ikke er lineær. Man kan eventuelt begrense seg til kun å tolke fortegnene til koeffisientene, som forteller om den enkelte variabelen påvirker sannsynligheten for at den avhengige variabelen er lik én, positivt eller negativt. Altså hvis fortegnet er positivt, vil en økning i den uavhengige variabelen resultere i at sannsynligheten for at den avhengige variabelen er lik én øke. To andre fremgangsmåter å tolke koeffisientene på er beskrevet under.

Odds-ratio

Odds-ratioen er forholdet mellom oddsen ved verdi f (på den avhengige variabelen) og verdi s (Kleinbaum 1994):

$$\theta = \frac{\frac{p_f}{1-p_f}}{\frac{p_s}{1-p_s}}$$

hvor p_s er sannsynligheten for at den *avhengige* variabelen er lik 1 (her: konkurs), ved en viss verdi på den *uavhengige* variabelen, mens p_f er sannsynligheten for at den *avhengige* variabelen er lik 1 ved en høyere verdi på den *uavhengige* variabelen. Hvis den *uavhengige* variabelen stiger med én enhet, viser odds-ratioen hvor mange ganger oddsen endres. Hvis odds-ratioen er høyere enn 1, stiger oddsen når den *uavhengige* variabelen stiger, mens en odds-ratio under 1 forteller at oddsen synker når den *uavhengige* variabelen stiger. Odds-ratioen kan aldri bli negativ. Fordelen med odds-ratio er at verdien er konstant gjennom hele logit-båndet (S-kurven). Ulempen er at den kan være vanskelig å tolke.

Sannsynligheter

Sannsynligheter er lettere å tolke enn odds-ratioer, men endring i sannsynlighet (som følge av endring i en uavhengig variabel) vil endre seg avhengig av hvor man er i logit-båndet (S-kurven). Er man midt i båndet (bratt kurve) vil en liten endring i en uavhengig variabel utgjøre en større forskjell enn om man er nede (slak kurve) i halen (Tuft 2000). Man kan finne et estimat på effekten av endring i en uavhengig variabel ved forskjellige punkter på S-kurven ved å ta koeffisienten og multiplisere denne med sannsynligheten for at den avhengige variabelen er 1 og sannsynligheten for at den er 0 (Tuft 2000):

$$b \times p(1-p)$$

Maksimal effekt av endring på en uavhengig variabel er når man befinner seg midt på S-kurven, hvor kurven er brattest og sannsynligheten for at den avhengige variabelen er 1, er lik 0,5. Det er også vanlig å se på effekten av koeffisienten når sannsynligheten er lik som

marginalfordelingen i datautvalget. Da får man et estimat på endring i sannsynlighet der hovedtyngden av datautvalget ligger (Tuft 2000).

Sannsynligheten for at den avhengige variabelen er lik 1 (her: konkurs) kan regnes ut ved følgende formel (Kleinbaum 1994):

$$P(Y=1) = \frac{1}{1+e^{(b_0+b_1x_1+\dots+b_nx_n)}}$$

3.5.6 Goodness of fit

Målet med å estimere en modell er å kunne gjøre statistisk inferens om en avhengig variabel basert på utvalgte uavhengige variabler. For å vurdere hvorvidt en estimert modell gir signifikante resultater, er én mulighet å teste hvor godt de estimerte verdiene sammenfaller med faktiske verdier. Det finnes ulike metoder for å gjøre dette og de omtales som mål på «goodness of fit». Vanligvis gjøres en vurdering av goodness of fit i to steg; først vurderes summen av feilleddene før de enkelte feilverdiene vurderes (Hosmer et al. 1991). I denne utredningen fokuseres det på metodene; Pearson, Deviance, Brown og Hosmer-Lemeshow. Disse testene utføres automatisk i Minitab ved testing av goodness of fit, men på bakgrunn av at resultatene bør tolkes og vurderes, vil de ulike testene gjennomgås. I denne gjennomgangen vil ikke hele utledningen bak de ulike testene gjennomgås, da matematikken bak er forholdsvis omfattende og således faller utenfor denne utredningens rammer.

Pearson kjikvadrattest og Deviance tar utgangspunkt i kovariatmønstre, mens Brown og Hosmer-Lemeshow tar utgangspunkt i predikerte sannsynligheter. En kort innføring i begrepene kovariatmønster og kovariater vil være hensiktsmessig før de enkelte metodene forklares nærmere.

Kovariatmønster

Et kovariatmønster kan defineres som et sett med verdier for ulike kovariater. En kovariat er en observert kombinasjon av de uavhengige variablene. Dersom to observasjoner har identiske verdier på alle de uavhengige variablene har de samme kovariatmønster (Tuft

2000). I denne utredningen blir kovariatmønstre ulike kombinasjoner av nøkkeltallene for konkurs- og ikke-konkursbedrifter. Det er i hovedsak to ulike typer slike mønstre, type 1- og type 2-mønstre. Type 1-mønstre finner sted når alle observasjonene (bedriftene i denne utredningen) har unike kovariater. Det betyr at antall kovariater (J) er likt antall observasjoner (n), $J = n$. Dette er vanlig når de uavhengige variablene er kontinuerlige variabler (Liu 2007). Type 2-mønstre defineres ved at noen av observasjonene har et sammenfallende mønster av de uavhengige variablene. Sistnevnte situasjon betyr at antall unike kovariater er mindre enn antall observasjoner, $J < n$.

Pearson kjikvadrat-test

Pearson kjikvadrat-test tester den estimerte modellen mot en optimal modell. Det vil si en modell som har like mange parametere som observasjoner. Modellen er basert på interpolering (rette linjer trekkes mellom de kjente verdiene) av dataene, noe som betyr at modellen passer dataene perfekt. Ved tilfeller hvor antall kovariater tilsvarer antall observasjoner er testobservatoren som følger (Hallett 1999):

$$\chi^2 = \sum_{i=1}^n \frac{(y_i - \pi(x_i))^2}{(x_i)(1 - \pi_i)} = n$$

y_i er verdien på den avhengige variabelen, x_i er antall bedrifter, mens $\pi(x_i) = y/m$. m er antall observasjoner innenfor en kovariat J . For tilfeller av type 1-mønstre, hvor $J=n$, er ikke Pearson kjikvadrat-test en hensiktsmessig test på hvor godt modellen estimerer. Grunnen til dette er at utvalgsstørrelsen (n) ikke er en tilfredsstillende testobservator (Hallett 1999).

Deviance

Ved å gjennomføre en Deviance-test, måler man forskjellen mellom log-likelihood til den estimerte modellen, og log-likelihood til den optimale modellen (observerte verdier). Ved type 1-kovariatmønstre, som er tilfelle i denne utredningens datautvalg, blir testobservatoren som følger (Hallett 1999):

$$D = -2 \sum_{i=1}^n \{\pi(x_i) \log \pi(x_i) + (1 - \pi(x_i)) \log (1 - \pi(x_i))\}$$

$\pi(x_i)$ og x_i betegner det samme som ved Pearson kjikvadrattest. I tillegg tar man logaritmen av disse. Ved type 1-kovariatmønstre ender testobservatoren altså opp med å ikke forklare forskjellen mellom de estimerte og observerte verdiene, men kun bruke de estimerte verdiene. Det betyr at Deviance-testen er ikke en god test på goodness of fit når $J = n$ (Hallett 1999).

Brown's score-test

Denne metoden sammenligner resultatene fra to estimerte modeller; henholdsvis en logistisk modell mot en spesifisert generell parametriske modell (bygger på normalfordelingen). Nettopp på grunn av at metoden tester hvorvidt en logistisk modell passer en modell blant de generelle parametriske modellene, og ikke tester mot alle mulige modeller, hevder le Cessie and van Houwelingen (1991) at den ikke er egnet som en generell goodness of fit-test. Testobservatoren er som følger (Hallett 1999):

$$B = s' C^{-1} s$$

der C er en estimert kovariansmatrise til s , og $s' = (s_1, s_2)$. s' er basisen i denne testen, og er en «score statistic». Matematisk vil det si at s' er den partiellderiverte av log-likelihood-funksjonen, med hensyn på de ulike modellene det testes mot.

Hosmer-Lemeshow

Hosmer-Lemeshow-testen grupperer sannsynlighetene fra den estimerte modellen i opptil ti grupper (g) avhengig av totalt antall observasjoner (n). Den første gruppen inneholder $n_1 = n/10$ observasjoner med de lavest estimerte sannsynlighetene fra modellen. Den andre gruppen inneholder $n_2 = n/10$ observasjoner med nest minst sannsynlighet etc. π_k representerer gjennomsnittlig estimert sannsynlighet i gruppe k for $y = 1$ (konkurs i denne utredningen), o_k betegner det faktiske antall observasjoner hvor $y = 1$ innenfor gruppe k . Antall frihetsgrader i Hosmer-Lemeshow-testen er $g-2$. Testobservatoren er som følger (Hosmer et al. 1991):

$$\hat{C} = \sum_{k=1}^g \frac{(o_k - o_k \pi_k)^2}{n_k \pi_k (1 - \pi_k)}$$

Metoden tester observert frekvens av en gitt tilstand (konkurs i dette tilfellet) innad i hver gruppe mot beregnet forventet frekvens i en kjikvadrattest. Dersom den estimerte modellen predikerer tilfredsstillende, vil kjikvadrattesten ikke være signifikant. Med andre ord vil p-verdien fra testen være høy i forhold til et gitt signifikansnivå. Hypotesene som testes i Hosmer-Lemeshow-testen er som følger:

H₀ = Forventet og observert fordeling i gruppene er like

H_A = Forventet og observert fordeling i gruppene er ulike

I tilfeller med type 1-kovariatmønstre kan det konkluderes med at det kun er Hosmer-Lemeshow-testen (blant testene nevnt over) som egner seg til test av goodness of fit.

3.6 Korrigerings av utvalgsskjevhet

Konkursprediksjonsmodeller blir typisk estimert fra ikke-tilfeldige utvalg, det vil si at bedriftene som inngår i studien ikke er tilfeldig trukket ut fra den totale populasjonen. Zmijewski (1984) beskriver to problemer dette kan føre til; «choice-based sample bias» og «sample selection bias». Førstnevnte problem beskriver en situasjon hvor antall konkursbedrifter er overrepresentert i utvalget sammenlignet med virkeligheten. Dette er et problem fordi enkelte metoder, som logit, forutsetter et representativt forhold mellom konkurs- og ikke-konkursbedrifter. Det andre problemet finner sted når det kun inkluderes bedrifter i utvalget som har fullstendig tilgjengelig informasjon. Spesielt når det gjelder konkursbedrifter, kan dette føre til at en del bedrifter utelates. Zmijewski (1984) beskriver videre at bakgrunnen for at disse problemene ofte oppstår er fordi man som forsker møter på ulike begrensninger og kostnader i forbindelse med datainnsamlingen. En av begrensningene vi har møtt på i denne utredningen, er det faktum at det er relativt få bedrifter som går konkurs hvert år. Da denne utredningen i tillegg møter på begrensninger i forhold til omfang og tid, er antall konkursbedrifter overrepresentert i forhold til virkeligheten. Dersom det ikke korrigeres for denne skjevheten, vil man kunne få liknende resultater som Zmijewski (1984).

Studien finner at probit-modellen, estimert på bakgrunn av et skjevt utvalg, overestimerer antall konkurs.

Skogsvik (2005) utleder en metode for korrigerings av utvalgsskjevhet ved logistisk regresjon. Denne metoden tar utgangspunkt i det estimerte resultatet fra probit-modellen i dette tilfellet, for så å korrigere denne. Det betyr at selve modellen ikke trenger å reestimeres. Dette kan være praktisk dersom utvalget modellen baserer seg på endrer seg over tid. Man kan da få representative resultater uten å måtte ha tilgang til hele datasettet (Skogsvik 2005).

Gjennom anvendelse av Bayes lov kommer Skogsvik (2005) frem til følgende modell for korrigerings av utvalgsskjevhet:

$$p = \left[1 + \frac{(1-\pi)}{\pi} * \frac{prop}{(1-prop)} * \frac{1-p^{prop}}{p^{prop}} \right]^{-1}$$

hvor p er korrigert sannsynlighet for konkurs. π representerer andelen konkursbedrifter i hele populasjonen, $prop$ er andelen konkursbedrifter i utvalget, mens p^{prop} er sannsynligheten predikert med den estimerte modellen.

4. Analyse

I denne delen vil først ulike aspekter rundt datainnsamling og -utvalg gjennomgås. Den gjennomførte komponentanalysen og påfølgende utvelgelse av nøkkeltall vil så kommenteres. Avslutningsvis vil de endelige estimerte modellene og tilhørende tolkninger gjennomgås, samt at modellenes prediksjonsevne testes.

4.1 Datautvalg

4.1.1 Innsamling av data

Datamaterialet i denne utredningen består av offentlig tilgjengelig regnskapsinformasjon. Dataene er lastet ned fra databasen RavnInfo.no hvor regnskapsinformasjonen til alle norske selskaper registrert i Brønnøysundregistret er tilgjengelig for flere år tilbake.

Bedriftene velges ut etter totalt seks kriterier. Kriteriene er satt med den hensikt at de innsamlede regnskapene i størst mulig grad skal være sammenlignbare. I tillegg krever utredningens omfang en viss avgrensning. Utvalget av konkursbedrifter består utelukkende av selskaper der det er åpnet konkurs.

Datautvalget er satt sammen på bakgrunn av følgende kriterier:

- 1) Bedriftene må benytte aksjeselskap (AS) som selskapsform (ikke ASA).
- 2) Totale balanseførte eiendeler må være minimum 15 millioner kroner.
- 3) Totale inntekter må minimum være 30 millioner kroner.
- 4) Antall ansatte må minimum være 10.
- 5) Konkursbedriftene må ha gått konkurs i perioden 01.01.2008 – 31.12.2011.
- 6) Konkursbedriftene må ha levert regnskap siste fire år før konkurs, mens ikke-konkursbedriftene må ha levert i hele den aktuelle perioden, det vil si fra 2004 til og med 2010.

Enkelte av konkursbedriftene har hatt noe svingninger hvor størrelsen på inntekter og eiendeler, samt antall ansatte har vært i underkant av grensene. På grunn av at det ikke benyttes data fra alle bedriftene for hvert år i estimeringen av modellene har disse bedriftene likevel blitt inkludert i utvalget i de tilfellene hvor svingningene har vært små. Dette har blitt gjort på grunn av begrensninger i tilgangen på data for konkursbedriftene.

4.1.2 Størrelse på datautvalget

I perioden 2008 – 2011 har andelen AS som har gått konkurs i forhold til totalt antall AS i Norge variert fra 1,06 % i 2008, til 1,51 % i 2009 (1: ssb.no, 3, 4, 5, 6: Pressemelding Brønnøysundregisteret, og Torgnesvik 2012) (se tabell 4.1). Gjennomsnittlig for hele perioden har 1,25 % av totalt antall AS gått konkurs årlig.

År	Antall konkurs	Antall AS	Andel
2008	2230	211439	1,05 %
2009	3214	213224	1,51 %
2010	2647	215076	1,23 %
2011	2638	219977	1,20 %
<i>Gj.snitt</i>			1,25 %

Tabell 4. 1: Antall AS konkurs årlig i forhold til totalt antall AS

Datautvalget som analysene i denne utredningen er basert på, består av regnskapsdata fra 43 konkursbedrifter, og 847 ikke-konkursbedrifter. For å få lik andel som i virkeligheten ville det krevd et totalt antall bedrifter på rundt 40 000, gitt antallet konkursbedrifter. Av hensyn til utredningens omfang brukes det et mindre datautvalg, slik at antall konkurser i utvalget overstiger det faktiske tall for perioden. Antall konkursbedrifter kunne vært redusert, slik at andelen i utvalget ville tilsvart andelen i virkeligheten. Det ble imidlertid vurdert dit hen at det kreves et visst antall for å omfavne flere bransjer blant konkursbedriftene, og på den måten få et mer representativt bilde av virkeligheten. De endelige resultatene vil korrigeres for denne utvalgsskjevheten.

