Titelblad

Titel: Konkursprognosticering af danske virksomheder

Uddannelse: Cand. Merc. Økonomistyring

Uddannelsessted: Aalborg Universitet

Semester: 10. Semester

Gruppe: 2

Vejleder: Allan Damsgaard

Afleveringsdato: 8. juni 2016

Normalsider: 99,93

Isabella Staun Bendtsen

Abstract

A company's bankruptcy is often related to large financial consequences for the shareholders and lenders, but often it is not the only stakeholders who are affected by the company's bankruptcy. This is why bankruptcy prediction models have been developed for decades, and serve as a way to try to predict the risk of company bankruptcy. The models are often based on financial ratios, and the preferred multivariate technique for the last couple of decades has been the logistic regression. Although the development of bankruptcy models is not a new phenomenon only a few models have been developed based on data from Danish companies, and this has been the motivation for preparing this thesis.

The thesis focus on how chosen financial ratios contribute to the prediction of bankruptcies in Danish companies and their contribution is analysed through a logistic regression. This focus resulted in that the thesis is based on a research problem according to Booth et al.'s (2008) method, which influences the structure of the thesis.

The analysis is based on a sample of 381 Danish companies where 271 of the companies are in normal operation, whereas the 110 companies are bankrupted. The total sample of 381 companies is divided into an analysis sample of 243 companies and a holdout sample of 138 companies. The chosen financial ratios consist of six ratios where five of them are Altman & Sabato's (2007) financial ratios, and these are Short-term debt/Equity book value, Cash/Total assets, EBITDA/ Total assets, Retained earnings/Total assets and EBITDA/Interest expenses. The last financial ratio is Equity book value/Total assets, which is included as a supplement to Altman & Sabato's (2007) financial ratios. The financial ratios are calculated from the companies' financial statements from 2011.

A descriptive statistics of the sample illustrates that the sample is not representative of the population if you look at the development in number of bankruptcies, the industrial distribution and type of business organisation, and additionally the two groups in the sample are not matched either.

The analysis is following Hair et al.'s (2010) six-step approach to multivariate model building for a logistic regression, and this leads to that a logistic regression model with two variables is estimated through a forward stepwise (Wald) method. In the first step the variable Retained earnings/Total assets is implemented in the logistic regression model, where after the variable

Cash/Total assets is implemented in the second step. The logistic regression model is subsequently assessed to have an acceptable *goodness-of-fit*. The interpretation of the logistic regression model results in the financial ratio Cash/Total assets has a positive relationship with the probability of companies going bankrupt, whereas the other variable Retained earnings/Total assets has a negative relationship with the probability of companies going bankrupt. The holdout sample is used to validate the logistic regression model, but a validation is not possible because of a low hit ratio and especially because the hit ratio is very low for the companies who are bankrupted.

The lack of validation indicates that a generalization of the logistic regression model is not possible, which probably is caused by the fact that the sample is not representative for the population. This reveals that there is a great development potential for the logistic regression model. These opportunities for improvement include optimizing and expanding the sample, a different method to choose the financial ratios and extending the statistical analysis with a correlation analysis. Furthermore it is interesting to consider if qualitative ratios should be included in the analysis, and if there is an interaction between some of the ratios.

In this thesis there is developed a bankruptcy prediction model, which illustrates that the financial ratio Cash/Total assets has a positive contribution to the prediction of bankruptcies and another financial ratio Retained earnings/Total assets has a negative contribution to the prediction of bankruptcies. The lack of validation of the logistic regression model results in the two financial ratios only contribute to the prediction of bankruptcies for the companies in the sample and not for all Danish companies.

Forord

Indledningsvis skal der lyde en tak til Allan Damsgaard for værdifuld, kompetent og kyndig vejledning ifm. udarbejdelsen af specialet.

Indholdsfortegnelse

1 Indledning

Aktionærer og långivere, der overvejer at investere i en virksomhed, vil drage nytte af, hvis det er muligt at kunne forudsige, om virksomheden vil gå konkurs inden for den nærmeste fremtid (Christensen & Nielsen, 2012). En konkursprognosticering af en given virksomhed vil kunne forhindre aktionærer og långivere i at foretage en investering, der formentlig vil gå tabt. Det skyldes, at en virksomheds konkurs ofte er forbundet med store økonomiske konsekvenser for aktionærer og långivere. På trods af de økonomiske konsekvenser for investorerne er disse parter langt fra de eneste, der påvirkes af en virksomheds konkurs. Alle virksomhedens interessenter vil hver især påvirkes af en konkurs, herunder fx ansatte, leverandører, ledelse, kunder og det offentlige (Ledernes Hovedorganisation, 1999). Disse konsekvenser indebærer bl.a., at de ansatte mister deres job, leverandørerne mister en indtægtskilde, og det offentlige mister arbejdspladser. Et eksempel på interessenters involvering kan ses i den meget omtalte konkurs af OW Bunker, hvor konkursboet efter OW Bunker nærmest blev begravet i henvendelser fra investorer, ledelse og andre interessenter (Erhardtsen et al., 2014).

Finanskrisen har efterhånden sluppet sit tag i økonomien og erhvervslivet, men alligevel går flere virksomheder konkurs. For første gang i tre år er antallet af konkurser i Danmark stigende, hvor i alt 4202 virksomheder drejede nøglen om i år 2015, hvilket er 5 % mere end året før (Simonsen, 2016). Denne situation beregner Martin Stabell, der er ekspert i risikovurderinger hos Bisnode, således:

"Dansk økonomi er stadig ikke oppe i fulde omdrejninger – og er muligvis endda ved at tabe lidt af luften. Der er betydelige udfordringer i mange brancher, og vi kan se på tallene, at der ikke skal mere end én dårlig måned til at vende en positiv udvikling til en negativ" (Simonsen, 2016).

Det skal dog påpeges, at antallet af konkurser er væsentligt lavere end i årene lige efter finanskrisen, men Stabell mener alligevel, at udviklingen er bekymrende. Det er især branchen landbrug, jagt, skovbrug og fiskeri, der er hårdt ramt, hvor antallet af konkurser i år 2015 steg med 60 % ift. året før (Simonsen, 2016). Det stigende antal af konkurser i Danmark øger interessen omkring konkursprognosticering, da det i denne situation vil få en større betydning, hvorvidt aktionærer og långivere er i stand til at kunne forudsige en virksomheds konkurs og derefter placere deres investering andetsteds. Det skyldes, at der med et stigende antal konkurser

alt andet lige er større risiko for, at en investering placeres i en virksomhed, der går konkurs inden for den nærmeste fremtid.

Konkursprognosticering er ikke et nyt fænomen, det har eksisteret i flere årtier. Mange forfattere har igennem de sidste 40-50 år udarbejdet forskellige modeller for at kunne forudsige virksomheders konkurs (Altman & Sabato, 2007). Blandt pionererne inden for konkursprognosticeringsmodeller er Beaver (1966) og Altman (1968), der har udviklet univariat og multivariat modeller til at forudsige konkurser vha. finansielle nøgletal (Altman & Sabato, 2007). Beavers (1966) univariat analyse indeholdte 30 forskellige nøgletal, der enkeltvis blev analyseret for at vurdere deres evne til at prognosticere en konkurs. Beaver (1966) betegnede det dog som økonomiske vanskeligheder i stedet for konkurs. Analysen blev foretaget på en stikprøve, der bestod af 79 virksomheder, der havde været i økonomiske vanskeligheder, og 79 sunde virksomheder (Beaver, 1966). Efter analysen af de 30 nøgletal blev de inddelt i seks kategorier, hvor Beaver (1966) identificerede det nøgletal, der havde færrest fejlklassifikationer, og dermed bedst adskilte de to grupper af virksomheder fra hinanden. Det bedste nøgletal i Beavers (1966) analyse var driftens Cash flow/Samlet gæld, da dette nøgletal kun fejlklassificerede 10 % af virksomhederne ud fra data et år før konkursen.

Altman (1968) var den første, der anvendte en multipel diskriminant analyse til at analysere konkursprognosticering. Analysen indeholdte 22 udvalgte nøgletal, og stikprøven bestod af 33 konkursramte virksomheder og 33 ikke-konkursramte virksomheder, der var udvalgt på baggrund af størrelse og branche. De 22 nøgletal blev inddelt i fem kategorier, og disse kategorier var likviditet, rentabilitet, gearing, soliditet og aktivitetsgrader. Derefter blev forskellige modeller med et nøgletal fra hver kategori estimeret og testet for klassifikationssikkerhed (Altman, 1968). Altmans (1968) multivariat analyse tillod i modsætning til Beavers (1966) univariat analyse, at modellen blev estimeret ud fra den kombination af nøgletallene, der bedst adskilte de to grupper af virksomheder. Den kombination af nøgletal, der bedst adskilte de to grupper af virksomheder, klassificerede 95 % af virksomhederne korrekt (Altman, 1968).

I de efterfølgende år var multipel diskriminant analyse den foretrukne teknik ift. konkursprognosticering, men mange forfattere gjorde opmærksom på, at der især var to forudsætninger ved multipel diskriminant analysen, som ofte ikke blev opfyldt ift. konkursprognosticering. Disse forudsætninger er, at de uafhængige variabler skal være

normalfordelte, og at de er lineære afhængige (Altman & Sabato, 2007). Ud fra disse problemer med multipel diskriminant analysen anvendte Ohlson (1980) for første gang en betinget logistisk regressionsmodel til konkursprognosticering. Den praktiske fordel ved at benytte logistisk regression er, at den ikke kræver de samme restriktive forudsætninger som multipel diskriminant analyse, og derudover tillader den uforholdsmæssige stikprøver (Altman & Sabato, 2007). Ohlson (1980) anvendte et datasæt med 105 konkursramte virksomheder og 2058 ikke-konkursramte virksomheder, og analysen blev baseret på ni indikatorer, herunder syv nøgletal og to binære variabler. Baggrunden for disse variabler var angiveligt, at det var dem, der hyppigst blev omtalt i litteraturen. Ohlsons (1980) model performede ift. klassifikationssikkerhed dårligere end de tidligere multiple diskriminant analyser, men der blev skabt et grundlag for at anvende logistisk regression frem for multipel diskriminant analyse. Fra et statistisk synspunkt skyldes det, at logistisk regression lader til at passe godt til karakteristikaene af konkursprognosticeringsproblemet, hvor den afhængige variabel er binær og med grupper, der er diskrete, ikke-overlappende og identificerbare (Altman & Sabato, 2007). Den logistiske regressionsmodel giver en score mellem 0 og 1, som svarer til sandsynligheden for konkurs for en given virksomhed. Derudover kan de estimerede koefficienter fortolkes separat som vigtigheden eller signifikansen af hver uafhængig variabel. Efter Ohlsons (1980) undersøgelse anvender størstedelen af den akademiske litteratur logistiske regressionsmodeller til konkurs-prognosticering (Altman & Sabato, 2007).

Ud fra ovenstående er konkursprognosticering analyseret i gennem mange år og af mange forskellige forfattere, men på trods af en grundig søgning er det svært at finde tilfælde af disse undersøgelser, der er foretaget på baggrund af data fra danske virksomheder. Den begrænsede tilstedeværelse af danske undersøgelser har betydning for aktionærer og långivere, der har interesse i danske virksomheder, da de således ikke har mulighed for at kunne benytte sig af en given konkursprognosticeringsmodel til at vurdere sin investering. Derfor er det interessant at udarbejde en konkursprognosticeringsmodel ud fra danske virksomheder, der bygger på nøgletal ligesom de tidligere undersøgelser. Derudover følger indeværende speciale tendensen i den akademiske litteratur og benytter sig af en logistisk regression som den statistiske teknik til analysen.

2 Problemfelt

Formålet med dette afsnit er at specificere specialets omdrejningspunkt, hvorfra der udledes en problemformulering. Derefter uddybes problemformuleringen, og der foretages en afgrænsning af rammerne for specialet.

Den begrænsede tilstedeværelse af konkursprognosticeringsmodeller, der er udarbejdet på baggrund af data fra danske virksomheder, og betydningen heraf har skabt en interesse hos forfatteren af indeværende speciale. Det skyldes, at en tidlig indikation af at en virksomhed formentlig vil gå konkurs, vil gøre det muligt at tage hånd om problemerne internt i virksomheden, og dermed evt. undgå virksomhedens konkurs. En evt. forhindring af en virksomheds konkurs på baggrund af en konkursprognosticeringsmodel kan potentielt medføre store samfundsøkonomiske besparelser. Derudover vil en mulighed for at kunne forudsige en konkurs hos en virksomhed kunne forhindre økonomiske tab for aktionærer, långivere og andre interessenter. En konkursprognosticeringsmodel vil således udgøre et vigtigt værktøj, hvis der skal foretages beslutninger ift. en given virksomhed, uanset om det vedrører karriemuligheder, samarbejdsaftaler eller investeringer. Derudover vil en mulighed for at kunne forudsige en virksomheds konkurs medføre, at en konkurs sjældent ville komme bag på de involverede interessenter.

Risikoen for konkurs og følgerne heraf får endnu større betydning, idet antallet af konkurser i Danmark, jf. afsnit 1 *Indledning*, for første gang i tre år er stigende, hvilket if. Stabell er en bekymrende udvikling. Det gør omdrejningspunktet for indeværende speciale endnu mere interessant og aktuelt, da en stigning i antallet af konkurser, øger den generelle risiko for, at virksomheder går konkurs. Udarbejdelsen af en konkursprognosticeringsmodel for danske virksomheder tager udgangspunkt i finansielle nøgletal, da nøgletal, jf. afsnit 1 *Indledning*, har været en del af konkursprognosticering helt tilbage til de allerførste analyser af Beaver (1966) og Altman (1968). Derudover er udarbejdelsen og tilgængeligheden af virksomheders finansielle nøgletal forholdsvis let og sammenlignelig, hvilket øger datagrundlaget til analysen. Den anvendte statistiske teknik til udarbejdelsen af konkursprognosticeringsmodellen er logistisk regression, idet den stemmer overens med karakteristikaene af konkursprognosticerings-problemet. Derudover er logistisk regression siden Ohlsons (1980) undersøgelse den mest anvendte metode inden for den akademiske litteratur på området (Altman & Sabato, 2007).

Indeværende speciale bærer præg af et undersøgelsesproblem, da det kredser om et konceptuelt problem, der opstår, fordi der er en uvidenhed omkring en given situation, hvorefter forfatterens intention er at øge sin viden omkring det pågældende ved at besvare et spørgsmål. Ovenstående har ledt til følgende undersøgelsesproblem, der ligeledes er specialets problemformulering.

Hvordan bidrager udvalgte nøgletal via en logistisk regression til forudsigelse af konkurser i danske virksomheder?

Problemformuleringens karakteristika af et undersøgelsesproblem er en del af strukturen i Booth et al. (2008), der uddybes i afsnit 5.1 *Undersøgelsesproblem*, hvor dets anvendelse og funktion ligeledes bearbejdes yderligere.

Ud fra problemformuleringen skal der foretages en statistisk analyse af danske virksomheder for at afgøre, om det er muligt at kunne forudsige konkurser. Dermed afgrænses der fra at inddrage udenlandske virksomheder til besvarelsen af problemformuleringen. Den statistiske analyse foretages dog kun på en stikprøve af danske virksomheder, idet det tidsmæssigt ikke har været muligt at bearbejde alle danske virksomheder i specialet. Indsamlingen af stikprøven uddybes i afsnit 5.2.1 *Dataindsamling*, hvorimod stikprøven præsenteres i afsnit 7.1.1 *Databearbejdning*. Den statistiske analyse skal jf. problemformuleringen udarbejdes via en logistisk regression, hvorfor der afgrænses fra at inddrage andre statistiske teknikker. Argumentationen for anvendelse af den logistiske regression som den mest velegnede teknik til besvarelse af problemformuleringen fremføres i afsnit 6.4 *Operationalisering af teorien*, efter at teorien om den logistiske regression er præsenteret.

Problemformuleringen specificerer, at den statistiske analyse skal foretages på baggrund af udvalgte nøgletal, og disse nøgletal er de fem nøgletal, der indgår i Altman & Sabatos (2007) undersøgelse samt soliditetsgraden. Altman & Sabato (2007) udarbejdede en konkurs-prognosticeringsmodel, der var specielt egnet til små og mellemstore virksomheder, men det er selve forarbejdet med udvælgelsen af de anvendte nøgletal, der er interessant for indeværende speciale. Altman & Sabatos (2007) forarbejde med de udvalgte nøgletal uddybes i afsnit 6.2.1 *Udvælgelse af nøgletallene*, hvorefter alle de seks udvalgte nøgletal uddybes i afsnit 6.2.3 *Nøgletallenes betydning*. Efterfølgende argumenteres der for valget af de seks nøgletal til indeværende speciale i afsnit 6.4 *Operationalisering af teorien*, samt årsagen til at soliditetsgraden inddrages som supplement til Altman & Sabatos (2007) fem nøgletal. De seks udvalgte nøgletal er alle kvantitative nøgletal, hvormed der afgrænses fra at inddrage kvalitative

nøgletal til besvarelsen af problemformuleringen. Denne afgrænsning skyldes ligeledes, at der sjældent inddrages kvalitative nøgletal i udarbejdelsen af konkursprognosticeringsmodeller. Det fremgår i afsnit 1 *Indledning*, hvor Beaver (1966) og Altman (1968) også udelukkende anvendte finansielle nøgletal til deres pionerundersøgelser, og det kan skyldes, at de kvantitative nøgletal er nemmere at håndtere i en statistisk analyse.

Problemformuleringen indledes med hvordan, hvilket stemmer overens med anbefalingen fra Booth et al. (2008) om, at en problemformulering bør starte med hvorfor eller hvordan. Det skal dog understreges, at hvordan nøgletallene bidrager kan resultere i, at de ikke bidrager. Der kan således ud fra problemformuleringen ikke antages, at alle de udvalgte nøgletal bidrager til forudsigelsen af konkurser i danske virksomheder.

Tidsmæssigt afgrænses den statistiske analyse til data fra år 2011, da der i det år er flest virksomheder, der er en del af det anvendte datasæt, hvilket uddybes i afsnit 7.1.1 *Databearbejdning*. Derfor medfører anvendelsen af år 2011, at analysen bygger på den størst mulige stikprøve. Anvendelsen af år 2011 medfører dog, at specialet bygger på data der er flere år gamle, samtidig med at det ud fra et enkelt år ikke er muligt at foretage en analyse af virksomhedernes udvikling.

3 Begrebsliste

Denne begrebsliste fungerer som en oversigt over de hyppigste anvendte statistiske begreber gennem specialet. Desuden kan den fungere som et opslagsværk, som læseren gentagende gange kan vende tilbage til, derfor er den struktureret i alfabetisk rækkefølge. Begrebslisten er udarbejdet for at undgå at bryde sammenhængen og forståelsen i den skrevne tekst, idet det ellers havde været nødvendigt at definere og uddybe begreberne løbende.