Bedriftene i utvalget har ikke blitt selektert spesifikt etter bransje, bortsett fra bedrifter innenfor finans- og forsikringsbransjen som er utelatt fra utvalget. Dette er gjort da man har antatt at disse har en annen finansiell struktur (jf. Ohlson 1980 og Sjøvoll 1999) som igjen

ville påvirket nøkkeltallene. Tabell 4.2 og tabell 4.3 viser fordelingen av bedriftene i hovednæringsgrupper for henholdsvis gruppen av ikke-konkurs- og konkursbedriftene.

Næringsområde ikke-konkursbedrifter	Kode	Antall	Andel av utvalg
Jordbruk, skogbruk og fiske	01-03	8	0,9 %
Bergverksdrift og utvinning	05-09	31	3,7 %
Industri	10-33	294	34,7 %
Elektrisitets-, gass-, damp- og varmtvannsforsyning	35	10	1,2 %
Vannforsyning, avløps-, og renovasjonsvirksomhet	36-39	9	1,1 %
Bygge- og anleggsvirksomhet	41-43	86	10,2 %
Varehandel	45-47	245	28,9 %
Transport og lagring	49-53	43	5,1 %
Overnatting og serveringsvirksomhet	55-56	15	1,8 %
Informasjon og kommunikasjon	58-63	50	5,9 %
Omsetning og drift av fast eiendeom	68	5	0,6 %
Faglig, vitenskapelig og teknisk tjenesteyting	69-75	16	1,9 %
Forretningsmessig tjenesteyting	77-82	23	2,7 %
Helse- og sosialtjenester	86-88	6	0,7 %
Kulturell virksomhet, underholdning og fritidsaktiviteter	90-93	3	0,4 %
Annen tjenesteyting	94-96	3	0,4 %
Sum		847	100,0 %

Tabell 4. 2: Oversikt over ikke-konkursbedrifter i datautvalget

Næringsområde konkursbedrifter	Kode	Antall	Andel av utvalg
Industri	10-33	16	37,2 %
Bygge- og anleggsvirksomhet	41-43	11	25,6 %
Varehandel	45-47	11	25,6 %
Transport og lagring	49-53	3	7,0 %
Overnatting og serveringsvirksomhet	55-56	1	2,3 %
Informasjon og kommunikasjon	58-63	1	2,3 %
Sum		43	100,0 %

Tabell 4. 3: Oversikt over konkursbedrifter i datautvalget

Utvalget av konkursbedrifter er noe ujevnt fordelt over de ulike årene i perioden som ligger til grunn (se tabell 4.4). En ujevn fordeling kan ha påvirkning på den endelige prediksjonsmodellen på grunn av ulike makrofaktorer. Dette vil imidlertid bli tatt hensyn til, noe som forklares nærmere under punktet om logit-analyse.

År	Antall konkurs	Andel av utvalg
2011	18	41,9 %
2010	11	25,6 %
2009	6	14,0 %
2008	8	18,6 %
Sum	43	100,0 %

Tabell 4. 4: Fordeling av konkursbedrifter i datautvalget pr år

4.2 Valg av uavhengige variabler

Hvilke variabler som inkluderes i modellen antas å være av stor betydning, som er bakgrunnen for at en relativt stor primærmengde av nøkkeltall er etablert. Med primærmengde menes den initiale mengden av nøkkeltall. Utgangspunktet for primærmengden av nøkkeltall er de som er definert av Skogsvik (1988). Videre har det blitt foretatt en vurdering fra forfatterens side om forventet relevans for denne studien, slik at det ikke har vært avgjørende at variabelen har vært benyttet av Skogsvik (1988) eller i andre studier tidligere. I tillegg er et nøkkeltall som ikke er benyttet i studier tidligere inkludert.

4.2.1 Operasjonalisering av regnskapene

Beregningen av de ulike nøkkeltallene er gjort med utgangspunkt i bedriftenes resultat- og balanseregnskap. Alle regnskap i databasen RavnInfo.no følger samme oppsett slik at det har ikke vært nødvendig å gjøre endringer i selve oppsettet. Tabell 4.5 viser hvordan resultatregnskapet er satt opp og hvilke symboler som er benyttet på de ulike postene i nøkkeltallene.

Inntekter		Ordinært resultat før skattekostnad	ORFS
Salgsinntekt	SI(1)	Skattekostnad på ordinært resultat	SK
Annen driftsinntekt	SI(2)	Ordinært resultat etter skattekostnad	ORES
Totale inntekter	TI	Ekstraordinær inntekt	EOI
		Ekstraordinær kostnad	EOK
		Skattekostnad på ekstraordinært resultat	EOSK
Kostnader	K	Resultat av ekstraordinære poster	
Endring i beholdning av varer under tilvirkning og ferdig tilvirkede varer	EB(1)	Årsresultat før minoritetsinteresser	R(3)
Endring i beholdning av egentilvirkede anleggsmidler	EB(2)	Minoritetsinteresser	MI
Varekostnad	K(1)	Årsresultat	ÅRES
Lønnskostnad	K(2)		
Driftsresultat før av- og nedskrivninger	R(1)		
Avskrivning på varige driftsmidler og immaterielle eiendeler	K(3)	Overføringer og disponeringer	
Nedskrivning av varige driftsmidler og immaterielle eiendeler	K(4)	Overføring til/fra fond for vurderingsforskjeller	OD(1)
Annen driftskostnad	K(5)	Overføring til/fra fond for urealiserte gevinster	OD(2)
Driftsresultat etter av- og nedskrivninger	R(2)	Utbytte	U
		Konsernbidrag	KB
		Aksjonærbidrag	AB
Finansinntekter	FI	Fondsemisjon	FE
Inntekt på investering i datterselskap	FI(1)	Udekket tap	UT
Inntekt på investering i annet foretak i samme konsern	FI(2)	Overføringer til/fra annen egenkapital	OD(3)
Inntekt på investering i tilknyttet selskap	FI(3)	Sum overføringer og disponeringer	R(6)
Renteinntekt fra foretak i samme konsern	FI(4)		
Annen renteinntekt	FI(5)		
Annen finansinntekt	FI(6)		
Verdipågang av markedsbaserte finansielle omløpsmidler	FI(7)		
Verdipågang andre finansielle instrumenter vurdert til virkelig verdi	FI(8)		
Finanskostnader	FK		
Verdireduksjon av markedsbaserte finansielle omløpsmidler	FK(1)		
Verdireduksjon andre finansielle instrumenter vurdert til virkelig verdi	FK(2)		
Nedskrivning av andre finansielle omløpsmidler	FK(3)		
Nedskrivning av finansielle anleggsmidler	FK(4)		
Rentekostnad til foretak i samme konsern	FK(5)		
Annen rentekostnad	FK(6)		
Annen finanskostnad	FK(7)		

Tabell 4. 5: Resultatregnskap med symboler

På samme måte viser tabell 4.6 balanseregnskapet. For å få datene mest mulig sammenlignbare ble det imidlertid gjort enkelte mindre tilpasninger av de nedlastede regnskapene før nøkkeltallene ble beregnet. Noen bedrifter bruker negativt fortegn på kostnadene, mens andre ikke gjør det. Dataene ble derfor kontrollert og eventuelt endret slik at bruken av fortegn er konsistent - det vil si at alle kostnader i denne utredningen er positive størrelser. Det finnes riktignok unntak, eksempelvis skattekostnad, som kan være både en positiv og negativ størrelse. I enkelte tilfeller ble det også oppdaget at summeringen av kostnadene og enkelte balanseposter ikke stemte overens med de ulike enkeltpostene. Det ble i disse tilfellene tatt for gitt at de enkelte tallene var korrekte slik at sum-postene ble endret.

Balanse			
Eiendeler	TK	EK/Gjeld	
Anleggsmidler	AM	Egenkapital	EK
Immaterielle eiendeler		Innskutt egenkapital	
Forskning og utvikling	FoU	Selskapskapital	SEK
Konsesjoner, patenter, lisenser, varemerker og lignende rettigheter	VM	Beholdning av egne aksjer	BEA
Utsatt skattefordel	USKF	Overkursfond	OK
Goodwill	GW	Annen innskutt egenkapital	AIEK
Andre immaterielle eiendeler	AIM		
		Opptjent egenkapital	OEK
Varige driftsmidler	VDM	Fond for vurderingsforskjeller	FVF
Tomter, bygninger og annen fast eiendom	FEI	Fond for urealiserte gevinster	FUG
Investerings eiendom	IEI	Annen Egenkapital	AEK
Maskiner og anlegg	MA	Udekket tap	UT
Skip, rigger, fly og lignende	SKF	Minoritetsinteresser	MI
Driftsløsøre, inventar, verkøytøy, kontormaskiner og lignende	DL		
		Langsiktig gjeld	LGJ
Finansielle anleggsmidler	FA	Pensjonsforpliktelser	PF
Investering i datterselskap	ID	Utsatt skatt	USK
Investering i annet foretak i samme konsern	IAFK	Andre avsetninger for forpliktelser	AFF
Lån til foretak i samme konsern	LFIK		
Investeringer i tilknyttet selskap	ITS	Annen langsiktig gjeld	ALGJ
Lån til tilknyttet selskap og felles kontrollert virksomhet	LTK	Konvertible lån	KVL(1)
Investeringer i aksjer og andeler	IA	Obligasjonslån	OBL
Obligasjoner og andre fordringer	OB	Gjeld til kredittinstitusjoner	GJK(1)
		Langsiktig konserngjeld	LKGJ
Omløpsmidler	OM	Ansvarlig lånekapital	ALK
Varer	VA	Øvrig langsiktig gjeld	ØLGJ
Varer	VA		
Biologiske eiendeler	BEI	Kortsiktig gjeld	KGJ
Sum varer	TVA	Konvertible lån	KVL(2)
Fordringer	FOR	Sertifikatlån	SL
Kundefordringer	KFOR	Gjeld til kredittinstitusjoner	GJK(2)
Andre fordringer	AFOR	Leverandørgjeld	LEV
Konsernfordringer	KOFOR	Betalbar skatt	BSK
Krav på innbetaling av selskapskapital	KRAV	Skyldige offentlige avgifter	SKOF
		Utbytte	UT
Investeringer		Kortsiktig konserngjeld	KKGJ
Aksjer og andeler i foretak i samme konsern	AFK	Annen kortsiktig gjeld	AKGJ
Markedsbaserte aksjer	MK		
Markedsbaserte obligasjoner	MO		
Andre markedsbaserte finansielle instrumenter	AMF		
Andre finansielle instrumenter	AF		
Sum investeringer	TI		
Bankinnskudd, kontanter og lignende	BK		

Tabell 4. 6: Balanseregnskap med symboler

4.2.2 Primærmengde av nøkkeltall

Primærmengden av nøkkeltall består av i alt 57 ulike nøkkeltall og er bestemt med utgangspunkt i de sju ulike kategoriene fra Skogsvik (1988). Nøkkeltallene er delt inn i kategoriene; lønnsomhet, kostnader, kapitalens omløpshastighet, likviditet, eiendelsstruktur, kapitalstruktur og vekst. Innenfor disse kategoriene er det definert ulike mål på totalrentabilitet, egenkapitalrentabilitet, andel rente- og skattekostnad, omløpshastighet på ulike eiendelskategorier, ulike mål på bedriftenes likviditet, strukturen på kapitalen og eiendelene, samt ulike mål på vekst.

Tabell 4.7 viser en detaljert beskrivelse av de ulike nøkkeltallene som inngår i primærmengden. I de nøkkeltallene hvor tall fra balansen inngår, er det benyttet et snitt for perioden. Det vil si et aritmetisk gjennomsnitt av inngående og utgående balanse på den aktuelle balanseposten. Nøkkeltall 20, skyldige offentlige avgifter/årsresutatet, er så vidt oss bekjent ikke benyttet i tidligere konkursprediksjonsstudier.

Kategori	Symbol	Definisjon	Likviditet		
Lønnsomhet					
1 Fortjenestemargin	FM	R(2)/TI	32 Likviditetsgrad 1 (1)	LG ₁ (1)	BK/TK
2 Totalkapitalrentabilitet (1)	TKR(1)	[ÅRES+FK*0,72]/TK	33 Likviditetsgrad 1 (2)	LG ₁ (2)	BK/KGJ
3 Totalkapitalrentabilitet (2)	TKR(2)	[ÅRES+SK+EOSK+FK]/TK	34 Likviditetsgrad 2 (1)	LG ₂ (1)	[OM-VA]/TK
4 Totalkapitalrentabilitet (3)	TKR(3)	[ÅRES+FK*0,72]/[TK-KGJ+SL+GJK(2)]	35 Likviditetsgrad 2 (2)	LG ₂ (2)	[OM-VA]/KGJ
5 Totalkapitalrentabilitet (4)	TKR(4)	[ÅRES+FK*0,72]/[TK-KGJ]	36 Likviditetsgrad 3 (1)	LG ₃ (1)	OM/KGJ
6 Totalkapitalrentabilitet (5)	TKR(5)	[ORES+FK*0,72]/TK	37 Kontantstrømsgrad (1)	KSG(1)	ORFS+K(3)+K(4)+FK+(EOI-EOK)/KGJ
7 Totalkapitalrentabilitet (6)	TKR(6)	[ORFS+FK]/TK	38 Kontantstrømsgrad (2)	KSG(2)	ORFS+K(3)+K(4)+FK+(EOI-EOK)/[FK(5)+FK(6)+FK(7)]
8 Totalkapitalrentabilitet (7)	TKR(7)	[ORFS+FK]/[TK-KGJ]	39 Kontantstrømsgrad (3)	KSG(3)	ORFS+K(3)+K(4)+FK+(EOI-EOK)/[KGJ+LGJ]
9 Totalkapitalrentabilitet (8)	TKR(8)	[ORFS+FK+K(3)+K(4)]/[TK-KGJ]			
10 Egenkapitalrentabilitet (1)	EKR(1)	[ÅRES+SK+EOSK]/EK	Eiendelstruktur		
11 Egenkapitalrentabilitet (2)	EKR(2)	ÅRES/EK	40 Andel varelager	AVA	TVA/TK
12 Egenkapitalrentabilitet (3)	EKR(3)	ORFS/EK	41 Andel arbeidskapital	AAK	[OM-KGJ]/TK
13 Egenkapitalrentabilitet (4)	EKR(4)	ORES/EK	42 Andel anleggsmidler	AAM	AM/TK
			43 Andel varige driftsmidler	AVDM	VDM/TK
			44 Størrelse (1)	ST(1)	ln(TK)
			45 Størrelse (2)	ST(2)	ln(TK-KGJ)
Kostnader			Kapitalstruktur		
14 Lønnsandel	LA	K(2)/K	46 Soliditet	SO	EK/TK
15 Avskrivningsandel	AA	K(3)/K	47 Andel kortsiktig gjeld	AKGJ	KGJ/TK
16 Rentekostnad(1)	RK(1)	[FK(5)+FK(6)+FK(7)]/[LGJ+KGJ]	48 Andel leverandørgjeld	ALEV	LEV/TK
17 Rentekostnad(2)	RK(2)	[FK(5)+FK(6)+FK(7)]/[LGJ+GJK(2)+SL]	49 Andel rentebærende gjeld	ARGJ	[LGJ+GJK(2)+SL]/TK
18 Skatteandel	SK	SK/ÅRES			
19 Aksjonær-/konsernbidragsandel	ABA	AB+KB/K	Vekst		
20 Skyldige offentlige avgifter	SKOF	SKOF/ÅRES	50 Vekst omsetning	VTI	ΔTI/TI _{t-1}
			51 Vekst eiendeler	VEI	ΔTK/TK _{t-1}
Kapitalens omløpshastighet			52 Vekst egenkapital	VEK	ΔEK/EK _{t-1}
21 Omløpshastighet eiendeler (1)	OHE(1)	TI/TK	53 Vekst kortsiktig gjeld	VKGJ	ΔKGJ/KGJ _{t-1}
22 Omløpshastighet eiendeler (2)	OHE(2)	TI/[TK-KGJ+(GJK(2)+SL)]	54 Vekst leverandørgjeld	VLEV	ΔLEV/LEV _{t-1}
23 Omløpshastighet eiendeler (3)	OHE(3)	TI/[TK-KGJ]	55 Vekst rentebærende gjeld	VRGJ	Δ[LGJ+GJK(2)+SL]/[LGJ+GJK(2)+SL] _{t-1}
24 Omløpshastighet kundefordringer	OHKFO	[FOR+KFOR+AFOR+KOFOR]/TI	56 Vekst anleggsmidler	VAM	ΔAM/AM _{t-1}
25 Omløpshastighet varelager (1)	OHV(1)	TVA/TI	57 Vekst varige driftsmidler	VVDM	ΔVDM/VDM _{t-1}
26 Omløpshastighet varelager (2)	OHV(2)	TVA/K			
27 Omløpshastighet kontanter	OHK	BK/TI			
28 Omløpshastighet likvide omløpsmidler	OHLO	[OM-VA]/TI			
29 Omløpshastighet omløpsmidler	OHOM	OM/TI			
30 Omløpshastighet arbeidskapital	OHAK	[OM-KGJ]/TI			
31 Størrelse arbeidskapital	STAK	[OM-KGJ]/TK			

Tabell 4. 7: Primærmengde nøkkeltall

4.2.3 Ekstremverdier

I flere tidligere studier setter forfatterne grenser for hvilke verdier nøkkeltallene kan ha. Skogsvik (1988) setter en minimumsbegrensning på tallet i nevneren til de enkelte nøkkeltallene, samt en maksimums- og/eller minimumsverdi på mange. En mulig forklaring på at observasjoner med verdier som klart skiller seg ut forekommer, er feil i dataene som benyttes.