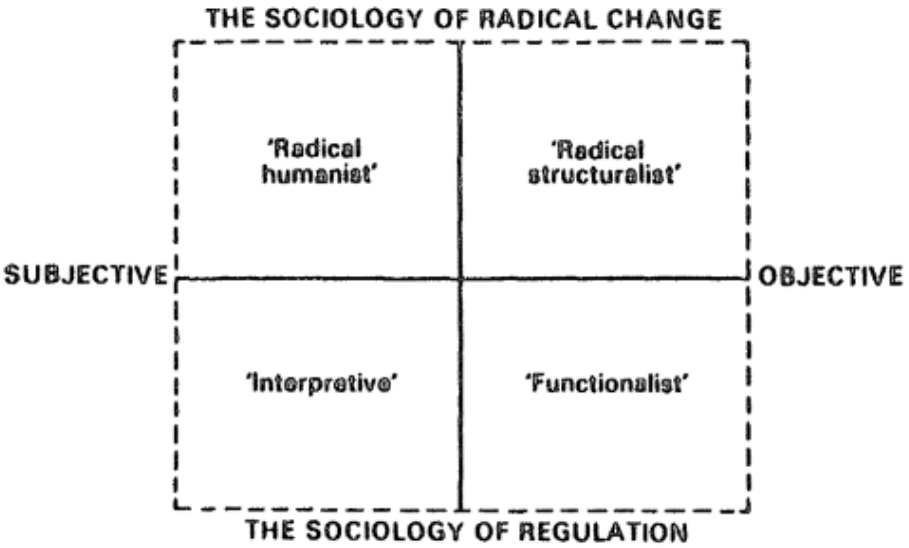
- **Den tilfældige variabel** – En lineær kombination af variabler dannet i den multivariate teknik ved at anvende empirisk data til et sæt af variabler, der er udvalgt af forskeren (Hair et al., 2010).
- **Goodness-of-fit** – Et udtryk for hvor godt den statistiske model passer til datasættet (Hair et al., 2010).
- **Hitrate** – Den procentdel af observationerne der er korrekt klassificeret af den logistiske regressionsmodel. Den er kalkuleret ved at tage antallet af observationer i diagonalen af klassifikationsmatrixen og dividere med det totale antal af observationer (Hair et al., 2010).
- **Hold-out stikprøve** – En gruppe af observationer der ikke anvendes til at estimere den logistiske regressionsmodel, men som i stedet bruges til at validere den logistiske regressionsmodel (Hair et al., 2010).
- **Klassifikationsmatrix** – Et værktøj til at vurdere den forudsigende evne af den logistiske regressionsmodel, og den er udarbejdet ved at krydstabulere faktisk gruppemedlemsskab med forudsagt gruppemedlemsskab. Matrixen indeholder diagonalt det antal af klassifikationer, der er korrekt klassificeret, hvorimod de ikke-diagonale antal repræsenterer fejlklassifikationerne (Hair et al., 2010).
- **Logistisk regression** – En speciel form for regression, hvor den uafhængige variabel er en binær og ikke-tal-variabel (Hair et al., 2010).
- **Maximum chance kriterie** – Mål for den forudsigende nøjagtighed i klassifikationsmatrixen, og det sammenligner hitraten med den procentdel af observationerne, der er i den største gruppe (Hair et al., 2010).
- **Maximum likelihood estimator (MLE)** – En procedure der anvendes gentagende gange for at finde de mest egnede estimater for koefficienterne i logistisk regression (Hair et al., 2010). MLE kan oversættes til maksimal sandsynligheds estimator.

- **Median** – Udtryk for det midtpunkt, der deler antallet af observationer i to lige store dele. For at beregne medianen skal observationerne kunne rangeres fra lav til høj, hvorefter medianen er værdien af den midterste observation (ved et lige antal observationer er det gennemsnittet mellem de to midterste) (Clement & Ingemann, 2011).
- **Multikollinearitet** – Udtrykker i hvilket omfang en variabel kan forklares ud fra andre variabler i analysen (Hair et al., 2010).
- **Multivariat** – Betyder flere variabler.
- **Odds** – Sandsynlighedsforholdet for, at en begivenhed forekommer ift. sandsynligheden for, at begivenheden ikke forekommer (Hair et al., 2010).
- **Outlier** – En observation der har en væsentlig forskel mellem den aktuelle værdi for den afhængige variabel og den forudsagte værdi. Derudover kan en outlier også være en observation, der er væsentlig forskellig ift. enten den afhængige eller uafhængige variabel (Hair et al., 2010).
- **Proportional chance kriterie** – Et kriterie for at vurdere hitraten, hvor den gennemsnitslige sandsynlighed for klassificering er beregnet ud fra alle gruppestørrelser (Hair et al., 2010).
- **Pseudo R^2** – En værdi for overordnet model fit der kan kalkuleres for den logistiske regression (Hair et al., 2010).
- **Standardafvigelse** – Udtryk for hvor tæt stikprøvens observationer ligger op ad hinanden. Den beregnes ved, at hver observations værdi fratrækkes middelværdien og kvadreres. Derefter summeres alle værdierne, hvorefter der divideres med antallet af observationer minus en for til sidst at tage kvadratroden af dette. Formlen er: $s = \sqrt{\frac{\sum(Y_i - \overline{Y})^2}{n-1}}$ (Clement & Ingemann, 2011).
- **Standardfejlen** – Den udtrykker usikkerheden på gennemsnittet af *n* målinger (Bendsen, 2016).
- **Statistisk inferens** – Indebærer at der skønnes en parameters størrelse.
- **Wald teststørrelsen** – Den teststørrelse der benyttes i logistisk regression for at vurdere signifikansen af de logistiske koefficienter (Hair et al., 2010).

4 Videnskabsteori

Dette afsnit præsenterer forfatterens videnskabsteoretiske ståsted, der afspejles i et af Burrell & Morgans (1979) fire paradigmer.

Burrell & Morgans (1979) opfattelse af samfundet er udgangspunktet for forfatterens videnskabsteoretiske ståsted. Burrell & Morgan (1979) har udarbejdet et klassifikationssystem ved at kombinere antagelsen om, at ændringer i samfundet kan foretages radikalt eller gennem regulativer med antagelsen om subjektivitet eller objektivitet. En kombination af de to dimensioner, hhv. radikal eller regulativ ændring samt subjektivitet eller objektivitet, har resulteret i en matrice. Denne matrice består af fire paradigmer, hhv. radikal humanisme, radikal strukturalisme, interpretivisme og funktionalisme, og de fire paradigmer kan ses i Figur 1.

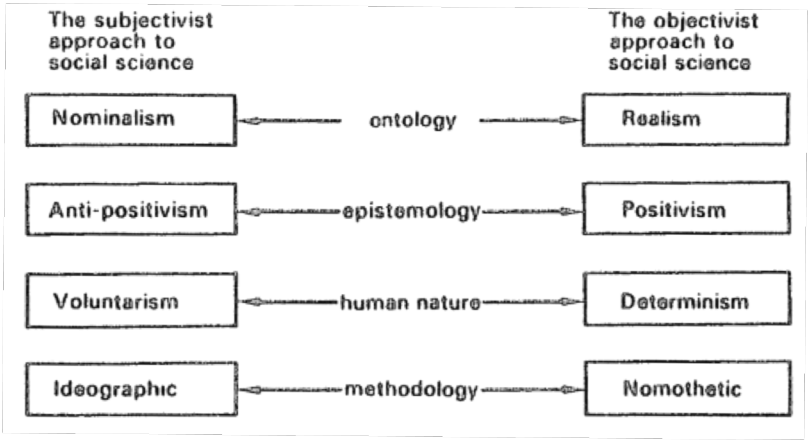


Figur 1 – Fire paradigmer til analyse af social teori (Burrell & Morgan, 1979, s. 22)

Hvert af de fire paradigmer bygger på sine egne unikke grundantagelser, hvilket har medført fire vidt forskellige socialvidenskabelige virkeligheder, der hver repræsenterer sin måde at anskue verden på. Derfor specificerer Burrell & Morgan (1979), at det ikke er muligt at kombinere de enkelte paradigmer, da de er modstridende i deres grundantagelser. Burrell & Morgan (1979) konkluderer således, at videnskaberne inden for socialvidenskaben kun befinder sig inden for et af de fire paradigmer.

Hvilket paradigme, der afspejler forfatterens videnskabsteoretiske ståsted, afgøres ud fra de to dimensioner, dvs. der skal tages stilling til, hvorvidt forfatterens udgangspunkt er subjektivt eller

objektivt, samt hvordan ændringer i samfundet opfattes. Forfatterens grundantagelse ift. ontologi, epistemologi, menneskelig natur samt metodologi er med til at præcisere, hvorvidt forfatterens udgangspunkt er subjektivt eller objektivt. De fire begreber indgår i Figur 2, der ligeledes illustrerer forskellen mellem et subjektivt eller objektivt udgangspunkt. En stillingtagen til ontologi, epistemologi, menneskelig natur samt metodologi øger ligeledes forfatterens bevidsthed omkring de valg, der foretages i udarbejdelsen af specialet.



Figur 2 – Et skema til at analysere antagelser om karakteren af samfundsvidenskaben (Burrell & Morgan, 1979, s. 3)

Ontologi omhandler "det værende", dvs. hvordan verden opfattes. Forfatterens synspunkt er her præget af realisme, der medfører, at virkeligheden eksisterer uden individets erkendelse (Burrell & Morgan, 1979). Dette synspunkt afspejler således objektivitet.

Epistemologi omhandler derimod, hvordan "det værende" erkendes, dvs. hvilke briller forfatteren tager på, når "det værende" analyseres. Der fokuseres således på forholdet mellem undersøgeren og det, der undersøges, samt hvordan viden kan anskaffes om det, der undersøges. Epistemologi betegnes derfor også som erkendelseslære eller erkendelsesteori, der omhandler den menneskelige erkendelses natur, grænser og betingelser. Forfatteren er præget af positivisme, hvilket indebærer, at erkendelsen udelukkende bygger på positive kendsgerninger. Positive kendsgerninger kendetegnes ved, at det er noget, der kan iagttages og observeres udefra (Burrell & Morgan, 1979). Forfatterens positivistiske synspunkt afspejler ligeledes objektivitet.

Den menneskelige natur specificerer, om individet har indflydelse på begivenheder. Determinisme, som præger forfatteren, hævder, at et menneskes gøren er givet på forhånd ud fra

den omverden og situation, som det befinder sig i (Burrell & Morgan, 1979). Forfatterens deterministiske synspunkt afspejler objektivitet.

Metodologi omhandler, hvordan undersøgeren helt praktisk arbejder for at opnå en forståelse samt skaffe viden og indsigt i genstandsfeltet. Forfatteren anvender igennem specialet en nomotetisk metode, der prioriterer vigtigheden af at basere forskning på systematiske værktøjer og teknikker. Den nomotetiske metode er kvantitativ og beskæftiger sig primært med data fra spørgeskemaundersøgelser, test eller andre former for strukturerede undersøgelsesmetoder (Burrell & Morgan, 1979). Dette harmonerer med, at den statistiske analyse i specialet er kvantitativ, og den tager udgangspunkt i data fra en database. Den nomotetiske metodologi stemmer overens med objektivitet.

Ud fra ovenstående grundantagelser ift. ontologi, epistemologi, menneskelig natur samt metodologi er forfatteren funderet i den objektive dimension. Den anden dimension bestemmes ved, at det vurderes, at specialet kun har mulighed for at kunne påvirke samfundet gennem regulativer. Det skyldes, at de statistiske resultater formentlig ikke vil resultere i radikale ændringer i samfundet, men derimod kan de bidrage med interessante synspunkter til, hvordan det er muligt at forudsige, hvorvidt en virksomhed går konkurs.

Grundantagelsernes objektivitet samt antagelsen om at specialets resultater kan ændre samfundet gennem regulativer medfører, at forfatterens videnskabelige ståsted befinder sig inden for det funktionalistiske paradigme. Dette paradigme er præget af en formel struktur, hvor mål realiseres gennem beregning, planlægning og kontrol (Burrell & Morgan, 1979). Det stemmer overens med indeværende speciale, da den statistiske analyse er præget af en formel struktur, hvor både beregning og kontrol har indflydelse på resultaterne.

5 Metode

Afsnittet præsenterer den metodiske tilgang i specialet, som indeholder en uddybning af undersøgelsesproblemet, den anvendte dataindsamling, analysestrategien med den tilhørende statistiske procedure samt de fire vurderingskriterier. Metodeafsnittets formål er at introducere de anvendte metodiske værktøjer, og afrundingsvis præsenteres specialets struktur samt designfigur.

5.1 Undersøgelsesproblem

Problemformuleringen er styrende for specialet, og da Booth et al. (2008) især fokuserer på fremgangsmåden vedrørende udarbejdelsen af et speciales problemformulering, indledes metodeafsnittet med dette. Booth et al. (2008) lægger vægt på, at et speciale bør tage udgangspunkt i interesser, hvorefter det indkredses til et emne. Derefter skal emnet indsnævres til et mere fokuseret emne, som kan skrives med få ord. Forfatteren skal dog være forsigtig med at indskrænke emnet for meget, da det kan resultere i datamangel (Booth et al., 2008). Ud fra det fokuserede emne bør der efterfølgende opstilles et spørgsmål, idet spørgsmålet bestemmer specialets datagrundlag, og det er en god måde at påbegynde arbejdet inden for et specifikt emne. Booth et al. (2008) anbefaler, at spørgsmålet starter med hvem, hvad, hvornår eller hvor, men det bedste er at starte med hvorfor eller hvordan.

Specialet tager udgangspunkt i forfatterens interesse for konkursprognosticering af virksomheder, og efterfølgende er interessen indkredset til et emne, der omhandler, den begrænsede tilstedeværelse af konkursprognosticeringsmodeller af danske virksomheder. Det fokuserede emne omhandler derfor en statistisk analyse af nøgletal for konkursramte virksomheder og normal drift virksomheder for at kunne opstille en konkursprognosticeringsmodel. Problemformuleringen er herefter udarbejdet ud fra det fokuserede emne.

Problemformuleringen defineres til at være et undersøgelsesproblem, hvilket medfører, at specialet løser et problem, der også er relevant for andre. Et undersøgelsesproblem kan betegnes som et konceptuelt problem, og det opstår, når der er noget ved verden, som man ikke helt forstår, og derfor gerne vil øge sin viden inden for dette område. Et undersøgelsesproblem løses ved at besvare et spørgsmål, der hjælper en med at forstå undersøgelsesproblemet bedre (Booth et al., 2008). Argumentationen for at specialet tager udgangspunkt i et undersøgelsesproblem er,

at det omhandler et konceptuelt problem, hvor forfatteren ønsker at øge sin viden inden for konkursprognosticering. Derudover indeholder specialet ikke en aktiv løsning, men derimod en besvarelse af et spørgsmål, der øger forståelsen, hvilket er kendetegnet for løsningen af et undersøgelsesproblem. En undersøgelse betegnes jf. Booth et al. (2008) som *anvendt*, hvis løsningen på undersøgelsesproblemet har praktiske konsekvenser. Det er vurderet, at undersøgelsen i indeværende speciale kan betegnes som *anvendt*, da det er muligt, at løsningen på undersøgelsesproblemet kan medføre en viden, der kan ændre personers handlinger.

Et undersøgelsesproblem består af to elementer, det ene element er en tilstand, og det andet element er uønskede konsekvenser eller omkostninger, der er forårsaget af den pågældende tilstand (Booth et al., 2008). Tilstanden i et undersøgelsesproblem er altid noget, som man ikke ved eller forstår. Et undersøgelsesproblem har ikke en håndgribelig omkostning, derfor betegnes dette element i stedet som en konsekvens. Konsekvensen af et undersøgelsesproblem er en anden ting, som man ikke forstår, fordi man ikke forstår den første ting, der er udgangspunktet for undersøgelsesproblemet. Uvidenheden om den anden ting, konsekvensen, er mere signifikant og betydningsfuld end uvidenheden om den første ting (Booth et al., 2008). Undersøgelses-problemets tilstand i specialet omhandler den uvidenhed, som forfatteren har ift., hvorvidt de udvalgte nøgletal bidrager til en forudsigelse af danske virksomheders konkurser. Konsekvensen ved undersøgelsesproblemet er aktionærers og långiveres manglende mulighed for at kunne forudsige en virksomheds konkurs.

Booth et al. (2008) pointerer, at et problem i denne henseende ikke bør undgås, derimod bør det opsøges ifm. en akademisk undersøgelse, da det indebærer, at dataindsamlingen bedre kan struktureres. Desuden sikrer anvendelsen af et problem, at der skabes et konkret fokus og en bedre struktur i specialet (Booth et al., 2008). Specialets problemformulering er gengivet herunder, da den medvirker til, at der skabes struktur i specialet. Derudover har den afgørende betydning for den anvendte dataindsamling og teori.

Hvordan bidrager udvalgte nøgletal via en logistisk regression til forudsigelse af konkurser i danske virksomheder?

5.2 Datakilder

Specialet bygger på en kvantitativ undersøgelsesmetode, jf. afsnit 4 *Videnskabsteori*, hvilket afspejles i de anvendte datakilder samt selve dataindsamlingen. Der anvendes både primære og

sekundære data i specialet. De primære data er data, som forfatteren selv udarbejder til specialet, hvorimod de sekundære data er data, der er oparbejdet til et andet formål (Heldbjerg, 2011). Der anvendes ikke primære data i sin rene form i specialet, men der udarbejdes et regneark i Excel på baggrund af sekundære data, hvori nøgletallene til analysen udregnes, og selve bearbejdningen af dette regneark får således en primær karakter, da det er udarbejdet til specialets formål. Dette medfører, at regnearket befinder sig et sted midt imellem de to datakarakteristika. De sekundære data består af de rå datasæt samt den teoretiske litteratur, der dækker over det statistiske felt, konkursteorien samt de nøgletal, der anvendes i analysen.

Der skelnes ligeledes mellem primære og sekundære kilder, hvor primære kilder er den empiri, som tilskrives den højeste værdi og prioritet, hvorimod sekundære kilder er den empiri, der tilskrives den laveste værdi og prioritet i specialet. Der foretages denne opdeling uanset, om det vedrører primære eller sekundære data (Heldbjerg, 2011). I indeværende speciale er de primære kilder de rå datasæt, der er hentet fra databasen, og det regneark, hvori nøgletallene er udregnet, samt den statistiske teori, der er grundlag for analysen. De sekundære kilder udgøres af den teori, der vedrører nøgletallene og konkursteorien.

Regnearket i Excel, der har karakter af at være primært data, kan kritiseres for at indeholde tastefejl, men da det kun er få oplysninger, der er indtastet manuelt, begrænser det kritikken. Kildekritikken af de sekundære data omhandler, at de ikke er udarbejdet specifikt til specialets formål. Hair et al. (2010), der udgør den primære statistiske litteratur i specialet, er anerkendt og velset, og derudover er den anbefalet gennem undervisere på universitetet, hvilket begrænser kritikken. Den teoretiske litteratur ift. nøgletallene tager udgangspunkt i det pensum, der har været tilknyttet uddannelsen, hvilket ligeledes begrænser kritikken.

5.2.1 Dataindsamling

Det datasæt, der anvendes til indeværende speciale, består af to separate datasæt, der er hentet fra Navne & Numre Erhverv (NN Erhverv), hvor adgangen til NN Erhverv er opnået gennem Aalborg Universitet (NN Erhverv, 2016). De to datasæt adskiller sig primært ved, at det ene datasæt vedrører virksomheder, der er opløst efter konkurs, hvorimod det andet datasæt vedrører virksomheder i normal drift. De rå datasæt fra NN Erhverv er vedlagt som Bilag 1 og Bilag 2, hvor Bilag 1 indeholder data fra de konkursramte virksomheder, og Bilag 2 indeholder data fra de virksomheder, der er i normal drift. De to datasæt er fremskaffet ved at anvende forskellige søgekriterier, således at antallet af virksomheder samt virksomhedernes data er tilpasset

specialets formål og omfang. De anvendte søgekriterier til de to datasæt fremgår af Figur 3, der er inddelt i de overskrifter, som findes inde på NN Erhverv. Den første kolonne i Figur 3 indeholder navnene på de enkelte søgekriterier, anden kolonne indeholder de søgekriterier, der er anvendt til de konkursramte virksomheder, og tredje kolonne indeholder de søgekriterier, der er anvendt til virksomhederne i normal drift.