Beregning av frekvensfordelinger for de ulike nøkkeltallene i denne utredningen viser at det forekommer enkelte ekstremverdier. For å minimere betydningen av mulige feil i dataene, ble disse korrigert. Ved å se på frekvensplott for hvert enkelt nøkkeltall ble hensiktsmessige grenseverdier satt. I de tilfellene hvor verdien på nøkkeltallet overskred den satte grenseverdien, ble verdien erstattet med grenseverdien.

Eksempelvis ble grenseverdiene til nøkkeltallet KSG(2) satt fra -100 til 500. Dette gjorde at totalt 167 observasjoner ble rettet til en av disse verdiene, noe som utgjør 3,26 % av antall observasjoner. Dette var det nøkkeltallet hvor det ble foretatt flest korrigeringer. Tabellen i appendiks 7.1 viser oversikten over de ulike grenseverdiene, samt hvor stor andel av nøkkeltallene som ble korrigert.

4.2.4 Komponentanalyse

Komponentanalysen ble foretatt i Minitab på et datautvalg bestående av nøkkeltall for totalt 5211 observasjoner. Ikke-konkursbedriftene bidro med data for hele analyseperioden (seks år), mens konkursbedriftene bidro med nøkkeltall for de tre siste årene før konkurs.

Det ble etablert totalt 57 komponenter i komponentanalysen - det samme som antall uavhengige variabler. Av disse hadde 14 egenverdi over 1, som igjen ble inkludert i videre analyse og utvelgelse av nøkkeltall. Disse 14 komponentene utgjør til sammen 77,24 % av den totale variansen, hvor den første komponenten er klart viktigst med sin andel av total varians på 20,10 %.

Tolkning av komponentene

For å kunne benytte resultatene fra komponentanalysen til å redusere antall nøkkeltall, bør de ulike komponentene tolkes. Som nevnt tidligere er denne tolkningen til dels en subjektiv vurdering. For noen komponenter er det forholdsvis enkelt å tolke hva de representerer, mens andre er noe mer uklare.

I første omgang ble egenvektorene til de ulike nøkkeltallene innad i komponentene vurdert mot hverandre. Flertallet av komponentene inneholder både positive og negative

egenvektorer, og i disse tilfellene ble de ulike gruppene av verdier vurdert opp mot hverandre. Tabell 4.8 viser de ulike komponentene og tolkningen av dem, samt noen av de karakteristiske nøkkeltallene tolkningen bygger på. Med karakteristiske nøkkeltall menes de med høyest egenvektor innad i komponentene, som således har størst betydning. Appendiks 7.2 viser en større oversikt over de mest innflytelsesrike nøkkeltallene og deres respektive egenvektorer innad i komponentene.

Komponent	Navn (tolkning)	Karakteristiske nøkkeltall
PC1	Rentabilitet	TKR(7), TKR(4), TKR(3), TKR(6), EKR(2), TKR(5), TKR(1), EKR(4)
PC2	Kortsiktig finansiering	KSG(1), ST(2), SO, KSG(3), OHE(3), OHE(1), ALEV, AKJG
PC3	Driftskapital	AAK, STAK, LG3(1), LG1(1), AA, AVDM, AAM
PC4	Omløpshastighet	OHK, OHKFO, OHLO, OHO
PC5	Varelager	LG1(1), LG1(2), LA, AVA, OHV(1), OHV(2)
PC6	Vekst	VRGJ, VTI, VLEV, VVDM, VKJG, VAM, VEI
PC7	Kapitalstruktur	ARGJ, EKR(4), OHE(1), OHE(3), OHE(2), ST(2), ST(1), SO, RK(1)
PC8	Aktivitetsnivå	KSG(2), OHAK, LG3(1), LG2(2), ARGJ, RK(1), OHE(2), OHE(3)
PC9	Langsiktig finansiering	RK(1), RK(2), OHV(2), LG1(2), OHV(1), OHK, KSG(2), LG1(1), OHE(3), OHE(2)
PC10	Rentekostnader	ST(2), ST(1), OHV(2), OHV(1), OHK, LG1(2), LG1(1), RK(1), RK(2)
PC11	Størrelse	VEK, KSG(1), SO, VRGJ, ST(2), ST(1)
PC12	Offentlige avgifter	SK, SKOF, ABA, VKGJ
PC13	Vekst kortsiktig finansiering	VLEV, VKGJ, SKOF, VVDM, VEK, VRGJ, VAM
PC14	Kostnader	ALEV, SO, SKOF, LA

Tabell 4. 8: Komponenter med egenverdi over 1

Komponent 1 gav tydelige utslag på alle rentabilitetsnøkkeltallene, både total- og egenkapitalrentabilitet, hvor de høyeste egenvektorene ble representert ved ulike totalrentabilitetstall. Denne komponenten ble derfor tolket til å forklare rentabilitet. *Komponent 2* inneholder i hovedsak ulike nøkkeltall for kontantstrømgrad og kortsiktig finansiering i form av andel leverandørgjeld og kortsiktig gjeld. I kontantstrømgrad inngår forøvrig kortsiktig gjeld som nevner. Komponentene gav utslag både på pluss- og minussiden, noe som indikerer at komponenten ser på kontrasten mellom ulike nøkkeltall. Nøkkeltallene for kortsiktig finansiering ble ansett som de viktigste ut ifra egenvektorene, slik at komponenten som helhet ble tolket til å representere kortsiktig finansiering.

Komponent 3 ble tolket som mål på driftskapital, da andel arbeidskapital og andel anleggsmidler gjorde særlig utslag med forskjellig fortegn. *Komponent 4* ble tolket som mål på omløpshastighet, da de nøkkeltallene som gjorde særlig utslag med tanke på egenvektorene er ulike mål på omløpshastighet. Alle disse nøkkeltallene har negative egenvektorer, noe som gjør komponenten enklere å tolke.

Komponent 5 ble tolket som mål på varelager, da AVA, OHV(1) og OHV(2) gjorde særlig utslag. I *komponent 6* gjorde ulike nøkkeltall for vekst særlig utslag, alle med samme fortegn. Derfor ble denne komponenten tolket som mål på vekst. I *komponent 7* gjorde ARGJ særlig utslag på den positive siden, mens SO og RK(2) gjorde særlig utslag på den negative siden. Det betyr, sett bort i fra foregående komponenter, at den største delen av variasjonen er blant de bedriftene som har stor rentebærende gjeld i forhold til soliditet og rentekostnader. Komponenten ble derfor tolket til å være et mål på kapitalstruktur. Blant de innflytelsesrike nøkkeltallene i *komponent 8* var det kun KSG(2) som hadde positiv egenvektor slik at nøkkeltallet i utgangspunktet ikke så ut til å være særlig viktig. Dersom man ser på absoluttverdien av egenvektoren til KSG(2) i forhold til de andre, ser man imidlertid at den er forholdsvis betydningsfull. Det er imidlertid OHE(1) og OHE(2) som har størst innflytelse, slik at komponenten ble tolket som et mål på aktivitetsnivå i bedriftene.

Komponent 9 har nøkkeltall med egenvektorer på både positiv og negativ side, slik at også her måtte de ulike nøkkeltallene vurderes opp mot hverandre. På den positive siden gjorde RK(1) og RK(2) særlig utslag, mot OHE(2) og OHE(3) på den negative siden. Samlet sett ble komponenten tolket som mål på langsiktig finansiering. RK(1) og RK(2) gjorde også særlig utslag i *komponent 10*, men nå på den negative siden. På den positive siden her gjorde ST(1) og ST(2) utslag, men samlet sett ble den negative siden vurdert som mest betydningsfull, slik at komponenten ble tolket som mål på rentekostnader. I *komponent 11* gjorde også ST(1) og ST(2) særlig utslag, men nå på den negative siden. Disse to ble vurdert opp mot VEK, som gjorde særlig utslag på den positive siden. Totalt sett ble komponenten vurdert som et mål på størrelse.

Komponent 12 var forholdsvis enkel å tolke, da det var to nøkkeltall som hadde egenvektorer med samme fortegn som gav særlig utslag; SK og SKOF. Denne komponenten ble tolket som et mål på offentlige avgifter. *Komponent 13* var imidlertid noe vanskeligere å tolke. Her var det store utslag både på positiv og negativ side blant egenvektorene. På den positive siden gjorde nøkkeltall som VLEV og VKGJ særlig utslag; disse var særlig betydningsfulle også i absolutte verdier. VRGJ og VAM hadde høye negative egenvektorer, dog ikke like høye i absolutte verdier som VLEV og VKGJ. Komponentene ble samlet sett tolket som et mål på vekst i kortsiktig finansiering.

Komponent 14, som er den siste komponenten med egenverdi høyere enn 1, var også noe vanskelig å tolke. ALEV gav særlig utslag på den positive siden, mens SO, SKOF og LA gjorde det på den negative siden. I absolutte verdier hadde SKOF og LA egenvektorene med høyest, og komponenten ble derfor tolket som et mål på kostnader.

4.2.5 Valg av nøkkeltall

Komponentanalysen ligger til grunn for reduksjon av primærmengden av nøkkeltall. Valg av nøkkeltall ble basert på følgende tre kriterier:

- 1) Ett nøkkeltall velges per komponent.
- 2) Nøkkeltallet må korrelere høyt med respektiv komponent.
- 3) Innbyrdes korrelasjon mellom valgte nøkkeltall skal ikke overstige 0,5 i absoluttverdi.

Korrelasjon mellom nøkkeltall og komponenter

I utgangspunktet ble det nøkkeltallet med høyest korrelasjon med komponenten valgt fra hver komponent. Grunnen til at høy korrelasjon mellom nøkkeltallet og komponenten vektlegges er fordi det vil si at nøkkeltallet representerer komponenten godt. Ett nøkkeltall (og ikke flere) per komponent velges for å unngå høy innbyrdes korrelasjon mellom variablene. Tabell 4.9 viser de ulike valgte nøkkeltallene med tilhørende korrelasjon mellom det enkelte nøkkeltall og den respektive komponenten.

Korrelasjon mellom valgt nøkkeltall og komponent					
Rentabilitet		Kortsiktig finansiering		Driftskapital	
PC1		PC2		PC3	
TKR(6)	0,8562	AKGJ	-0,7337	AAK	0,8738
Omløpshastighet		Varelager		Vekst	
PC4		PC5		PC6	
OHO	-0,8293	OHV(2)	-0,8287	VEI	-0,8282
Kapitalstruktur		Aktivitetsnivå		Langsiktig finansiering	
PC7		PC8		PC9	
RK(2)	-0,4713	OHE(3)	-0,4239	OHE(2)	-0,3670
Rentekostnader		Størrelse		Offentlige avgifter	
PC10		PC11		PC12	
RK(1)	-0,4307	ST(1)	-0,4296	SK(1)	0,6324
Vekst kortsiktig finansiering		Kostnader			
PC13		PC14			
VLEV	0,4630	LA	-0,5775		

Tabell 4. 9: Valgte nøkkeltall og deres korrelasjon med respektive komponenter

De seks første komponentene, med unntak av *komponent 2*, har nøkkeltall med korrelasjon over 0,80 (i absoluttverdi). I utgangspunktet ble nøkkeltallene med høyest absolutt korrelasjon i forhold til komponenten valgt, men for *komponent 1 og 10* ble andre nøkkeltall valgt. Bakgrunnen for dette kommenteres i neste delkapittel. Det var dog ikke stor forskjell mellom de valgte nøkkeltallenes korrelasjon og de nøkkeltallene med høyest korrelasjon i forhold til de respektive komponentene.

Innbyrdes korrelasjon mellom nøkkeltall

De 14 nøkkeltallene som ble valgt ut etter komponentanalysen ble så sjekket for innbyrdes korrelasjon. Nøkkeltallene bør ikke korrelere for sterkt med hverandre, da meningen er at nøkkeltallene skal dekke ulike dimensjoner av dataene. Grensen for hvilken innbyrdes korrelasjon som tillates er, som nevnt, i utgangspunktet satt til 0,50 (absoluttverdi), som i Skogsvik (1988).

TKR(7) korrelerer relativt høyt med flere andre nøkkeltall (dog under 0,50): 0,37, 0,30 og 0,38 med hhv. OHE(2), OHE(3) og AKGJ. Det ble derfor foretatt en test av innbyrdes korrelasjon for de neste nøkkeltallene i *komponent 1*, rangert etter korrelasjon med komponenten. Resultatet av dette var at TKR(6) hadde lavere korrelasjon med samtlige

nøkkeltall, men fortsatt høy korrelasjon med *komponent 1* og ble således valgt fremfor TKR(7). Det er små forskjeller hva egenvektorer angår blant totalrentabilitetstallene, slik at det ble vurdert som viktigere å velge et nøkkeltall med lav innbyrdes korrelasjon fremfor høyest mulig egenvektor. OHE(2) og OHE(3) korrelerer høyt med hverandre, men også med AKGJ. Disse to nøkkeltallene ble derfor byttet ut med henholdsvis LG₁(1) og KSG(2), som har relativt moderat korrelasjon med resten av de valgte nøkkeltallene. De øvrige nøkkeltallene som ble valgt ut etter komponentanalysen ble beholdt. Tabell med fullstendig oversikt over innbyrdes korrelasjon ligger i appendiks 7.3. Det kan nevnes at ingen av de valgte nøkkeltallene har innbyrdes korrelasjon høyere enn 0,36.

Etter kontroll for innbyrdes korrelasjon står følgende nøkkeltall igjen:

TKR(6), LA, RK(1), RK(2), SK, OHV(2), OHO, AAK, ST(1), AKGJ, VLEV, VEI, LG₁(1) og KSG(2). Ved å gjennomføre komponentanalyse og kontrollere for innbyrdes korrelasjon har antall nøkkeltall blitt redusert fra 57 til 14.

Levens test

Før en t-test utføres bør det testes hvorvidt observasjonene i gruppene av konkurs- og ikke-konkursbedrifter har lik varians, slik at riktig testobservator blir brukt. En modifisert Levens-test ble utført i Minitab for dette formål. For $\alpha = 0,05$ kan ikke nullhypotesen om lik varians forkastes for LA, RK(1), OHO, AAK, og ARGJ. Appendiks 7.5 viser oversikt over resultatene fra Levens test. Hypotesene på likningsform blir som følger:

$$H_0: \sigma_1^2 = \sigma_0^2$$

$$H_A: \sigma_1^2 \neq \sigma_0^2$$

σ_1^2 er variansen til konkursbedriftene, mens σ_0^2 er variansen til ikke-konkursbedriftene.