Søgekriterie	**Indtastningsværdi**	
Firmasøgning		
Status	Selskabet er opløst efter konkurs.	Selskabet er i normal drift.
Salg og marketing		
Etableringsdato		Til 30-06-2011
Antal ansatte CVR-nr.	Fra 1	Fra 1 til 100
Økonomi og regnskab		
Dato for regnskabsafslutning		Til 31-12-2014
Omsætning	Fra -1000000	Fra -1000000
Bruttofortjeneste		Fra -1000000
Vareforbrug		Fra -1000000
EBITDA	Fra -1000000	Fra -1000000
Primært resultat		Fra -1000000
Afskrivninger		Fra -1000000
Resultat (før skat)		Fra -1000000
Årets resultat		Fra -1000000
Skat af årets resultat		Fra -1000000
Finans indtægter		Fra -1000000
Finans udgifter	Fra -1000000	Fra -1000000
Overført resultat	Fra -1000000	Fra -1000000
Omsætningsaktiver		Fra -1000000
Varebeholdning		Fra -1000000
Varedebitorer		Fra -1000000
Likvider	Fra -1000000	Fra -1000000
Egenkapital	Fra -1000000	Fra -1000000
Hensættelser		Fra -1000000
Langfristet gæld		Fra -1000000
Kortfristet gæld	Fra -1000000	Fra -1000000
Varekreditorer		Fra -1000000
Yderligere kriterier		
Fravalg af reklamebeskyttede	Tilvalgt	Tilvalgt
Firmaer med regnskabspligt	Tilvalgt	Tilvalgt

Figur 3 – Søgekriterier til datasættene (Egen tilvirkning)

De anvendte søgekriterier til konkursramte virksomheder indeholder et minimumskrav på én ansat, derudover er der anvendt et minimumskrav til virksomhedens omsætning, EBITDA,

finans udgifter, overført resultat, likvider, egenkapital og kortfristet gæld på -1000000. Disse regnskabsposter er udvalgt ud fra de nøgletal, der skal anvendes til den statistiske analyse. Det store beløb, der er i 1.000 kr., er valgt, således at stort set ingen frasorteres pga. beløbsgrænsen, men derimod hvis der ikke er et beløb i de angivne regnskabsposter. Beløbet er negativt, da det sikrer, at der medtages virksomheder med både negative og positive beløb i de udvalgte regnskabsposter. Derudover er der fravalgt virksomheder, der er omfattet af reklamebeskyttelse samt tilvalgt virksomheder, der har regnskabspligt.

Søgekriterierne til virksomheder i normal drift indeholder de samme som ved konkursramte virksomheder, men derudover er der anvendt flere søgekriterier for at begrænse antallet af virksomheder. De ekstra søgekriterier omfatter en øvre grænse for virksomhedens etableringsdato, der er fastlagt til d. 30-06-2011, således at virksomheder, der er etableret efter denne dato, ikke medtages. Datoen er fastlagt ud fra en vurdering om, at virksomheder, der er etableret efter denne dato, formentlig ikke vil aflægge et regnskab for år 2011, og da der hentes data for de sidste fem år, er det essentielt, at et regnskab for år 2011 er med. Derudover er der anvendt en maksimumsgrænse for antallet af medarbejdere, og denne er fastsat til 100 for at begrænse antallet af virksomheder i normal drift. Der er også angivet en dato for regnskabsafslutning til d. 31-12-2014 for at sikre, at perioden for virksomheder i normal drift bliver ensartet med de konkursramte virksomheder samt for at begrænse antallet af virksomheder. Der er ligeledes tilføjet et minimumskrav på -1000000 til en del andre regnskabsposter, der fremgår af Figur 3, og disse minimumskrav er udelukkende anvendt for at begrænse antallet af virksomheder. Betydningen af disse søgekriterier debatteres i afsnit 7.1.1 *Databearbejdning* og afsnit 7.1.2 *Deskriptiv statistik*.

Efter der er fundet de virksomheder, der passer til søgekriterierne, skal der tages stilling til, hvilke data der skal eksporteres. Til det formål er der udarbejdet et eksportkriterie, der er anvendt til begge datasæt, da det således er nemmere at sammensætte de to datasæt til et datasæt samt sammenligne dataene. Det anvendte eksportkriterie ses i Figur 4, og det er opdelt i fire underkategorier, hhv. basisdata, resultatopgørelse, aktiver og passiver. Basisdata indeholder generel virksomhedsinformation, der kan være relevant for analysen, og baggrunden for valget af indholdet i de tre andre underkategorier skyldes de nøgletal, der anvendes til den statistiske analyse. Det skal dog understreges, at ift. aktiver er alle tilgængelige poster en del af eksportkriteriet. Foruden det anvendte eksportkriterie i Figur 4, blev der afkrydset, at der skulle

eksporteres de sidste fem års nøgletal, således at der var flere års data at arbejde med. Bilag 1 og Bilag 2 er således et resultat af de anvendte søgekriterier samt eksportkriteriet.

Basisdata	• Firmanavn • CVR-nr. • Antal ansatte (CVR-nr.) • Etableringsdato • Virksomhedsform • Branchekode • Branchebetegnelse • Status
Resultatopgørelse	• EBTIDA • Finans udgifter • Overført resultat
Aktiver	• Anlægsaktiver • Grunde og bygninger • Omsætningsaktiver • Varebeholdning • Varedebitorer • Likvider • Goodwill • Immaterielle anlægsaktiver • Materielle anlægsaktiver • Finansielle anlægsaktiver • Periode afgræsninger • Driftsmateriel og inventar
Passiver	• Egenkapital • Kortfristet gæld

Figur 4 – Eksportkriterie til datasættene (Egen tilvirkning)

5.2.2 SPSS

Til udarbejdelsen af den statistiske analyse i specialet anvendes SPSS, der er et program til statistisk databehandling. Programmet er erhvervet gennem Aalborg Universitet, der har indkøbt en licens til alle studerende (Aalborg Universitet, 2016). SPSS er ejet af IBM, og det er en del af IBM SPSS Statistics, der er en integreret familie af produkter, der indeholder et væld af moduler, der gør det muligt at anvende de specialiserede funktioner, som der er behov for (IBM, 2016). SPSS anvendes af samfundsvidenskabelige forskere, men det er også udbredt i erhvervslivet, og derfor findes det naturligt at benytte SPSS som hjælpeværktøj til at besvare specialets problemformulering.

5.3 Analysestrategi

I afsnit 4 *Videnskabsteori* fremgår det, at forfatteren er funderet i det funktionalistiske paradigme, og det har betydning for den analysestrategi, der anvendes i specialet, da det funktionalistiske paradigme fokuserer på steps og fastlagte procedurer. Indledningsvis kan der dog nævnes, at den overordnede analysestrategi er deduktiv, da specialet tager udgangspunkt i teorien, hvoraf der udledes en hypotese om det forventede resultat, hvorefter denne testes på empirien (Heldbjerg, 2011).

En succesfuld færdiggørelse af en multivariat analyse involverer jf. Hair et al. (2010) mere end blot udvælgelsen af en korrekt metode, derfor har Hair et al. (2010) udarbejdet en sekstrinsmodel til at hjælpe forskeren. Formålet med sekstrinsmodellen er ikke at opstille et sæt af procedurer, der skal følges, men derimod at fremstille en række retningslinjer, der understreger en modelkonstruktions tilgang. Sekstrinsmodellen sætter rammerne for udvikling, tolkning og validering af enhver multivariat analyse (Hair et al., 2010). Den anvendte analysestrategi i specialet stemmer således overens med denne sekstrinsmodel, da det jf. Hair et al. (2010) sikrer en mere succesfuld færdiggørelse af en multivariat analyse. Sekstrinsmodellen visualiseres i Figur 5, og efterfølgende vil de seks trin i modellen uddybes særskilt. Det skal dog påpeges, at en mere detaljeret gennemgang af de seks trin for logistisk regression vil fremgå i afsnit 6.3.2 *Sekstrinsmodellen for logistisk regression*.

Trin 1	Definere undersøgelsesproblemet, formålet og den anvendte multivariat teknik
Trin 2	Udvikle analyseplanen
Trin 3	Evaluere forudsætningerne for multivariat teknikken
Trin 4	Estimere multivariat modellen og vurdere modellens overordnede fit
Trin 5	Fortolke den tilfældige variabel
Trin 6	Validere multivariat modellen

Figur 5 – Sekstrinsmodellen (Egen tilvirkning)

5.3.1 Trin 1 – Definere undersøgelsesproblemet, formålet og den anvendte multivariat teknik

Enhver multivariat analyse starter med at definere undersøgelsesproblemet og analysemålsætningerne i konceptuelle termer, og dette sker før variabler og foranstaltninger specificeres. Forskeren skal, uanset om det er en akademisk eller anvendt undersøgelse, først betragte problemet i konceptuelle termer ved at definere koncepterne og identificere det fundamentale forhold, der skal undersøges (Hair et al., 2010). En konceptuel model behøver ikke være kompleks og detaljeret, i stedet kan den nøjes med at være en simpel repræsentation af det forhold, der skal undersøges. Hvis forskningsmålet er et afhængighedsforhold, er forskeren nødt til at specificere de afhængige og uafhængige koncepter (Hair et al., 2010). Forskeren identificerer først de ideer og emner, der har interesse, fremfor at fokusere på konkrete foranstaltninger, der skal benyttes. Dette minimerer risikoen for, at relevante koncepter vil blive udeladt i bestræbelserne på at udvikle foranstaltninger og definere enkeltdelene i undersøgelsesdesignet (Hair et al., 2010). Når formålet og den konceptuelle model er specificeret, skal forskeren vælge den passende multivariat teknik. Denne beslutning foretages på baggrund af valget af afhængighedsmetode, hvorefter forskeren skal vælge den specifikke teknik baseret på de måletekniske egenskaber af de afhængige og uafhængige variabler (Hair et al., 2010).