T - test

For å ytterligere redusere antall nøkkeltall ble det foretatt en tosidig t-test for å teste for lik gjennomsnittsverdi i gruppen av henholdsvis konkurs- og ikke-konkursbedrifter. Dette ble gjort for hvert av de 14 utvalgte nøkkeltallene. Bakgrunnen var at man anså det som lite hensiktsmessig å ha med nøkkeltall som ikke kunne skille mellom konkurs- og ikke-

konkursbedrifter. Kun de nøkkeltallene som viste signifikant forskjell i gjennomsnittsverdi (P-verdi $< 0,05$), ble benyttet i videre analyse. Nullhypotesen i denne testen er at det er lik gjennomsnittsverdi på det aktuelle nøkkeltallet i begge grupper av bedrifter, mens alternativhypotesen er at gjennomsnittsverdien er ulik. Hypotesene på likningsform blir som følger:

$$H_0: \mu_0 = \mu_1$$

$$H_A: \mu_0 \neq \mu_1$$

hvor μ_0 er gjennomsnittsverdien for et gitt nøkkeltall i gruppen av ikke-konkursbedrifter, mens μ_1 er tilsvarende for konkursbedriftene. For variablene LA, RK(1), OHO, AAK, og ARGJ ble t-test med antakelse om lik varians foretatt. For de resterende nøkkeltallene ble antagelse om ulik varians i de to gruppene lagt til grunn.

For nøkkeltallene LA, RK(2), SK, OHO og VLEV kunne ikke nullhypotesen forkastes på et signifikansnivå for $\alpha = 0,05$ - altså gav t-testen for disse en p-verdi større enn 0,05. Appendiks 7.6 viser en fullstendig oversikt over resultatet fra t-testene.

Etter komponentanalyse og t-test stod det igjen ti nøkkeltall. De gjenstående nøkkeltallene som benyttes i videre analyse er: TKR(6), RK(1), OHV(2), AAK, ST(1), AKGJ, VEI, LG₁(1), ARGJ og KSG(2).

4.3 Logit-analyse

I denne delen blir først resultatet av regresjonen presentert, inkludert en enkel vurdering om fortegnene til koeffisientene gir mening, og videre om forutsetningene for regresjon er tilfredsstillt. Deretter går vi videre inn på tolkning av koeffisientene i form av odds-ratio og de enkelte koeffisientens betydning for konkurssannsynlighet. Til slutt vurderes det hvor godt modellen predikerer konkurs i form av goodness of fit-tester og en analyse av feilklassifiseringer basert på utvalgsdataene.

4.3.1 Resultat av regresjon

Logit-regresjonen ble foretatt i Minitab, hvor utskrift av koeffisientene, p-verdier og log-likelihood oppgis. Den sistnevnte faktoren oppgis for modellen som helhet. Det ble også foretatt ulike tester på goodness of fit, som kommenteres under punkt 4.3.2..

Analysen ble gjennomført for prognoseperioder på henholdsvis ett, to og tre år. Med det menes det at det ble estimert tre ulike modeller som predikerer sannsynligheten for konkurs henholdsvis ett, to og tre år frem i tid.

For at konkursbedriftene og ikke-konkursbedriftene skulle bli sammenliknet over samme tidsperiode, er det for hver konkursbedrift et forholdsmessig likt utvalg av ikke-konkursbedrifter over den samme perioden i datasettet. Av antall konkursbedrifter gikk 47 % konkurs i 2011, 26 % i 2010, 14 % i 2009 og 19 % i 2008. For prognoseperiode på ett år betyr det at 47 % er 2010-tall, 26 % er 2009-tall, 14 % er 2008-tall og 19 % er 2007-tall for både konkurs- og ikke-konkursbedriftene

I utgangspunktet ble alle de ti variablene omtalt i siste avsnitt under punkt 4.2.5 inkludert i regresjonen. Nøkkeltall som ikke var signifikant forskjellig fra null for $\alpha = 0,05$ ble ekskludert og regresjonen ble gjennomført på nytt. Denne prosessen ble gjentatt frem til modellen utelukkende bestod av koeffisienter signifikant forskjellig fra null.

Som nevnt under punkt 3.5.4, om signifikanttesting, er det noe usikkerhet knyttet til om en z-test er en gyldig signifikantstest av koeffisientene ved eventuell ulik fordeling blant variablene. Det ble derfor testet for ulike fordelinger blant nøkkeltallene i Minitab. Ut fra testene var det vanskelig å fastslå fordelingen til de ulike nøkkeltallene. På bakgrunn av dette ble alle nøkkeltallene også testet mot likelihoodratio-testen for med større sikkerhet kunne si at alle de inkluderte variablene er signifikante. Forskjeller i log-likelihood ble testet for regresjon med alle variablene mot regresjon med alle variablene unntatt én. Dette ble gjentatt for alle variablene og resultatet ble det samme som ved z-testen for alle tre prognoseårene.

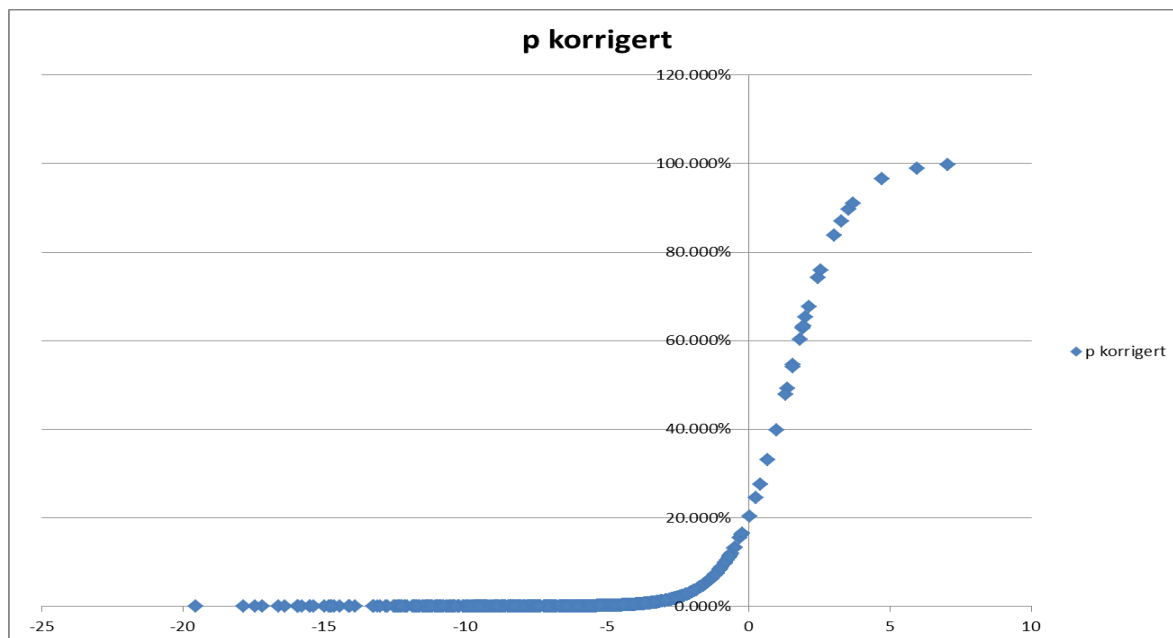
Ett års prognoseperiode

Modellen for ett års prognoseperiode ble estimert ved hjelp av data ett år før konkurs for konkursbedriftene og en forholdsmessig lik fordeling av data fra ikke-konkursbedrifter for samme år. Regresjonsprosessen ble gjentatt frem til samtlige koeffisienter var signifikant forskjellig fra null for $\alpha = 0,05$. For koeffisientene til RK(1), LG₁(1), KSG(2), AAK, og VEI kan ikke nullhypotesen om at koeffisientene er lik null for $\alpha = 0,05$ forkastes. Dette samsvarer med Likelihoodratio-testen. Etter eliminering av ikke-signifikante koeffisienter gav nøkkeltallene TKR(6), OHV(2), ST(1) og AKGJ og følgende lineær sammenheng:

$$L = 16,7113 - 8,97576 * TKR(6) + 4,26569 * OHV(2) - 1,98331 * ST(1) + 4,04992 * AKGJ$$

En første vurdering av modellen ble gjort ved å se hvorvidt fortegnene til koeffisientene gav fornuftige resultater. Det er intuitivt at høyere rentabilitet (TKR(6)) reduserer sannsynligheten for konkurs, mens andel kortsiktig gjeld (AKGJ) øker sannsynligheten for konkurs. Rentabilitet beskriver bedriftens lønnsomhet, og en bedrift som ikke er lønnsom vil ikke overleve over tid. Det å ha en stor andel kortsiktig gjeld utgjør en risiko, fordi kortsiktig gjeld er en løpende forpliktelse som må betales innen relativt kort tid. Økt OHV(2) (varelager/kostnader) øker ifølge modellen sannsynligheten for konkurs, mens økt størrelse (ST(1)) reduserer sannsynligheten for konkurs. OHV(2) kan øke ved at enten varelageret øker eller at kostnadene reduseres. Det kan bety at økt OHV(2) ikke nødvendigvis bør gi en høyere konkurssannsynlighet og at resultatet derfor ikke er like intuitivt som resten. Det at større total kapital (ST(1)) gir redusert sannsynlighet for konkurs virker fornuftig da det antas at en større bedrift har større reserver å tære på i dårlige tider. Dette samsvarer med funn i flere tidligere studier, blant annet finner også Ohlson (1980) at størrelse er en viktig variabel.

En viktig forutsetning for logistisk regresjon er formen på S-kurven. I dette tilfellet vil punktene på kurven være korrigerede sannsynligheter fra modellen på Y-aksen mot logit-verdiene på X-aksen. Korrigert sannsynlighet er beregnet først ved å regne om logit-verdien til sannsynlighet, for så å korrigere denne for utvalgsstørrelsen (jf. Skogsvik 2005). Figur 4.1 viser S-kurven til modellen for ett års prognoseperiode, og ut i fra plottet ser forutsetningen om S-form ut til å være oppfylt. Naturlig nok er det betydelig flere observasjoner i venstre hale grunnet større andel ikke-konkursbedrifter i datasettet, som følgelig har lavere konkurssannsynlighet.



Figur 4. 1: S-kurve ett års prognoseperiode

To års prognoseperiode

På samme måte som for ett års prognoseperiode ble modellen for to års prognoseperiode estimert på bakgrunn av data to år før konkurs for konkursbedriftene og en forholdsmessig lik andel data fra ikke-konkursbedrifter for de samme årene. Etter eliminering av ikke-signifikante koeffisienter for $\alpha = 0,05$ gav nøkkeltallene TKR(6), RK(1), LG₁(1), KSG(2), ST(1) og AKGJ følgende lineær sammenheng:

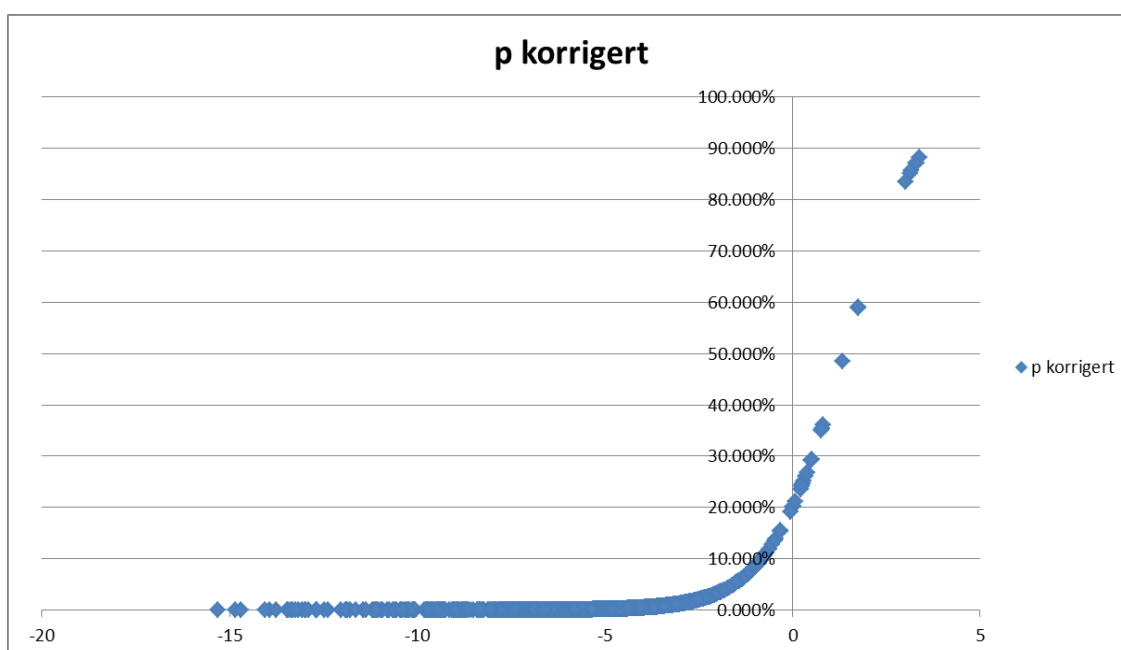
$$L = 14,5509 - 7,55622*TKR(6) + 11,3671*RK(1) - 8,15395*LG_1(1) + 0,0031207*KSG(2) - 1,65789*ST(1) + 3,0924*AKGJ$$

I forhold til prognoseperiode for ett år er variablene RK(1), LG₁(1) og KSG(2) inkludert, mens OHV(2) er utelatt. KSG(2) har dog en veldig lav koeffisient og fortegnet gir ikke et meningsfullt resultat. Man skulle tro høyere kontantstrømgrad gir lavere sannsynlighet for konkurs. Denne variabelen ble derfor utelukket og regresjonen ble foretatt på nytt med følgende resultat:

$$L = 13,5507 - 6,78894*TKR(6) + 9,85507*RK(1) - 5,25059*LG_1(1) - 1,56903*ST(1) + 3,02310*AKGJ$$

I denne likningen gir fortegnene til samtlige koeffisienter meningsfulle resultater. En høyere $LG_1(1)$ resulterer i lavere sannsynlighet for konkurs, mens høyere $RK(1)$ øker sannsynligheten. Det virker fornuftig at bedre likviditet reduserer konkurssannsynligheten da bedriften i større grad vil være i stand til å håndtere sine forpliktelser. Økte rentekostnader øker sannsynligheten for konkurs, noe som også virker fornuftig. En økning i rentekostnader kan enten skyldes opptak av gjeld eller økt rente på eksisterende gjeld. Rentekostnadene må betales uansett og de vil således representere en risiko. $TKR(6)$, $ST(1)$ og $AKGJ$ er, som for ett års prognoseperiode, også viktig for predikere konkurssannsynlighet to år frem i tid.

Figur 4.2 viser korrigerte sannsynligheter fra den estimerte modellen plottet mot logit-verdiene. Sammenlignet med S-kurven til modellen for ett års prognoseperiode er ikke S-formen like tydelig.