5.3.2 Trin 2 – Udvikle analyseplanen

I Trin 1 blev den konceptuelle model fastlagt, og der blev udvalgt en multivariat teknik. Derfor fokuseres der i dette trin på implementeringsproblemstillinger. For hver teknik er det nødvendigt, at forskeren udvikler en analyseplan, der adresserer de særlige problemstillinger, der er gældende for teknikkens formål og design (Hair et al., 2010). Problemstillingerne omfatter generelle overvejelser såsom min. eller ønsket stikprøvestørrelse, tilladte eller påkrævet typer af variabler (tal vs. ikke-tal) og vurderingsmetoder. Disse problemstillinger løser specifikke detaljer, færdiggør modeludformningen og krav til dataindsamlingen (Hair et al., 2010).

5.3.3 Trin 3 – Evaluere forudsætningerne for multivariat teknikken

Dataindsamlingen er foretaget ud fra Trin 2, og derfor er den næste opgave ikke at estimere multivariat modellen, men derimod at evaluere de underliggende forudsætninger. Alle multivariate teknikker har underliggende forudsætninger, og disse er både statistiske og

konceptuelle, hvilket betydeligt påvirker deres evne til at repræsentere multivariate forhold (Hair et al., 2010). Teknikker, der er baseret på statistisk inferens, har forudsætninger såsom normalitet, linearitet og ligestilling af varianser i afhængighedsforhold, og disse forudsætninger skal være opfyldt (Hair et al., 2010). Hver teknik har også en række konceptuelle forudsætninger, der vedrører problemstillinger såsom modeludformning og de typer af forhold, der er repræsenteret. Før der forsøges enhver form for modelestimering, skal forskeren sikre sig, at både de statistiske og konceptuelle forudsætninger er opfyldt (Hair et al., 2010).

5.3.4 Trin 4 – Estimere multivariat modellen og vurdere modellens overordnede fit

Efter at forudsætningerne er opfyldt, fortsætter analysen med den faktiske estimering af multivariat modellen samt en vurdering af modellens overordnede fit. I estimeringsprocessen kan forskeren vælge mellem forskellige muligheder for at opfylde specifikke datakarakteristika eller for at maksimere modellens fit til dataene (Hair et al., 2010). Efter at modellen er estimeret, vurderes dens overordnede fit for at sikre, hvorvidt den opnår acceptable niveauer ift. de statistiske kriterier, og ofte vil en model blive bearbejdet flere gange i et forsøg på at opnå et bedre overordnede fit (Hair et al., 2010). Det er nødvendigt, at modellens fit er acceptabel, før analysen fortsættes. Uanset hvad modellens overordnede fit er, skal forskeren afgøre, hvorvidt resultaterne er uretmæssigt påvirket af outliers, hvilket indikerer, at resultaterne er usikre eller ikke generaliserbare. Disse anstrengelser sikrer, at resultaterne er robuste og stabile (Hair et al., 2010).

5.3.5 Trin 5 – Fortolke den tilfældige variabel

Når modellen har et acceptabelt fit, er det næste trin at fortolke den tilfældige variabel, og denne fortolkning afslører karakteren af multivariat forholdet. Fortolkningen af effekterne for de enkelte variabler er foretaget ved at undersøge de estimerede koefficienter for hver variabel i den tilfældige variabel (Hair et al., 2010). Fortolkningen kan føre til yderligere bearbejdning af variablerne, hvori modellen er genestimeret og genfortolket. Formålet er at identificere empiriske beviser for multivariat forholdet i stikprøvedataene, der kan generaliseres til den totale population (Hair et al., 2010).

5.3.6 Trin 6 – Validere multivariat modellen

Før resultaterne accepteres, skal forskeren udsætte dem for en sidste diagnostisk analyse, der vurderer graden af generaliserbarhed af resultaterne ud fra tilgængelige valideringsmetoder. Forsøget med at validere modellen er målrettet hen imod en demonstration af generaliserbarheden af resultaterne til den totale population. Disse diagnostiske analyser tilføjer lidt til fortolkningen af resultaterne, men analyserne kan ses som en forsikring mod, at resultaterne er de mest deskriptive af dataene, og dermed er de generaliserbare til populationen (Hair et al., 2010).

5.3.7 Opsummering

For hver multivariat teknik kan sekstrinsmodellen opdeles i to sektioner, hvor første sektion indeholder Trin 1-3, og anden sektion indeholder Trin 4-6. Den første sektion omhandler de problemstillinger, der opstår, mens estimeringen af den faktiske model forberedes. Den anden sektion omhandler problemstillinger vedrørende modelestimering, fortolkning og validering. Specialets analysestrategi medfører således en bestemt logik i udarbejdelsen af den statistiske analyse, da analysen gennemgår sekstrinsmodellen af Hair et al. (2010).

5.4 Kvantitative vurderingskriterier

Inden for kvantitative undersøgelser er der fire vurderingskriterier, der anvendes til at fastslå kvaliteten af specialets resultater. De fire vurderingskriterier er intern validitet, ekstern validitet, reliabilitet og objektivitet.

5.4.1 Intern validitet

Intern validitet vedrører indvendig gyldighed, der omhandler, hvor sikker er den undersøgelse, der er foretaget i specialet. Intern validitet omhandler således den undersøgte stikprøve, samt hvor stor sikkerheden er af denne. Hvis sikkerheden er stor, er specialets resultater gyldige (Heldbjerg, 2011). I afsnit 7.2.6 *Trin 6 – Validering af resultaterne* foretages der en validering af analysens resultater på baggrund af en hold-out stikprøve, der er en del af den totale stikprøve, hvilket medfører, at den vedrører den interne validitet (Hair et al., 2010). Analysens logistiske regressionsmodel valideres ikke i analysens Trin 6, hvilket medfører en lav intern validitet for specialets resultater.

5.4.2 Ekstern validitet

Ekstern validitet vedrører udvendig gyldighed, der omhandler, hvorvidt resultaterne af stikprøven kan overføres til populationen (Heldbjerg, 2011). Dette vurderingskriterie indikerer således, om stikprøven er repræsentativ for populationen. Afsnit 7.2.6 *Trin 6 – Validering af resultaterne* i analysen omhandler, som tidligere nævnt, primært den interne validitet, hvorfor valideringen i analysen kun giver en indikation af den eksterne validitet. Den lave interne validitet, der blev nævnt i afsnit 5.4.1 *Intern validitet*, resulterer i en indikation af, at den eksterne validitet er tilsvarende lav. Den lave eksterne validitet understreges af, at der i afsnit 7.1.2 *Deskriptiv statistik* vurderes, at stikprøven på flere parametre ikke er repræsentativ for populationen.

5.4.3 Reliabilitet

Reliabilitet betegnes også som pålidelighed, der vedrører, hvorvidt undersøgelsen kan foretages på ny af andre mennesker med opnåelse af samme resultat (Heldbjerg, 2011). Hvis andre mennesker benytter sig af de samme valg ift. videnskabsteoretiske ståsted, metodevalg, datasæt, teori, og multivariat teknik er det muligt, at der kan opnås identiske resultater. Den grundige gennemgang af indsamlingen af datasættene i afsnit 5.2.1 *Dataindsamling* øger ligeledes reliabiliteten, idet det gør det nemmere at fremskaffe de samme datasæt. Der er dog den hindring, at virksomheder kan ændre deres status, hvorfor NN Erhverv ofte opdaterer deres database. Derfor er det muligt, at de enkelte grupper i stikprøven enten er forøget eller formindsket, hvis der indsamles en ny stikprøve. Det er dog muligt, ud fra det anvendte datasæt til indeværende speciale at søge efter de enkelte virksomheder i databasen, således at der kan sammensættes en stikprøve, der er tilsvarende den, der benyttes i dette speciale. Dette medfører, at specialet overordnet set har en forholdsvis høj reliabilitet på trods af, at NN Erhverv opdaterer deres database løbende, hvilket kan begrænse tilgængeligheden af den anvendte stikprøve.