Figur 4. 2: S-kurve to års prognoseperiode

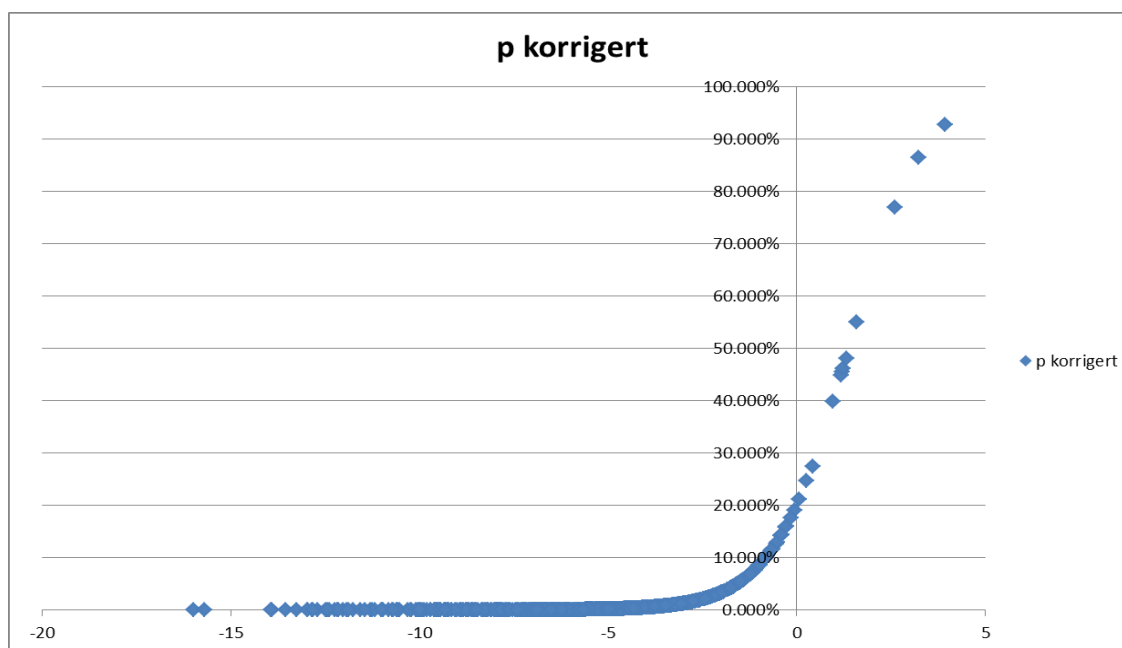
Tre års prognoseperiode

Analysen for tre års prognoseperiode ble utført på tilsvarende måte som de to foregående analysene. Datautvalget ble satt sammen av data for tre år før konkurs for konkursbedriftene og med data fra et forholdsmessig likt antall ikke-konkursbedrifter for samme år. Etter at

ikke-signifikante koeffisienter ble eliminert og regresjonen gjennomført på nytt gav nøkkeltallene TKR(6), LG₁(1), ST(1) og VEI følgende resultat:

$$L = 17,8382 - 4,34656*TKR(6) - 7,93162*LG_1(1) - 1,78603*ST(1) + 1,58257*VEI$$

Koeffisientenes fortegn gir også for denne modellen fornuftige resultater. Som for de to foregående modellene gir høyere TKR(6) og ST(1) redusert konkurssannsynlighet. Det samme er tilfellet for LG₁(1) som inngår i modellen for både to og tre års prognoseperiode. Det er imidlertid interessant å se at VEI kun er med i modellen for tre års prognoseperiode. Høyere vekst på eiendelene øker i følge modellen risikoen for konkurs på litt lengre sikt. Et scenario kan være at mange bedrifter har ambisjoner om høy vekst, hvor veksten lånefinansieres, noe som gir problemer senere.



Figur 4. 3: S-kurve tre års prognoseperiode

Figur 4.3 viser korrigert sannsynlighet fra modellen for tre års prognoseperiode plottet mot logit-verdiene. S-kurven er ikke like tydelig som for ett års prognoseperiode, men noe tydeligere enn for to års prognoseperiode.

Odds-ratio

Som nevnt under metode-delen er det vanskelig å tolke hver enkelt koeffisient siden dette ikke er lineær regresjon. I dette avsnittet gjør vi et forsøk på å tolke hver enkelt koeffisient ved hjelp av å se på odds-ratio. Alle odds-ratioene ligger i regresjonsutskriftene i appendiks 7.7-7.9.

Ett års prognoseperiode

For ett års prognoseperiode har TKR(6) og ST(1) odds-ratio under 1, noe som er naturlig siden fortegnene er negative. Oddsen reduseres når variabelen øker, som igjen vil si at sannsynligheten for konkurs reduseres når variabelen øker. TKR(6) har odds-ratio på tilnærmet null (med to desimaler). Det lave tallet har sammenheng med et lavt tall i koeffisienten (-8,976) som vil si at sannsynligheten for konkurs reduseres ved økt TKR(6).

ST(1) har odds-ratio på 0,14 og en koeffisient på -1,983. Det vil si at oddsen endrer seg med 0,14 ganger for én enhet økning i variabelen. Sammenlignet med TKR(6) vil en endring med én enhet ha mindre effekt for ST(1) på grunn av høyere koeffisient og høyere odds-ratio. Det må dog understrekes at de fleste bedriftene i datasettet vårt ligger mellom 10 og 14 for ST(1) (ln til sum eiendeler) og -0,10 og 0,40 for TKR(6). Da er det naturlig at én enhet endring i TKR(6) gjør større utslag på sannsynlighet for konkurs enn ST(1).

OHV(2) og AKGJ har relativt like koeffisienter og høye odds-ratio tall, henholdsvis 4,266 og 71,21 for OHV(2) og 4,050 og 57,39 for AKGJ. Det vil si at oddsen for konkurs øker med henholdsvis 71,21 og 57,39 ganger for én enhet økning i variablene, noe som igjen vil si at sannsynligheten for konkurs er 71,21 og 57,39 ganger i forhold til sannsynligheten for å ikke gå konkurs ved én enhets økning i variabelen.

To års prognoseperiode

For to års prognoseperiode gir TKR(6) og ST(1) omtrent samme resultat som for ett års sikt med henholdsvis 0 og 20,55 i odds-ratio. $LG_1(1)$ har odds-ratio på 0,01 som ikke er unaturlig når koeffisienten er såpass lav (-5,25). Odds-ratioen til RK(1) har ekstremt høy verdi, grunnet at størsteparten av datasettet ligger mellom 0 og 0,1 så én enhets endring er ekstremt for denne variabelen.

Tre års prognoseperiode

For tre års prognoseperiode gir TKR(6), ST(1) og LG₁(1) omtrent samme resultat som i de andre prognoseperiodene med henholdsvis 0,01, 0,17 og 0,00 i odds-ratio. VEI er den eneste variabelen med positiv koeffisient og odds-ratio over 1 (4,87).

Sannsynlighet

I dette avsnittet tolkes koeffisientene ved å regne ut endring i sannsynlighet ved én enhets endring i den uavhengige variabelen. Denne effekten er dog avhengig av hvilket på S-kurven (figur 3.1) man tar utgangspunkt i. Ved konkurssannsynlighet på 50 % er effekten for hver koeffisient størst (bratteste punktet på S-kurven). De fleste bedriftene er ikke i nærheten av å ha så høy sannsynlighet for konkurs - derfor mener vi dette utgangspunktet er uinteressant (se appendiks 7.10-7.12). Vi velger heller å se på effekten av hver enkelt koeffisient hvor de fleste bedriftene befinner seg på S-kurven. Datasettet har 43 konkursbedrifter og 890 bedrifter totalt, som gir en gjennomsnittlig konkurssannsynlighet på 4,83 % (uten korrigerings for utvalgsskjevhet). Vi kan regne ut effekten av hver enkelt koeffisient på dette punktet ved hjelp av denne formelen:

$$b \times p(1-p)$$

Ett års prognoseperiode

For ett års prognoseperiode er effekten ved én enhets endring i hver enkelt uavhengig variabel, gitt alt annet likt, som følger:

$$TKR(6): -8,976 \times 0,0483(1-0,0483) = -0,413$$

$$OHV(2): 4,266 \times 0,0483(1-0,0483) = 0,196$$

$$ST(1): -1,983 \times 0,0483(1-0,0483) = -0,091$$

$$AKGJ: 4,050 \times 0,0483(1-0,0483) = 0,186$$

Én enhet økning (100 prosentpoeng) i TKR(6), gitt alt annet likt, vil redusere sannsynligheten for konkurs med 0,413 eller 41,3 prosentpoeng, mens for ST(1) reduseres sannsynligheten med 9,1 prosentpoeng. Én enhet endring i variabelen vil naturlig nok gi større utslag for TKR(6) enn ST(1) da ST(1) har større variasjon. OHV(2) og AKGJ påvirker

sannsynligheten for konkurs relativt likt med henholdsvis en økning på 19,6 og 18,6 prosentpoeng hvis nøkkeltallene øker med én enhet.

To års prognoseperiode

For to års prognoseperiode er effekten ved én enhets endring i hver enkelt uavhengig variabel, gitt alt annet likt, som følger:

$$TKR(6): -6,789 \times 0,0483(1-0,0483) = -0,312$$

$$RK(1): 9,855 \times 0,0483(1-0,0483) = 0,453$$

$$LG_1(1): -5,250 \times 0,0483(1-0,0483) = -0,241$$

$$ST(1) -1,569 \times 0,0483(1-0,0483) = -0,0721$$

$$AKGJ: 3,023 \times 0,0483(1-0,0483) = 0,139$$

Sammenlignet med ett års prognoseperiode er effekten av TKR(6), ST(1) og AKGJ redusert. Én enhets endring i disse nøkkeltallene vil påvirke sannsynligheten for konkurs mindre enn i modellen for ett år. RK(1), som ikke er med i ett års prognoseperiode, har betydelig effekt på endring i sannsynlighet med 45,3 prosentpoeng. Det må imidlertid nevnes at maksimal variasjon i datasettet for denne variabelen er 0,40, som ikke er i nærheten av én enhet. LG₁(1) er heller ikke med i modellen for ett års prognoseperiode. Også denne variabelen har en relativt stor effekt på konkurssannsynligheten ved én enhets endring, dog mindre enn RK(1) og TKR(6).

Tre års prognoseperiode

For tre års prognoseperiode er effekten ved én enhets endring i hver enkelt uavhengig variabel, gitt alt annet likt, som følger:

$$TKR(6): -4,347 \times 0,0483(1-0,0483) = -0,200$$

$$LG_1(1): -7,931 \times 0,0483(1-0,0483) = -0,365$$

$$ST(1): -1,786 \times 0,0483(1-0,0483) = -0,082$$

$$VEI: 1,583 \times 0,0483(1-0,0483) = 0,073$$

Sammenliknet med de to andre prognoseperiodene er $LG_1(1)$ den som gir størst effekt ved endring i én enhet. VEI gir relativt liten effekt grunnet at spredningen er litt større enn ved de andre. I datasettet finnes det tall opp mot 2 og ned mot -0,5. De fleste ligger dog mellom -0,2 og 0,3. TKR(6) er mindre betydningsfull enn i de to andre prognoseperiodene, mens ST(1) har omtrent samme effekt.

Tolkning av koeffisientene oppsummert

Ser man modellene for alle tre prognoseperiodene under ett, er rentabilitet og størrelse på bedriften viktig, uansett om det er på kort (ett år) eller lengre sikt (tre år). Andel kortsiktig gjeld er viktig for å predikere på kort og medium lang sikt, mens likviditet er viktig på medium og lang sikt. Vekst i eiendelene er kun viktig ved predikering på lengre sikt. En tolkning av dette kan være at, gitt alle andre variabler like, bedrifter med høy andel kortsiktig gjeld har høyere sannsynlighet for å gå konkurs på kort sikt, mens en bedrift med høy vekst i eiendelene har høyere sannsynlighet for å gå konkurs på lengre sikt. Det er også verdt å nevne at TKR(6) og ST(1) gjorde klart størst utslag i likelihood-ratio testen, noe som understreker betydningen av disse to variablene. Appendiks 7.7-7.9 viser resultatene fra Minitab.

Odds-ratioene gir noe begrenset informasjon om koeffisientene. Som ved å tolke fortegnene kan man si noe om hvorvidt variabelen påvirker sannsynligheten i positiv eller negativ retning. For alle prognoseperiodene har TKR(6) svært lave odds-ratio tall, mens ST(1) har noe høyere, dog fortsatt veldig lave tall, som kan si noe om betydningen av disse to koeffisientene.

Å se på endring i sannsynlighet for hver enkelt koeffisient er lettere å tolke, men man må ta hensyn til hvor på S-kurven man er. Det er dog interessant å se at effekten av TKR(6) minker med antall år man predikerer for, mens $LG_1(1)$ har større betydning for å predikere tre år sammenliknet med to år. For prognoseperiode på to år er effekten av RK(1) størst, mens for tre års prognoseperiode er effekten av $LG_1(1)$ størst. ST(1) har omtrent samme effekt uavhengig av hvilken prognoseperiode som benyttes.

Som vi har indikert over, kan det være meningsløst å se på odds-ratio og effekten av hver enkelt koeffisient for hvert enkelt prognoseår alene. Alle variablene blir sammenlignet mot én enhets endring. Dette er en betydelig endring for noen nøkkeltall, men ikke engang et standardavvik for andre (jf. RK(1) og ST(1)). Å sammenligne effekten av de forskjellige variablene mot hverandre er derfor vanskelig. Ved å sammenligne samme variabel over flere perioder kan man derimot se hvor betydningsfull den enkelte variabel er relativt til de andre periodene.

4.3.2 Resultater fra modellene

For å kunne vurdere resultatene modellene gir, har vi utført en statistisk test av goodness of fit og en analyse av feilklassifisering. Analysen av feilklassifisering inkluderer analyse av andel type 1- og type 2-feil, samt total feilandel.

Goodness of fit estimerte modeller

De ulike goodness of fit-testene tester hvor godt predikerte verdier stemmer overens med observerte verdier. Hypotesene kan for goodness of fit-testene generelt defineres slik:

H₀: Modellen forklarer bra

H_A: Modellen forklarer ikke bra

Vi vil vektlegge Hosmer-Lemeshow-testen i og med at ikke alle testene, som nevnt, er like godt egnet i vårt tilfelle. Bakgrunnen for dette er at henholdsvis Pearson og Deviance-testene ikke er et godt mål på goodness of fit dersom antall kovariater er lik antall observasjoner. Dette er tilfellet i de estimerte modellene, noe som kan sees ut ifra Minitab-utskriftene (appendiks 7.7-7.9). Antall frihetsgrader tilsvarer her antall observasjoner fratrukket det antall uavhengige variabler som inngår i de tre estimerte modellene (jf. Tufte 2000). Vi velger også å se bort fra Brown's test, jf. kritikken til le Cessie og van Houwelingen (1991).

Ett års prognoseperiode

Appendiks 7.7 viser resultatene fra Hosmer-Lemeshow-testen. Denne testen fordeler de predikerte resultatene inn i ti grupper basert på prosentiler hvor målet er å få et mest mulig likt antall i de ulike gruppene. I appendiks 7.7 ser vi at dette stemmer bra for ikke-konkursbedriftene (Value = 0), mens for konkursbedriftene (Value = 1) er det noe avvik mellom gruppene. Sistnevnte skyldes at det er relativt få observasjoner av konkursbedrifter i datautvalget, men totalt sett vil validiteten av testen være tilfredsstillende.

Resultatene fra denne testen kan betraktes som en krysstabell hvor man har to verdier på avhengig variabel ganger antall grupper (Tuft 2000). Dette betyr at man kan beregne kjikvadratverdien på egenhånd, men da totalen for alle gruppene er oppgitt i Minitab gjøres ikke det. Antall frihetsgrader tilsvarer antall grupper minus antall verdier på avhengig variabel slik at $DF = 10 - 2 = 8$. Kjikvadratverdien til testen er 8,014, mens p-verdien er 0,432. Dette betyr at for et signifikansnivå hvor $\alpha = 0,05$ kan ikke nullhypotesen om at observert og forventet fordeling er like forkastes.

To års prognoseperiode

Appendiks 7.8 viser resultatene for Hosmer-Lemeshow-testen utført på den estimerte modellen for to års prognosesikt. På samme måte som for modellen for ett års prognoseperiode grupperes de predikerte resultatene i ti grupper. Tabellen viser at grupperingen er forholdsvis lik som for ett års prognosesikt, noe som betyr at testen vil gi valide resultater. Antall frihetsgrader er også for denne modellen 8, mens kjikvadratverdien er 14,83. P-verdien til denne testen er 0,062, noe som betyr at den så vidt er signifikant på et signifikansnivå hvor $\alpha = 0,05$. Nullhypotesen om at observert og forventet fordeling er like kan altså ikke forkastes.

Tre års prognoseperiode

Appendiks 7.9 viser, på samme måte som for de to foregående modellene, resultatet fra Hosmer-Lemeshow-testen utført i Minitab. Det er også for denne modellen forholdsvis godt samsvar mellom forventet antall og observert antall i de ulike gruppene. Som for de to foregående testene er også for denne testen antall frihetsgrader lik 8. Videre gir testen en kjikvadratverdi på 3,665 og en p-verdi på 0,886. Ut fra dette kan heller ikke her

nullhypotesen om lik observert og forventet fordeling forkastes på et signifikansnivå hvor $\alpha = 0,05$.

Analyse av type 1- og type 2-feil

Når grensen for klassifisering som konkurs skal settes, må man gjøre en avveining rundt feilklassifiseringskostnadene og hvordan de ulike feilklassifiseringene vektet. Det er vanskelig å vurdere kostnaden på type 1- og type 2-feil, men tidligere litteratur har funnet at type 1-feil i de fleste tilfeller er mest kostbart (Altman et al. 1981) slik at målet i denne utredningen har vært å finne en grense som minimerer denne andelen, samt total feilandel. Total feilandel er funnet ved følgende formel:

$$\text{Total feilandel} = [(B_0/B) * f^0(p) + (B_1/B) * f^1(p)]$$

$f^0(p)$ representerer andel type 2-feil, altså feilklassifiserte ikke-konkursbedrifter. $f^1(p)$ representerer andel type 1-feil, altså andel feilklassifiserte konkursbedrifter. Disse blir vektet med henholdsvis andelen ikke-konkurs og konkursbedrifter i datautvalget, (B_0/B) og (B_1/B) .

Grensesettingen ble gjort ved å prøve og feile. Konkurs sannsynligheten for alle konkursbedriftene ble testet frem til et tilfredsstillende resultat ble oppnådd. Utgangspunktet var å få andelen type 1-feil lavest mulig, helst under 10 %, samtidig som hensynet til total feilandel ble ivarettatt.

Ett års prognoseperiode

For ett års prognoseperiode ble en grense (G) på 1,31 % satt. Dette resulterte i andel type 1-feil på 9,30 %, andel type 2-feil på 11,81 %, og en total feilandel på 15,62 %. Testet på det opprinnelige datautvalget klassifiserer modellen for ett års prognoseperiode 84,38 % korrekt. Tabell 4.10 viser en oversikt over ulike G og feilandeler.

Type 1-feil	Type 2-feil	Totalt	G
0,00 %	58,09 %	60,11 %	0 %
4,65 %	42,98 %	45,51 %	0,14 %
6,98 %	18,89 %	22,47 %	0,65 %
9,30 %	15,11 %	18,76 %	0,88 %
9,30 %	11,81 %	15,62 %	1,31 %
11,63 %	11,81 %	15,51 %	1,32 %

Tabell 4. 10: Andel feilklassifiseringer ett års prognoseperiode

To års prognoseperiode

Klassifiseringsgrensen for to års prognoseperiode ble satt til 1,94 %. Dette gav en andel type 1-feil på 11,63 %, andel type 2-feil på 9,92 %, og en total feilandel på 13,71 %. Dette tilsvarer at modellen predikerer 86,29 % korrekt når man tester på det opprinnelige datautvalget. Tabell 4.11 viser en oversikt over ulike G og feilandeler for to års prognoseperiode.

Tilsynelatende predikerer denne modellen noe bedre totalt sett enn modellen for ett års prognoseperiode på tross av at forutsetningene for regresjonen ikke er like tilfredsstillende (noe mer utydelig S-kurve). På en annen side kan man i appendiks 7.10 og 7.11 se at det er flere konkursbedrifter med høy konkurssansynlighet for ett års prognoseperiode i forhold til to års prognoseperiode. Dette indikerer at modellen for ett år likevel predikerer bedre enn modellen for to år.

Type 1-feil	Type 2-feil	Totalt	G
0 %	64,58 %	66,29 %	0,07 %
4,65 %	50,89 %	53,03 %	0,16 %
4,65 %	46,52 %	48,88 %	0,22 %
6,98 %	38,84 %	41,46 %	0,34 %
11,63 %	21,49 %	24,72 %	0,83 %
11,63 %	9,92 %	13,71 %	1,94 %
13,95 %	9,92 %	13,60 %	1,95 %

Tabell 4. 11: Andel feilklassifiseringer to års prognoseperiode

Tre års prognoseperiode

For tre års prognoseperiode ble klassifiseringsgrensen satt til 1,16 %. Det gir en andel type 1-feil på 9,30 %, en andel type 2-feil på 20,43 %, og en total feilklassifiseringsandel på 23,82 %. Modellen for tre års prognoseperiode predikerer altså 76,18 % korrekt når den testes på det opprinnelige datautvalget. Dette resultatet er noe svakere enn for de to foregående modellene. Tabell 4.12 viser en oversikt over ulike grenser og feilklassifiseringsandeler.

Type 1-feil	Type 2-feil	Totalt	G
0 %	65,76 %	67,42 %	0,08 %
4,65 %	43,92 %	46,40 %	0,34 %
6,98 %	33,06 %	35,96 %	0,59 %
9,30 %	27,04 %	30,11 %	0,78 %
9,30 %	20,43 %	23,82 %	1,16 %
11,63 %	20,43 %	23,71 %	1,17 %
13,95 %	20,19 %	23,37 %	1,20 %

Tabell 4. 12: Andel feilklassifiseringer tre års prognoseperiode

Oppsummert

Gitt antakelsen om at type 1-feil er mest kostbart, predikerer alle tre estimerte modeller forholdsvis bra. Andelen feilklassifiserte bedrifter ser ut til å være i samme størrelsesorden som i tidligere studier hvor logistisk regresjon er benyttet (jf. litteraturgjennomgangen). I utgangspunktet er det naturlig å anta at det er lettere å predikere over en kortere tidsperiode. Modellen for to års prognoseperiode gir likevel lavest total feilandel.

Ser man på plottene over beregnet konkurssannsynlighet for konkursbedriftene i appendiks 7.10-7.12 ser man betydelig høyere andel konkursbedrifter med høy konkurssannsynlighet for prognoseperioden på ett år, sammenlignet med to og tre år. Nesten 50 % av konkursbedriftene i prognoseperioden på ett år har konkurssannsynlighet over 20 %, mot 35 % og 17 % for henholdsvis to og tre år. Tilsynelatende predikerer modellen for to års prognoseperiode best når total feilandel er lagt til grunn. Ser man på plottene over konkurssannsynlighet for konkursbedriftene ser imidlertid modellen for ett års prognoseperiode ut til å predikere best.

5. Diskusjon

5.1 Oppsummering og konklusjon

Formålet med denne utredningen er å lage en generell modell som kan predikere konkurssannsynlighet på bakgrunn av offentlig tilgjengelig regnskapsinformasjon. Med generell så menes en modell som kan benyttes på tvers av bransjer. Det er viktig at modellen er enkel å bruke og ikke krever tilgang på data som er vanskelig tilgjengelig, samtidig som modellen er pålitelig og enkel å tolke.

Datautvalget består av 890 norske bedrifter hvorav 43 har gått konkurs i perioden 2008-2011. For å predikere ett, to og tre år før konkurs er regnskapstall for perioden 2005-2010 brukt. Vi har sett bort fra finans- og forsikringsbransjen i denne utredningen grunnet deres finansielle struktur (jf. Ohlson 1980 og Sjøvoll 1999). Alle bedriftene i datasettet er aksjeselskaper, med eiendeler på minimum 15 millioner kroner, minimum 10 ansatte og inntekter over 30 millioner kroner.

Metodene benyttet til å konstruere og estimere modellen er prinsippal komponentanalyse, t-test og logistisk regresjon (her: logit). Komponentanalyse ble brukt for å finne de nøkkeltallene som har størst betydning for analysen, og reduserte antall nøkkeltall fra 57 til 14. T-test ble så gjennomført for å skille ut de nøkkeltallene som viste signifikant forskjell i gjennomsnittsverdier mellom konkurs- og ikke-konkursbedrifter. Etter at komponentanalyse og t-test var gjennomført gjenstod ti nøkkeltall som ble benyttet i den logistiske regresjonen: TKR(6), RK(1), OHV(2), AAK, ST(1), AKGJ, VEI, LG₁(1), ARGJ og KSG(2). Alle nøkkeltallene bortsett fra AAK, ARGJ og KSG(2) er med i minst én av prognoseperiodene. Koeffisientene til AAK og ARGJ var ikke signifikant forskjellig fra null, mens KSG(2) hadde svært lav koeffisient og et lite intuitivt fortegn. Disse ble derfor utelatt fra modellene.

TKR(6) er klart viktigst i modellen da dette nøkkeltallet både gjorde stort utslag i komponentanalysen og regresjonen. Rentabilitetsnøkkeltallene står for mesteparten av variasjonen i dataene (komponentanalyse) og TKR(6) er et viktig nøkkeltall i alle tre prognoseperiodene. Viktigheten av rentabiliteten er dog minkende jo lengre prognoseperiode. Naturligvis vil høyere rentabilitet redusere konkurssannsynligheten.

Størrelsen på selskapet (ST(1)) er også med i alle tre prognoseperiodene – noe som understreker viktigheten av dette nøkkeltallet. Jo større selskapet er, jo lavere er risikoen for å gå konkurs – da større selskaper sjeldnere går konkurs. Det er også rimelig å anta at de har større reserver å tære på. Dette kan dog skape et problem når man skal sammenligne selskaper som varierer mye i størrelse. Dette kommer vi tilbake til under punkt 5.2.

For å predikere ett år frem i tid er også OHV(2) (varelager/kostnader) og AKGJ (andel kortsiktig gjeld) viktige. Et varelager som vokser i forhold til kostnadene er en mulig indikasjon på at innkjøp/produksjon ikke er tilpasset etterspørsel, noe som kan føre til likviditetsproblemer og igjen insolvens. På en annen side kan OHV(2) øke fordi kostnadene reduseres, slik at tolkningen av dette nøkkeltallets betydning for konkurssannsynligheten kan være noe problematisk. AKGJ er også viktig for å predikere to år frem i tid (dog viktigst for ett år). Kortsiktig gjeld er en forpliktelse for bedriften som må betales innen relativt kort tid. En høy andel vil kunne utgjøre en risiko og således øke konkurssannsynligheten.

Rentekostnader (RK(1)) representerer noe av den samme risikoen som AKGJ. Høye rentekostnader betyr at man har mye gjeld og/eller høy rentesats, som kan øke konkursrisikoen ved at man har en betydelig forpliktelse. Likviditetsgraden (LG₁(1)) er viktig for å predikere både to og tre år frem i tid, dog viktigst for tre år. Lavere likviditet utgjøre større risiko for konkurs, fordi man da ikke vil være i stand til å møte forpliktelser i like stor grad, og faren for insolvens øker.

Sammen med TKR(6), LG₁(1) og ST(1) er også vekst i eiendeler (VEI) viktig for å predikere tre år frem i tid. I følge modellen vil høy vekst i eiendelene øke risikoen for å gå konkurs. Bakgrunnen for dette kan være at vekst ofte kan kreve usikre investeringer, gjerne finansiert med gjeld, som kan øke risikoen for konkurs. Tabell 5.1 viser en samlet oversikt over hvilke nøkkeltall som inngår i de tre estimerte modellene.

Nøkkeltall	Ett år	To år	Tre år
TKR(6)	x	x	x
OVH(2)	x		
ST(1)	x	x	x
AKGJ	x	x	
RK(1)		x	
LG ₁ (1)		x	x
VEI			x

Tabell 5. 1: Oversikt over hvilke nøkkeltall som inngår i de estimerte modellene

Under punkt 4.3.2 kom vi frem til at å predikere *to* år frem i tid, under antakelsen om lavest total feilandel, gir bedre resultat enn å predikere *ett* og *tre* år frem i tid. For prognoseperiode ett år satte vi en klassifiseringsgrense på 1,31 %. Total feilandel var da 15,62 %, mens type 1-feil og type 2-feil utgjorde hhv 9,30 % og 11,81 %. Dette er lignende resultater vi har sett i tidligere undersøkelser. For to års prognoseperiode kom vi frem til en klassifiseringsgrense som var noe høyere (1,94 %), men som resulterte i en noe lavere total feilandel, 13,71 %. Type 1-feil var dog noe høyere (11,63 %) enn for ett års prognoseperiode. Med antakelse om at type 1-feil er mest kostbart (jf. Altman et al. 1981) kan det likevel konkluderes med at modellen for ett år prognoseperiode er noe bedre. Denne konklusjonen forsterkes dersom man også ser på plottene i appendiks 7.10 og 7.11, som viser konkurssannsynligheten for konkursbedriftene henholdsvis ett og to år før konkurs. Det er klart flest konkursbedrifter med høy konkurssannsynlighet for prognoseperioden på ett år, sammenlignet med to års prognoseperiode.

Tre års prognoseperiode predikerer dårligere enn de to andre med total feilandel på 23,82 %, hvor type 1-feil og type 2-feil er henholdsvis 9,30 % og 20,43 %. Dette med en lavere klassifiseringsgrense på 1,16 %. Fra plottet i appendiks 7.12 ser man også at det er klart færre konkursbedrifter med høy konkurssannsynlighet sammenlignet med prognoseperiodene ett og to år.

Når det gjelder fastsettelsen av en klassifiseringsgrense, bør dette vurderes i forhold til kostnaden av type 1-feil og type 2-feil, hvor type 1-feil i de fleste tilfeller antas å være mest kostbart. Grensene vi har satt er kun retningslinjer basert på vårt datasett og våre antagelser rundt feilklassifiseringskostnader, hvor vi i tillegg vet hvilke bedrifter som har gått konkurs eller ikke. Ved bruk av modellene bør det vurderes om klassifiseringsgrensene gir et meningsfullt resultat og eventuelt tilpasse disse grensene.

Analysen av feilandelene og plottene av konkurssannsynlighetene til konkursbedriftene indikerer at modellene ikke er egnet til å gi en eksakt konkurssannsynlighet. Vår anbefaling er derfor at modellene brukes til å sammenligne den/de bedriften(e) man er interessert i over flere år for å vurdere bedriften(e)s utvikling. Eventuelt kan resultatet for én bedrift vurderes opp mot andre bedrifter slik at man kan vurdere relativ risiko. Best resultat får man hvis bedriftene ikke varierer for mye i størrelse, grunnet størrelseskomponenten i modellen. Størrelsesvariabelen diskuteres ytterligere under punkt 5.2.

Denne utredningen ble innledet med et spørsmål om hvorvidt det er mulig å estimere en god, generell modell for konkursprediksjon inntil tre år forut for konkurs. Ved å konstruere og estimere en modell for henholdsvis ett, to og tre års prognoseperiode, kan dette på bakgrunn av ovenstående analyse og diskusjon kunne bekreftes. Det er fullt mulig, men det er imidlertid visse begrensninger som man bør være klar over. Noen av disse er nevnt over, mens noen kommenteres i neste delkapittel.

I appendiks 7.13 ligger en oppsummering av fremgangsmåten for å beregne konkurssannsynlighet. Det krever noe regning, men det skal være rimelig intuitivt og greit å utføre disse operasjonene.

5.2 Kritikk av modellene

Vi har i denne utredningen kommet frem til tre modeller som både er forholdsvis enkle å anvende og gir forholdsvis gode resultater. Det er imidlertid viktig å være klar over visse begrensninger/svakheter ved de estimerte modellene.

Når det gjelder sammensetningen av datautvalget, er det er to fremgangsmåter man kunne benyttet. Man kan enten velge seg ut et utvalg bedrifter og så identifisere henholdsvis friske bedrifter og konkursbedrifter, eller man kan velge ut konkursbedriftene først. Da det er relativt få konkurser i Norge, har vi funnet sistnevnte fremgangsmåte mest hensiktsmessig. Dette kan imidlertid ha ført til at utvalget ikke representerer virkeligheten på en optimal måte.