5.4.4 Objektivitet

Det sidste vurderingskriterie omhandler, hvorvidt der er uafhængighed mellem forfatterens opfattelse og de opnåede resultater i analysen (Heldbjerg, 2011). Forfatterens videnskabsteoretiske ståsted er funktionalismen, der er præget af objektivitet. Derfor er indeværende speciale udarbejdet ud fra en objektiv tankegang, hvilket også understreges i, at der, som nævnt ovenfor, er en forholdsvis høj reliabilitet. Udvælgelsen af søgekriterierne til

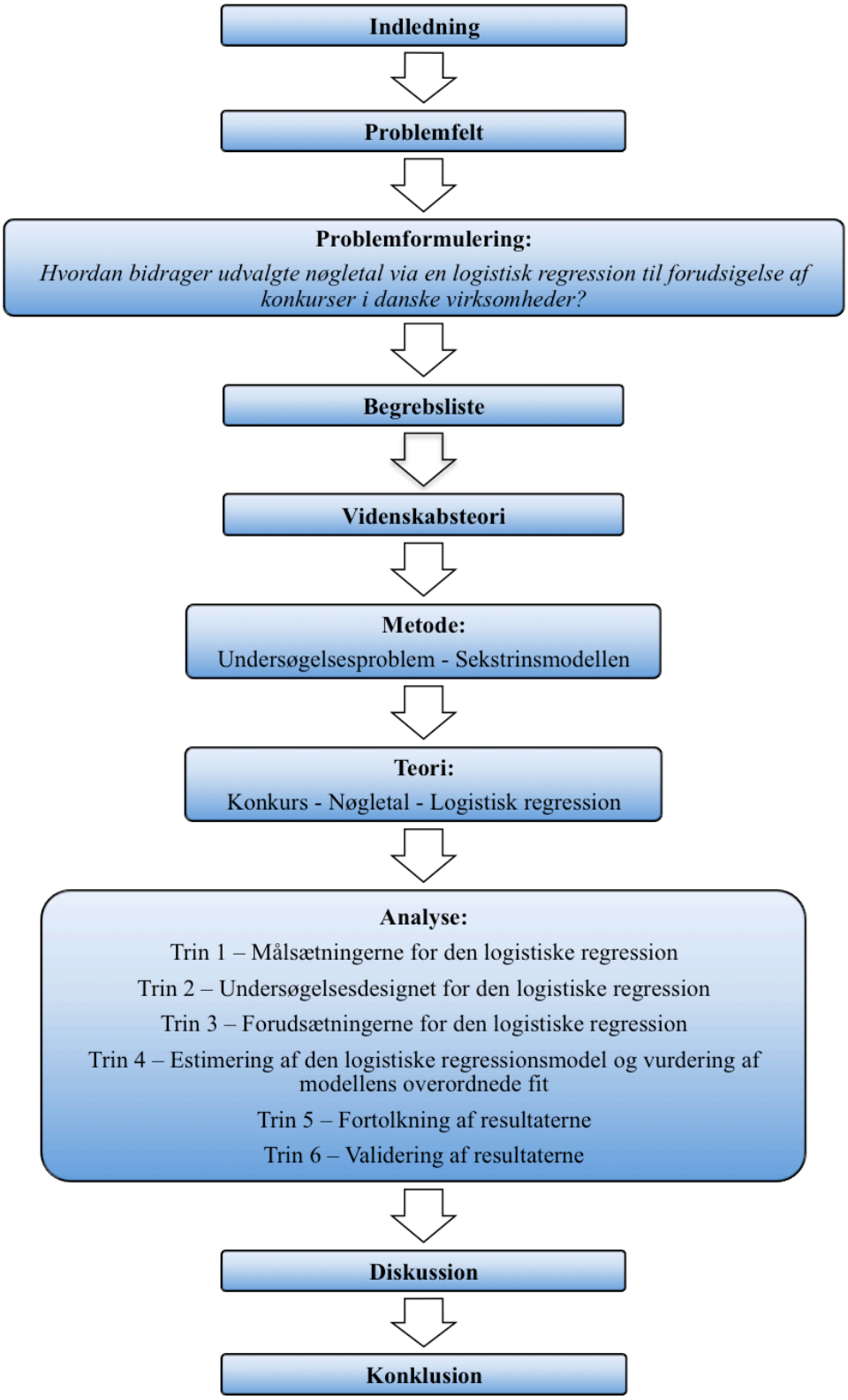
dataindsamlingen samt fortolkningen af de statistiske resultater er dog præget af en mindre objektivitet, idet udvælgelsen af søgekriterierne er foretaget ud fra en generel vurdering, og derudover er det svært at fortolke resultater uden at inddrage subjektive synspunkter. Objektiviteten for indeværende speciale karakteriseres dog til at være forholdsvis høj.

5.5 Specialets struktur

Specialet begyndes med en indledning i Kapitel 1, der introducerer konkursprognosticering samt de tidligere undersøgelser inden for feltet. Problemfeltet i Kapitel 2 konkretiserer emnet for specialet, hvorefter specialets problemformulering udledes. Desuden fastsættes rammerne for specialet i Kapitel 2. I Kapitel 3 opstilles en begrebsliste med de hyppigste anvendte statistiske begreber gennem specialet, hvorefter forfatterens videnskabsteoretiske ståsted præsenteres i Kapitel 4. Den metodiske tilgang til besvarelsen af specialets problemformulering præsenteres i Kapitel 5, hvor specialets undersøgelsesproblem ligeledes uddybes. Derudover introduceres de anvendte datakilder samt dataindsamlingen af disse. Analysens hovedteorier præsenteres i Kapitel 6, og dette kapitel er opdelt i tre underafsnit, der beskæftiger sig med hvert sit teoriområde, hhv. konkurs, nøgletal og logistisk regression.

Kapitel 7 udgør specialets analyse, hvor der indledes med en bearbejdning af de indsamlede data samt en deskriptiv statistisk af stikprøven. Derefter foretages der en logistisk regressionsanalyse, der følger specialets analysestrategi. Trin 1 omhandler målsætningerne for den logistiske regression, og Trin 2 vedrører undersøgelsesdesignet for den logistiske regression, herunder udvælgelse af variablerne samt opdeling af stikprøven. Trin 3 omhandler forudsætningerne for anvendelse af logistisk regression, og i Trin 4 estimeres den logistiske regressionsmodel, hvorefter der foretages en vurdering af modellens overordnede fit. I Trin 5 fortolkes modellens resultater, og i Trin 6 valideres resultaterne.

I Kapitel 8 diskuteres analysens resultater og anvendelsen af disse, samt hvordan resultaterne er fremkommet og kan forbedres. Derudover foretages der en sammenligning mellem resultaterne af indeværende speciale og de tidligere undersøgelser inden for konkursprognosticering. Konklusionen i Kapitel 9 afrunder specialet med en besvarelse af problemformuleringen. Specialets struktur visualiseres i Figur 6.



Figur 6 – Designfigur (Egen tilvirkning)

6 Teori

I dette afsnit præsenteres hovedteorierne, der anvendes i analysen. Afsnittet er opdelt i tre underafsnit, hhv. konkurs, nøgletal og logistisk regression, der til sammen udgør det teoretiske grundlag for at kunne besvare specialets problemformulering. Teoriafsnittet afrundes med en operationalisering af de enkelte teorier.

6.1 Konkurs

Forud for en evt. konkurs kan virksomheden anmode skifteretten om lov til at gå i betalingsstandsning, hvilket giver virksomheden ret til at udskyde sine betalinger til et senere tidspunkt (Ledernes Hovedorganisation, 1999). Det giver virksomheden ro, hvor det er muligt at få styr på økonomien, og dermed vurderes mulighederne for at føre aktiviteterne videre (Schack, 2009). En betalingsstandsning har altid en fast udløbsdato, hvorefter virksomheden skal betale sine regninger. Skifteretten har dog mulighed for at forlænge betalingsstandsningen til maks. et år, hvis der er en god grund. Under en betalingsstandsning udpeges der en tilsynsførende, der ofte er en advokat, som skal observere, hvorvidt virksomhedens ledelse udnytter betalingsstandsningen ansvarligt og fornuftigt (Ledernes Hovedorganisation, 1999). Der er ikke en direkte sammenhæng mellem betalingsstandsning og konkurs, da en betalingsstandsning ikke altid ender i en konkurs, og desuden kan en virksomhed gå konkurs uden at have været i betalingsstandsning. Det er dog ofte under en betalingsstandsning, at virksomheden indser, at det ikke er muligt at redde virksomheden, hvorefter den erklæres konkurs (Ledernes Hovedorganisation, 1999). Derfor resulterer hovedparten af betalingsstandsninger i konkurs (Schack, 2009).

Konkurs er en metode til at lukke en virksomhed, hvis den ikke formår at betale sine regninger, såfremt det ikke blot er et forbigående fænomen (Ledernes Hovedorganisation, 1999). En konkurs er dermed det ultimative økonomiske indgreb mod en skyldner (Bang-Pedersen et al., 2014). Efter § 17 i konkursloven (KL) skal en skyldner erklæres konkurs, hvis denne er insolvent, og det kan begæres af skyldner selv eller en kreditor, og en skyldner kan være fysiske såvel som juridiske personer (Schiøtz & Jakobsen, 2008). I KL § 17, stk. 2 defineres begrebet insolvent således:

"En skyldner er insolvent, hvis han ikke kan opfylde sine forpligtelser, efterhånden som de forfalder, medmindre betalingsudygtigheden må antages blot at være forbigående"
(Justitsministeriet, 2014).

Insufficiens eller teknisk insolvens, hvor passiverne overstiger aktiverne, er dermed ikke tilstrækkelig for at kunne erklære en skyldner konkurs, idet skyldneren stadig kan betale sine månedlige ydelser. Det tekniske aspekt henviser til, at det kun er på papiret, at skylderen er insolvent, såfremt at alle aktiver og lån skal realiseres øjeblikkeligt. Insufficiens indikerer dog, at skylderen ikke på lang sigt kan betale sine kreditorer (Bang-Pedersen et al., 2014).

En skyldner kan erklæres konkurs, hvis der indgives en konkursbegæring til skifteretten (Schiøtz & Jakobsen, 2008). En konkursbegæring skal være skriftlig med en original underskrift, derudover skal det være muligt at identificere skyldneren samt vurdere, hvorvidt skyldneren er insolvent (Bang-Pedersen et al., 2014). Derefter vurderer skifteretten, hvorvidt betingelserne for en konkurs er opfyldt, hvis dette er tilfældet afsiges en konkursdekret (Ledernes Hovedorganisation, 1999). Konkursen indtræder i det øjeblik, hvor skifteretten afsiger konkursdekretet (Bang-Pedersen et al., 2014). Konkursdekret er skifterettens erkendelse af, at en skylder tages under konkursbehandling (Danmarks Domstole, 2016). Det er klokkeslættet for afsigelsen af konkursdekretet, der er afgørende, idet retsvirkningerne regnes fra dette klokkeslæt. Efter afsigelsen af konkursdekretet offentliggøres det i Statstidende (Bang-Pedersen et al., 2014).