Modellen tar ikke hensyn til når regnskapene offentliggjøres, noe som betyr at det ikke kontrolleres for om de enkelte bedriftene har gått konkurs før eller etter offentliggjøringen. Det antas at alle regnskapene er tilgjengelige ved slutten av året. Dette kan muligens være et problem for de bedriftene som har problemer i utgangspunktet, hvor revisjonen av regnskapene kan ta lengre tid og det således vil gå enda lengre tid fra slutten av regnskapsåret til regnskapene blir gjort tilgjengelig.

I tidligere studier av nøkkeltallsbasert konkursprediksjon er det benyttet ulike definisjoner av konkurs. Begrepet «failure» er ofte benyttet, noe som kan innebære en ren legalistisk tolkning, mens andre for eksempel inkluderer bedrifter som har mottatt tilskudd fra det offentlige for å kunne opprettholde driften. Vi har benyttet en enkel tolkning av konkursbegrepet og kun inkludert bedrifter hvor det er åpnet konkurs. Bakgrunnen for dette valget er tilgangen til datamateriale. Når formålet er å estimere en modell for å predikere konkurstrisiko, kan det argumenteres for at konkursbegrepet burde vært utvidet. Dette kommer av at bedrifter kan ha velvillige kreditorer, som i enkelte situasjoner går med på løsninger for å opprettholde driften og unngå å ta tap i tilfeller der bedriften egentlig er konkurs.

Det kan også argumenteres for at tidsperioden som ligger til grunn er noe kort, spesielt med tanke på de spesielle omstendighetene knyttet til den globale finanskrisens utvikling de senere årene. Som følge av finanskrisen fikk man en økning i antall konkurser i 2009 (se figur 1.1). Det ser imidlertid ut til at dette har stabilisert seg på et noe lavere nivå per 4. kvartal 2011, dog på et noe høyere nivå enn før finanskrisen inntraff.

Tabell 4.2 og 4.3 viser fordelingen av hvilke bransjer henholdsvis ikke-konkursbedriftene og konkursbedriftene i datautvalget representerer. Som det fremgår av tabellen består utvalget av ikke-konkursbedrifter av langt flere bransjer enn konkursbedriftene. Dette kan påvirke modellene til å ikke være så generelle som vi hadde ønsket i utgangspunktet.

Som nevnt tidligere kan størrelsesvariabelen utgjøre et problem hvis man sammenligner bedrifter med stor variasjon i størrelse. En stor bedrift vil gjøre betydelig utslag på størrelsesvariabelen og redusere konkurssannsynligheten i modellen. Selv en bedrift som har problemer vil kunne få en lav konkurssannsynlighet grunnet størrelsen på bedriften. Betydningen av størrelsen i modellen kan ha blitt påvirket av at snittstørrelsen blant

konkursbedriftene er mindre enn blant ikke-konkursbedriftene. Vi testet imidlertid dette ved å ta ut de største bedriftene i utvalget av ikke-konkursbedrifter. Regresjonen resulterte i omtrent den samme modellen, så vi forkastet denne hypotesen. Vi har likevel anbefalt å sammenligne bedrifter av omtrentlig samme størrelse for unngå en eventuell problematikk rundt størrelsesvariabelen.

5.3 Forslag til videre forskning

Modellene i denne utredningen er utelukkende estimert på bakgrunn av data fra offentlige regnskap. Det er imidlertid rimelig å anta at bedriftenes konkurssannsynlighet også vil påvirkes av makroøkonomiske forhold. Bernhardsen (2001) er inne på temaet, men ellers er det, så vidt oss bekjent, forsket lite på dette området med norske data. Jacobsen og Kloster (2005) finner imidlertid at antall konkurser i Norge er påvirket av konkurranseevne, realrente, konjunktursvingninger, samt utvikling i fortjenestemarginer. Det kunne derfor vært interessant å estimere en konkursprediksjonsmodell hvor disse makroindikatorne inkluderes i tillegg til regnskapsdata. For at dette skal være hensiktsmessig bør nok tidsperioden utvides noe sammenlignet med denne utredningen, samtidig som man bør ha data på månedsbasis.

6. Referanser

6.1 Bøker

- ALDRICH, J. H. & NELSON, F. D. 1984. *Linear probability, logit and probit models*, Beverly Hills, Sage
- ALTMAN, E. I., AVERY, R. B., A., E. R. & SINKEY JR, J. F. 1981. *Applications of Classification Techniques in Business, Banking and Finance*, Greenwich, Connecticut, Jai Press Inc.
- EKLUND, T. & KNUTSEN, K. 1997. *Regnskapsanalyse med årsoppgjør: aktiv bruk av regnskapet*, Oslo, Universitetsforlaget AS.
- JOLLIFFE, I. T. 1986. *Principal component analysis*, New York, Springer.
- KELLER, G. 2009. *Managerial Statistics*, Mason, USA, South-Western Cengage Learning.
- KLEINBAUM, D. G. 1994. *Logistic regression: a self-learning text*, New York, Springer.
- MANLY, B. F. J. 2005. *Multivariate statistical methods: A primer*, Boca Raton, Fla., Chapman and Hall/CRC.
- SYDSÆTER K. & ØKSENDAL B. 2006. *Lineær algebra. 4. utg.* Oslo, Gyldendal Norsk Forlag

6.2 Artikler

- ALTMAN, E. I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance*, 23, 589-609.
- ALTMAN, E. I. 1977. ZETA-Analysis: A new model to identify bankruptcy risk of corporation. *Journal of Banking and Finance*, 29-54.
- BEAVER, W. H. 1966. Financial Ratios as Predictors of Failure. *Journal of Accounting Research*, 4, 71-111.
- BROWN, M. B. & FORSYTHE, A. B. 1974. Robust Tests for the Equality of Variances. *Journal of the American Statistical Association*, 69, 364 - 367.
- HOSMER, D. W., TABER, S. & LEMESHOW, S. 1991. The Importance of Assessing the Fit of Logistic Regression Models: A Case Study. *American Journal of Public Health*, 81, 1631-1635.

-
- JACOBSEN, D.H. & KLOSTER T.B. 2005. Hva påvirker konkursutviklingen? *Penger og kreditt nr. 3/2005*, 206-214.
- JOY, O. M. & TOLLEFSON, J. O. 1975. On the financial applications of discriminant analysis. *Journal of Financial & Quantitative Analysis*, 10, 723-739.
- LE CESSIE, S. & VAN HOUWELINGEN, T. J. 1991. A Goodness-of-fit itest for Binary Regression Models, Based on Smoothing Methods. *Biometrics*, 47, 1267-1282.
- OHLSON, J. A. 1980. Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18, 109-131.
- SCOTT, J. 1981. The Probability Of Bankruptcy: A Comparison of Empirical Predictions and Theoretical Models. *Journal of Banking and Finance*, 317-344.
- SKOGSVIK, K. 2005. On the Choice-Based Sample Bias in Probabilistic Business Failure Prediction *SSE/EFI Working Paper Series in Business Administration*
- TAMARI, M. 1966. Financial Ratios as a Means of Forecasting Bankruptcy. *Finanzwirtschaftliche Kennzahlen als Mittel zur Vorhersage von Insolvenzen.*, 6, 15-34.
- TUFTE, P. A. 2000. En intuitiv innføring i logistisk regresjon. *Statens Institutt for Forbruksforskning Arbeidsnotat nr. 8 - 2000*.
- WILCOX, J. W. 1976. The Gambler's Ruin Approach to Business Risk. *Sloan Management Review*, 18, 33-46.
- ZMIJEWSKI, M. E. 1984. Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22, 59-82.

6.3 Utredninger

- BERNHARDBSEN, E. 2001. *A Model of Bankruptcy Prediction*. Cand. Oecon, University of Oslo.
- FINANSKRISEUTVALGET. 2011. *Bedre rustet mot finanskriser*. Norges offentlige utredninger 2011:1. Oslo, Finansdepartementet.
- HALLETT, D. C. 1999. *Goodness of fit tests in logistic regression*. Master of Science, University of Toronto.
- LIU, Y. 2007. *On goodness of fit of logistic regression model. An abstract of a dissertation*. Doctor of Philosophy, Kansas State University.
- SJØVOLL, E. 1999. *Assessment of Credit Risk in the Norwegian Business Sector*. Cand. Polit, University of Bergen.

SKOGSVIK, K. 1988. *Prognos av finansiell kris med redovisningsmått : en jämförelse mellan traditionell och inflationsjusterad redovisning*, Stockholm, Handelshögskolan i Stockholm. Ekonomiska forskningsinstitutet.

6.4 Internettider

1: Statistisk Sentralbyrå, Statistikkbanken,
<http://statbank.ssb.no/statistikkbanken/Default_FR.asp?PXSid=0&nvl=true&PLanguage=0&tilside=selectvarval/define.asp&Tabellid=08563>, (Nedlastet 23.02.2012)

2: Brønnøysundregisteret, Pressemelding årsstatistikk fra Konkursregisteret 2008
<<http://www.brreg.no/presse/pressemeldinger/2009/01/konkurs2008.html>>, (Nedlastet 23.02.2012)

3: Brønnøysundregisteret, Pressemelding årsstatistikk 2008
<http://www.brreg.no/presse/pressemeldinger/2009/01/fr-statistikk_08.html>, (Nedlastet 23.02.2012)

4: Brønnøysundregisteret, Pressemelding årsstatistikk 2009,
<http://www.brreg.no/presse/pressemeldinger/2010/01/fr-statistikk_09.html>, (Nedlastet 23.02.2012)

5: Brønnøysundregisteret, Pressemelding årsstatistikk 2010,
<http://www.brreg.no/presse/pressemeldinger/2011/01/fr-statistikk_10.html>, (Nedlastet 23.02.2012)

6: Brønnøysundregisteret, Pressemelding årsstatistikk 2011,
<http://www.brreg.no/presse/pressemeldinger/2012/01/fr-statistikk_n.html>, (Nedlastet 23.02.2012)

6.5 Andre referanser

Konkursloven. 1984. *Lov om gjeldsforhandling og konkurs av 6. august 1984 nr. 58*. Tilgjengelig fra <<http://www.lovdatab.no>> (Nedlastet 26.02.2012)

Torgnesvik, L. 2012. E-post med konkursstatistikk for aksjeselskaper. Konsulent, Konkursregisteret.

7. Appendiks

7.1 Grenseverdier nøkkeltall

Nøkkeltall	Grenseverdier	Antall over grenseverdi	Prosentvis
FM	+/-100 %	24	0.46%
TKR(1)	-50 %, +100 %	20	0.38%
TKR(2)	-50 %, +100 %	26	0.50%
TKR(3)	-200 %, +400 %	40	0.76%
TKR(4)	-200 %, +400 %	44	0.84%
TKR(5)	-40 %, +80 %	31	0.59%
TKR(6)	-40 %, +100 %	34	0.65%
TKR(7)	-300 %, +500 %	40	0.76%
TKR(8)	-200 %, +600 %	39	0.75%
EKR(1)	-300 %, +600 %	76	1.45%
EKR(2)	-300 %, +400 %	75	1.43%
EKR(3)	-300 %, +600 %	76	1.45%
EKR(4)	-300 %, +400 %	74	1.42%
LA	Ingen grenser	0	0.00%
AA	+25 %	50	0.96%
RK(1)	+40 %	16	0.31%
RK (2)	+400 %	94	1.80%
SK	+/-300 %	58	1.11%
ABA	-30 %, +100 %	14	0.27%
SKOF	+/-1000 %	148	2.83%
OHE(1)	0, +10	24	0.46%
OHE(2)	+/-100	16	0.31%
OHE(3)	+/-100	20	0.38%
OHKFO	0, +5	29	0.55%
OHV(1)	0, +1,5	22	0.42%
OHV(2)	0, +1	23	0.44%
OHK	0, +1,5	24	0.46%
OHLO	0, +12	27	0.52%
OHO	0, +8	26	0.50%
OHAK	+/-15	15	0.29%
STAK	+/-2	2	0.04%
LG ₁ (1)	+0,80	7	0.13%
LG ₁ (2)	+3	13	0.25%
LG ₂ (1)	+2	2	0.04%
LG ₂ (2)	+10	27	0.52%
LG ₃ (1)	+6	22	0.42%

Nøkkeltall	Grenseverdier	Antall over grenseverdi	Prosentvis
KSG(1)	-3, +5	10	0.19%
KSG(2)	-100, +500	173	3.31%
KSG(3)	+1,5	29	0.55%
AVA	+0,7	40	0.76%
AAK	+1	2	0.04%
AAM	Ingen grenser	0	0.00%
AVDM	Ingen grenser	0	0.00%
ST(1)	Ingen grenser	0	0.00%
ST(2)	Ingen grenser	0	0.00%
SO	Ingen grenser	0	0.00%
AKGJ	Ingen grenser	0	0.00%
ALEV	+55 %	36	0.69%
ARGJ	+80 %	34	0.65%
VTI	+500 %	49	0.94%
VEI	+200 %	51	0.98%
VEK	+500%	75	1.43%
VKGJ	+500 %	23	0.44%
VLEV	+500 %	66	1.26%
VRGJ	+700 %	134	2.56%
VAM	+500 %	89	1.70%
VVDM	+500 %	72	1.38%

7.2 Egenverdier og egenvektorer

Rentabilitet PC1		Kortsiktig finansiering PC2		Driftskapital PC3	
TKR(7)	0.2643	KSG(1)	0.2709	AAK	0.3691293
TKR(4)	0.2631	ST(2)	0.2283	STAK	0.3676003
TKR(3)	0.2621	SO	0.2111	LG ₃ (1)	0.302135
TKR(8)	0.2592	KSG(3)	0.2098	LG ₂ (2)	0.2953493
EKR(3)	0.2531	LG ₂ (2)	0.1954	LG ₂ (1)	0.2800
TKR(6)	0.2530	LG ₃ (1)	0.1926	LG ₁ (2)	0.1856846
EKR(1)	0.2529	OHE(2)	-0.1942	LG ₁ (1)	0.1747362
TKR(2)	0.2521	OHE(3)	-0.2038	AA	-0.22084
EKR(2)	0.2511	OHE(1)	-0.2297	AVDM	-0.245357
TKR(5)	0.2487	ALEV	-0.2341	AAM	-0.305892
TKR(1)	0.2462	AKGJ	-0.2867		
EKR(4)	0.2425				
Egenverdi	11.457	Egenverdi	6.549	Egenverdi	5.604
Andel	20.10%	Andel	11.49%	Andel	9.83%
Kumulativ	20.10%	Kumulativ	31.59%	Kumulativ	41.42%

Omløpshastighet PC4		Varelager PC5		Vekst PC6	
OHK	-0.2840	LG ₁ (1)	0.2928	VRGJ	-0.2740
OHKFO	-0.4378	LG ₁ (2)	0.2905	VTI	-0.274755
OHLO	-0.4473	LA	0.2419	VLEV	-0.297228
OHO	-0.4476	AVA	-0.4372	VVDM	-0.368771
		OHV(1)	-0.4418	VKGJ	-0.381392
		OHV(2)	-0.4672	VAM	-0.391308
				VEI	-0.4890
Egenverdi	3.433	Egenverdi	3.146	Egenverdi	2.8689993
Andel	6.02%	Andel	5.52%	Andel	5.03%
Kumulativ	47.44%	Kumulativ	52.96%	Kumulativ	58.00%

Kapitalstruktur PC7		Aktivitetsnivå PC8		Langsiktig finansiering PC9	
ARGJ	0.3055	KSG(2)	0.2405	RK(1)	0.3029
EKR(4)	0.1970	OHAK	-0.2129	RK(2)	0.2279
EKR(2)	0.1953	LG ₃ (1)	-0.2359	OHV(2)	-0.2022
EKR(3)	0.1916	LG ₂ (2)	-0.2392	LG ₁ (2)	-0.2156
EKR(1)	0.1914	ARGJ	-0.2518	OHV(1)	-0.2333
OHE(1)	-0.1952	RK(1)	-0.2554	OHK	-0.2468
OHE(3)	-0.2075	OHE(2)	-0.3189	KSG(2)	-0.2481
OHE(2)	-0.2291	OHE(3)	-0.3249	LG ₁ (1)	-0.2551
ST(2)	-0.2591			OHE(3)	-0.2975
ST(1)	-0.2657			OHE(2)	-0.3051
SO	-0.2931				
RK(2)	-0.3442				
Egenverdi	1.875	Egenverdi	1.702	Egenverdi	1.447
Andel	3.29%	Andel	2.99%	Andel	2.54%
Kumulativ	61.29%	Kumulativ	64.27%	Kumulativ	66.81%

Rentekostnader PC10		Størrelse PC11		Offentlige avgifter PC12	
ST(2)	0.2327	VEK	0.2195	SK	0.5753
ST(1)	0.2192665	KSG(1)	0.2154	SKOF	0.4933
OHV(2)	-0.1655615	SO	-0.1890	ABA	-0.1954
OHV(1)	-0.1723094	VRGJ	-0.2064	VKGJ	-0.2140
OHK	-0.2202015	ST(2)	-0.3358		
LG ₁ (2)	-0.247069	ST(1)	-0.3836		
LG ₁ (1)	-0.2591632				
RK(1)	-0.3761488				
RK(2)	-0.4362717				
Egenverdi	1.311	Egenverdi	1.254	Egenverdi	1.208
Andel	2.30%	Andel	2.20%	Andel	2.12%
Kumulativ	69.11%	Kumulativ	71.31%	Kumulativ	73.43%

Vekst kortsiktig finansiering PC13		Kostnader PC14	
VLEV	0.4334	ALEV	0.3821
VKGJ	0.3855	SO	-0.2362
SKOF	0.3308	SKOF	-0.3442
VVDM	-0.2215	LA	-0.5685
VEK	-0.2286		
VRGJ	-0.2698		
VAM	-0.3209		
Egenverdi	1.141	Egenverdi	1.032
Andel	2.00%	Andel	1.81%
Kumulativ	75.43%	Kumulativ	77.24%

7.3 Korrelasjon mellom nøkkeltall og komponent

Rentabilitet PC1		Kortsiktig finansiering PC2		Driftskapital PC3	
TKR(7)	0.8946	KSG(1)	0.6933	AAK	0.8738
TKR(4)	0.8906	ST(2)	0.5842	STAK	0.8702
TKR(3)	0.8870	SO	0.5403	LG ₃ (1)	0.7152
TKR(8)	0.8772	KSG(3)	0.5368	LG ₂ (2)	0.6991
EKR(3)	0.8566	LG ₂ (2)	0.5000	LG ₂ (1)	0.6628
TKR(6)	0.8562	LG ₃ (1)	0.4929	LG ₁ (2)	0.4396
EKR(1)	0.8559	OHE(2)	-0.4969	LG ₁ (1)	0.4136
TKR(2)	0.8532	OHE(3)	-0.5215	AA	-0.5228
EKR(2)	0.8499	OHE(1)	-0.5877	AVDM	-0.5808
TKR(5)	0.8417	ALEV	-0.5992	AAM	-0.7241
TKR(1)	0.8332	AKGJ	-0.7337		
EKR(4)	0.8210				

Omløpshastighet PC4		Varelager PC5		Vekst PC6	
OHKFO	-0.8111	LG ₁ (1)	0.5194	VEK	-0.3310
OHLO	-0.8288	LG ₁ (2)	0.5153	VRGJ	-0.4641
OHO	-0.8293	AVA	-0.7756	VTI	-0.4654
		OHV(1)	-0.7837	VLEV	-0.5034
		OHV(2)	-0.8287	VVDM	-0.6246
				VKGJ	-0.6460
				VAM	-0.6628
				VEI	-0.8282

Kapitalstruktur PC7		Aktivitetsnivå PC8		Langsiktig finansiering PC9	
ARGJ	0.4183	KSG(2)	0.3137	RK(1)	0.3644
EKR(4)	0.2697	LG3(1)	-0.3078	RK(2)	0.2741
EKR(2)	0.2674	LG2(2)	-0.3122	LG1(1)	-0.3068
ST(2)	-0.3547	ARGJ	-0.3286	OHE(3)	-0.3578
ST(1)	-0.3638	RK(1)	-0.3333	OHE(2)	-0.3670
SO	-0.4014	OHE(2)	-0.4161		
RK(2)	-0.4713	OHE(3)	-0.4239		

Rentekostnader PC10		Størrelse PC11		Offentlige avgifter PC12	
ST(2)	0.2664	VEK	0.2458	SK(1)	0.6324
ST(1)	0.2511	KSG(1)	0.2412	SKOF	0.5422
LG1(2)	-0.2829	SK(1)	0.2194		
LG1(1)	-0.2968	ARGJ	0.2055		
RK(1)	-0.4307	SO	-0.2116		
RK(2)	-0.4996	VRGJ	-0.2311		
		ST(2)	-0.3761		
		ST(1)	-0.4296		

Vekst kortsiktig finansiering PC13		Kostnader PC14	
VLEV	0,4630	ALEV	0,3881
VKGJ	0,4118	SKOF	-0,3496
VAM	-0,3428	LA	-0,5775

7.4 Innbyrdes korrelasjon blant valgte nøkkeltall

	TKR(6)	LA	RK(1)	RK(2)	SK	OHV(2)	OHO
TKR(6)	1.000						
LA	0.006	1.000					
RK(1)	-0.023	-0.030	1.000				
RK(2)	0.100	-0.017	0.297	1.000			
SK	0.051	-0.077	0.001	-0.013	1.000		
OHV(2)	-0.033	-0.228	0.106	-0.028	0.030	1.000	
OHO	-0.013	0.060	0.132	0.060	-0.072	0.227	1.000
AAK	0.077	-0.082	0.009	-0.036	0.017	0.251	0.055
ST(1)	-0.033	-0.096	0.149	0.034	-0.026	0.055	0.287
AKGJ	0.110	-0.053	-0.223	0.156	0.030	-0.105	-0.083
VLEV	0.026	-0.009	0.008	0.005	0.020	-0.021	0.049
VEI	0.201	0.007	0.031	0.001	-0.008	-0.034	0.024
LG ₁ (1)	0.225	0.158	-0.200	0.010	-0.008	-0.211	-0.006
KSG(2)	0.317	0.083	-0.280	-0.089	-0.017	-0.126	-0.016

	AAK	ST(1)	AKGJ	VLEV	VEI	LG ₁ (1)	KSG(2)
AAK	1.000						
ST(1)	-0.214	1.000					
AKGJ	-0.339	-0.259	1.000				
VLEV	-0.037	-0.007	0.009	1.000			
VEI	-0.020	-0.065	0.047	0.355	1.000		
LG ₁ (1)	0.245	-0.217	0.156	0.001	0.005	1.000	
KSG(2)	0.090	-0.089	0.116	0.009	0.030	0.324	1.000

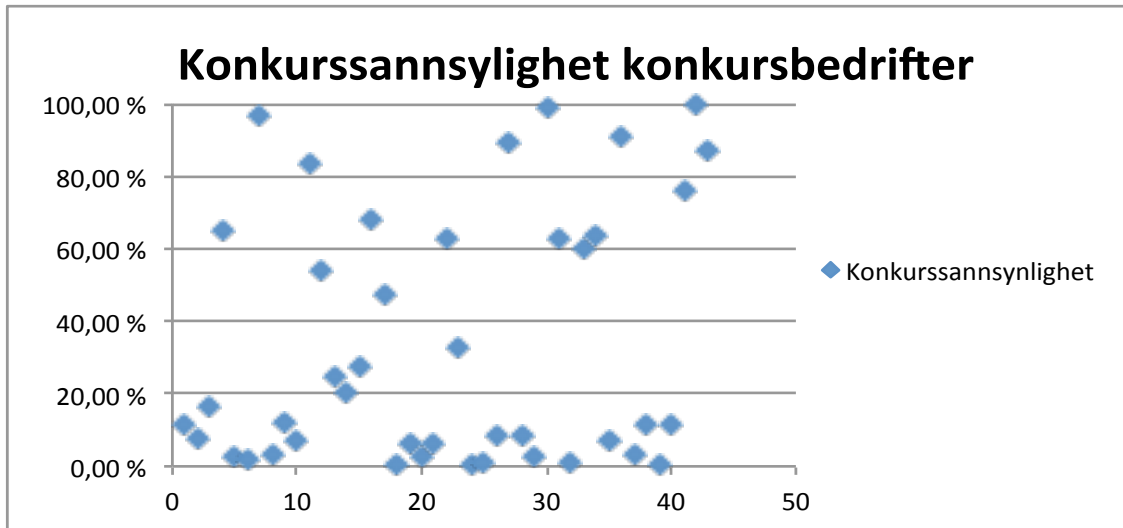
7.5 Levens test for lik varians

Levens test for lik varians						
Nøkkeltall	TKR(6)	LA	RK(1)	RK(2)	SK	OHE(2)
F-verdi	21.69	2.44	0.05	1.76	4.18	5.63
P-verdi	0.00	0.12	0.82	0.18	0.04	0.02
Nøkkeltall	OHE(3)	OHV(2)	OHO	AAK	ST(1)	AKGJ
F-verdi	24.36	4.92	0.00	0.07	36.74	11.21
P-verdi	0.00	0.03	0.99	0.80	0.00	0.00
Nøkkeltall	VLEV	VEI	LG ₁ (1)	ARGJ	KSG(2)	LG ₃ (1)
F-verdi	8.41	86.37	22.59	2.33	13.54	18.37
P-verdi	0.00	0.00	0.00	0.13	0.00	0.00

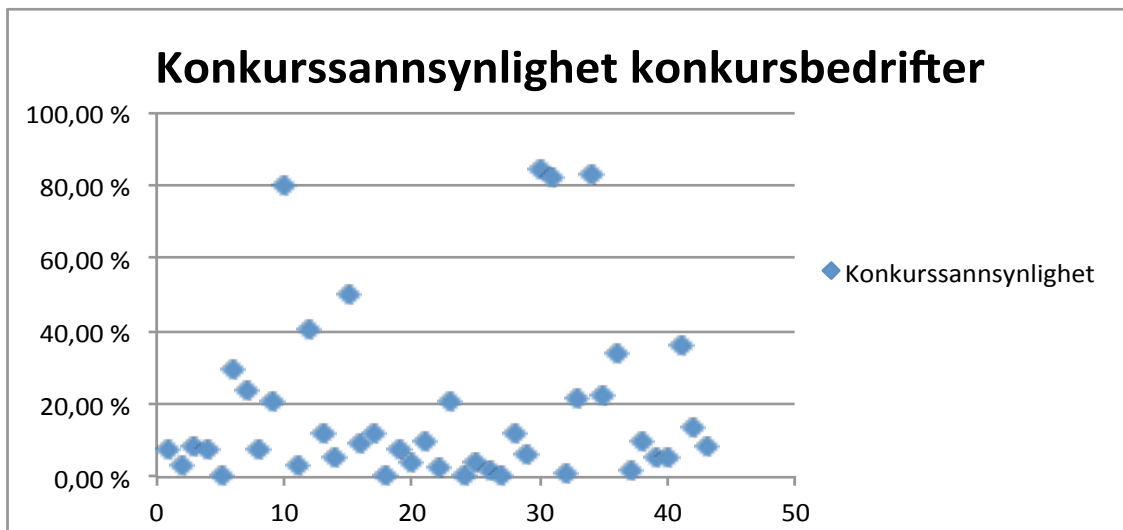
7.6 T-test for likt gjennomsnitt

T-test for likt gjennomsnitt konkurs/ikke-konkurs						
Nøkkeltall	TKR(6)	LA	RK(1)	RK(2)	SK	OHE(2)
T-verdi	9.96	1.51	-3.09	0.97	-0.30	-1.50
P-verdi	0.00	0.13	0.00	0.33	0.77	0.14
Nøkkeltall	OHE(3)	OHV(2)	OHO	AAK	ST(1)	AKGJ
T-verdi	-2.65	-2.63	-0.28	4.66	19.07	-10.70
P-verdi	0.01	0.01	0.78	0.00	0.00	0.00
Nøkkeltall	VLEV	VEI	LG ₁ (1)	ARGJ	KSG(2)	LG ₃ (1)
T-verdi	-1.64	-2.49	6.43	-4.49	10.23	9.82
P-verdi	0.10	0.01	0.00	0.00	0.00	0.00

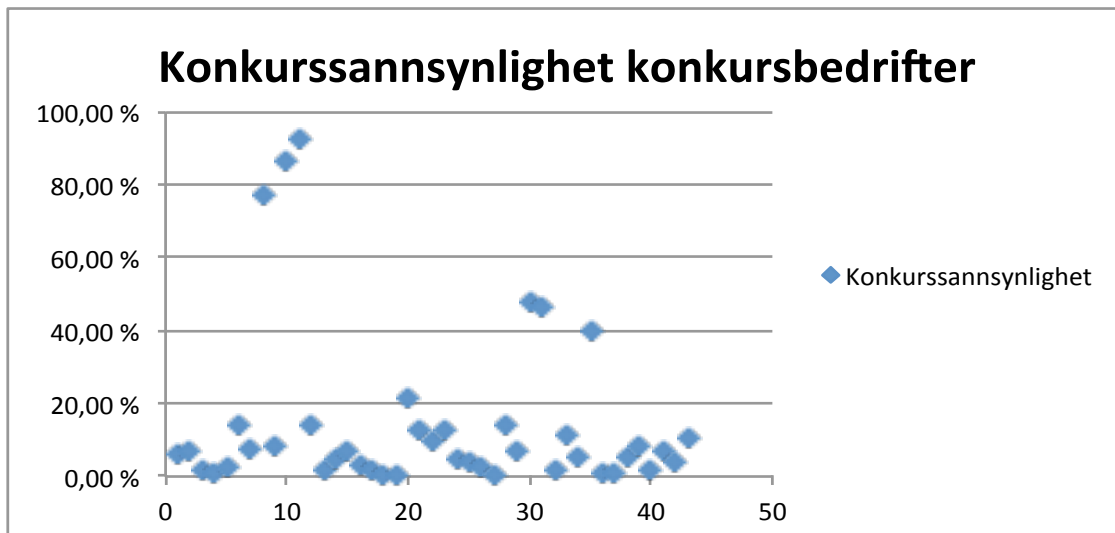
7.10 Konkursansynlighet for konkursbedrifter ett års prognoseperiode



7.11 Konkursansynlighet for konkursbedrifter to års prognoseperiode



7.12 Konkursansynlighet for konkursbedrifter tre års prognoseperiode



7.13 Fremgangsmåte beregning av konkurssannsynlighet

For å bruke modellen må de enkelte nøkkeltallene settes inn for å få logiten, L. Se tabell 4.7 for definisjon av nøkkeltallene.

Ett års prognoseperiode:

$$L = 16,7113 - 8,97576*TKR(6) + 4,26569*OHV(2) - 1,98331*ST(1) + 4,04992*AKGJ$$

To års prognoseperiode:

$$L = 13,5507 - 6,78894*TKR(6) + 9,85507*RK(1) - 5,25059*LG_1(1) - 1,56903*ST(1) + 3,02310*AKGJ$$

Tre års prognoseperiode:

$$L = 17,8382 - 4,34656*TKR(6) - 7,93162*LG_1(1) - 1,78603*ST(1) + 1,58257*VEI$$

Logiten må så regnes om til konkurssannsynlighet ved hjelp av denne formelen:

$$p^{prop} = \frac{1}{1+e^{(L)}}$$

Deretter må det korrigeres for utvalgsskjevhet:

$$P(\text{konkurs}) = \left[1 + \frac{(1-\pi)}{\pi} * \frac{0,0483}{(1-0,0483)} * \frac{1-p^{prop}}{p^{prop}} \right]^{-1}$$

Hvor π er andel konkursbedrifter i populasjonen som vil variere over tid. Vi har brukt 1,25% i denne utredningen. 0,0483 er andel konkursbedrifter i utvalget brukt for å lage modellen.