En virksomhed, der er under konkurs, har ikke et formål om at drive en aktiv erhvervsvirksomhed, men afvikler derimod aktiviteterne, således at værdierne kan gøres op og fordeles. Den øverste ledelse i virksomheden skiftes ud med en advokat, og der tilføjes "under konkurs" til virksomhedens navn (Ledernes Hovedorganisation, 1999). Advokaten, der udgør virksomhedens øverste ledelse, er i de første tre uger midlertidig bestyrer, hvorefter han bliver kurator. Det er kuratoren, der bestemmer i virksomheden, efter at konkursdekretet er afsat (Bang-Pedersen et al., 2014). De kreditorer, der har penge til gode i virksomheden, skal indgive deres krav til kuratoren, idet han skal fordele de indkomne beløb mellem kreditorerne, efter at virksomhedens værdier er opgjort og solgt (Ledernes Hovedorganisation, 1999). Afviklingen af et konkursbo er en forholdsvis langvarig affære, idet ukomplicerede sager tager ca. 1-2 år, hvorimod enkeltstående tilfælde kan tage helt op til 10-15 år (Ledernes Hovedorganisation,

1999). Det er således først efter afviklingen af et konkursbo, at en virksomhed betegnes som opløst efter konkurs.

6.2 Nøgletal

Nøgletal udarbejdes for at give et kort og samlet helhedsindtryk af virksomhedens økonomi og udvikling, men der er stor forskel på, om nøgletallene skal anvendes internt eller eksternt. Årsrapporten er udarbejdet i overensstemmelse med regnskabslove og standarder, og dermed følger nøgletallene i årsrapporten et bestemt regelsæt. De valgte regnskabsprincipper og –metoder er derfor af afgørende betydning for den afbildning af virksomhedens økonomi, som det eksterne regnskab giver (Christensen & Nielsen, 2012). Det interne regnskab afspejler derimod et styringsbehov, og det er ikke drevet af love og regler, men i stedet af en driftsøkonomisk logik. Der er således en væsentlig forskel mellem det eksterne og det interne regnskab, idet virksomheden ift. det interne regnskab selv kan kontrollere, hvilke informationer, der skal gøres tilgængelige, samt hvordan og hvornår de skal gøres tilgængelige (Christensen & Nielsen, 2012).

Generelt illustrerer nøgletal i det eksterne regnskab et øjebliksbillede, da værdierne af de enkelte poster ændrer sig løbende over året, hvorfor tidspunktet for udregning af nøgletallet er afgørende. Derudover kan nøgletal i det eksterne regnskab manipuleres med inden for de fastsætte regnskabslove og standarder. Indeværende speciale tager udgangspunkt i nøgletal fra danske virksomheders eksterne regnskab, derfor er der risiko for, at de anvendte nøgletal kan være manipuleret. Det har dog ikke været muligt inden for de fastsætte rammer for indeværende speciale at undersøge og frasortere evt. manipulerede nøgletal.

6.2.1 Udvælgelse af nøgletallene

I afsnit 2 *Problemfelt* blev det præsenteret, at de udvalgte nøgletal til indeværende speciale er Altman & Sabatos (2007) fem nøgletal samt soliditetsgraden. Derfor uddybes Altman & Sabatos (2007) udvælgelsesprocedure for de fem nøgletal, hvorefter den teoretiske argumentation for inddragelse af soliditetsgraden fremføres.

I Altman & Sabato (2007) fremgår det, at et stort antal af mulige nøgletal er identificeret i litteraturen, som værende brugbare til at kunne forudsige virksomheders konkurs. Altman & Sabato (2007) anvender kun kvantitative variabler, da deres database ikke indeholder kvalitative variabler. Altman & Sabato (2007) anvender, ligesom tidligere studier, fem nøgletalskategorier,

der beskriver de vigtigste aspekter af en virksomheds finansielle profil. Disse nøgletalskategorier er gearing, likviditet, rentabilitet, dækning og aktivitet. For hver kategori har Altman & Sabato (2007) fremstillet et antal af nøgletal, der er identificeret i litteraturen som værende mest succesful ift. at kunne forudsige virksomheders konkurs. Disse nøgletal fremgår af første kolonne i Figur 7.

Variables examined	Variables manually selected	Variables entered in the model	Accounting ratio category
Short term debt/ Equity (book value)	Short term debt/ Equity (book value)		
Equity (book value)/ Total liabilities		Short term debt/ Equity book value	Leverage
Liabilities/Total assets	Liabilities/Total assets		
Cash/Total assets	Cash/Total assets		
Working capital/Total assets			
Cash/Net sales	Working capital/ Total assets	Cash/Total assets	Liquidity
Intangible/Total assets			
EBIT/Sales			
EBITDA/Total assets	EBITDA/Total assets		
Net income/Total assets		EBITDA/Total assets	Profitability
Retained earnings/ Total assets	Retained earnings/ Total assets		
Net income/Sales			
EBITDA/Interest expenses	EBITDA/Interest expenses	Retained earnings/ Total assets	
EBIT/Interest expenses	EBIT/Interest expenses		Coverage
Sales/Total assets	Sales/Total assets		
Account payable/Sales	Account receivable/ Liabilities	EBITDA/Interest expenses	Activity
Account receivable/Liabilities			

Figur 7 – Udvælgelsesprocessen af variablerne (Altman & Sabato, 2007, s. 341)

Derefter er de potentielle nøgletalsindikatorer kalkuleret for at observere nøjagtigheden for hver nøgletalsindikator, hvorefter der arbitrært er udvalgt to variabler fra hver gruppe med den højeste nøjagtighed (Altman & Sabato, 2007). Analysen i Altman & Sabato (2007) tog ikke højde for en mulig korrelation mellem variablerne i hver gruppe, og derfor blev der udvalgt to variabler fra hver nøgletalskategori i stedet for en. De ti udvalgte variabler fremgår af kolonne to i Figur 7. Derefter blev der udført en statistisk forward stepwise udvælgelsesproces for de ti udvalgte variabler, hvor der først blev udvalgt de mest relevante variabler, og derefter blev der tilføjet den stepwise udvælgelsesproces (Altman & Sabato, 2007). Efterfølgende blev den fulde model estimeret for at eliminere de mindst hjælpsomme variabler en efter en, indtil der til sidst

kun var de brugbare variabler tilbage, dvs. dem med en p-værdi under det valgte signifikansniveau. I dette tilfælde har Altman & Sabato (2007) valgt et signifikansniveau på 20 %. Ud af de ti variabler blev fem variabler udvalgt, da de sammen gjorde det overordnede bedst ift. at kunne forudsige en virksomheds konkurs. De fem variabler fremgår af kolonne tre i Figur 7.

Nøgletalskategorierne i Altman & Sabato (2007) er, som tidligere nævnt, gearing, likviditet, rentabilitet, dækning og aktivitet. Hvor gearing afspejler forholdet mellem gæld og egenkapital i virksomheden, likviditet udtrykker virksomhedens kortsigtede betalingsevne, og rentabilitet illustrerer virksomhedens indtjeningsevne (Schack, 2009). Derudover omhandler dækning virksomhedens evne til at dække omkostningerne, og aktivitet afspejler virksomhedens aktivitetsniveau. På trods af at disse nøgletalskategorier dækker bredt, omhandler ingen af dem virksomhedens finansielle styrke, og dermed evnen til at modstå evt. tab, hvilket må anses at være relevant ift. at kunne forudsige virksomheders konkurs. Derfor inddrages soliditet som en nøgletalskategori med soliditetsgraden som nøgletal, idet soliditetsgraden illustrerer virksomhedens finansielle styrke og modstandsdygtighed over for evt. tab (Jensen et al., 2010).

6.2.2 Elementer i nøgletallene

I dette afsnit vil de enkelte elementer i de udvalgte nøgletal defineres, herunder hvad elementerne består af. Elementerne er aktivsum, likvider, egenkapital, kortfristet gæld, EBITDA, finans udgifter og overført resultat.

6.2.2.1 Aktivsum

En aktivsum består af en virksomheds samlede aktiver, herunder både anlægsaktiver og omsætningsaktiver. Anlægsaktiver betegnes også som faste aktiver, og det er aktiver, der forventes at medføre økonomiske fordele for virksomheden i mere end et år (Hawawini & Viallet, 2011). I årsregnskabsloven defineres anlægsaktiver som aktiver, der er bestemt til vedvarende eje eller brug for virksomheden (Christensen & Nielsen, 2012). Anlægsaktiverne opdeles i håndgribelige og uhåndgribelige aktiver, hvor de håndgribelige aktiver udgøres af grunde, bygninger, maskiner, inventar mm., hvorimod de uhåndgribelige aktiver udgøres af goodwill, patenter mm. (Hawawini & Viallet, 2011). Udover de materielle anlægsaktiver og de immaterielle anlægsaktiver eksisterer der også finansielle anlægsaktiver, der omfatter alle former for finansielle fordringer, som virksomheden besidder mhp. vedvarende eje. De finansielle

anlægsaktiver omfatter bl.a. aktier i datterselskaber og associerede selskaber. Omsætningsaktiver er den anden del af aktivsummen, og disse aktiver forventes omsat inden for kort tid. Omsætningsaktiverne består bl.a. af varebeholdninger, tilgodehavender, likvide midler og periodeafgrænsningsposter (Christensen & Nielsen, 2012).

6.2.2.2 Likvider

Likvider eller likvide midler inkluderer kontanter og indestående i banker samt kortsigtede investeringer med en forfaldstid inden for et år. Disse kortfristede investeringer er normalvis kendt som værdipapirer, og de indebærer kun en lille risiko, samtidig med at de er meget likvide. Det medfører, at værdipapirerne nemt kan sælges, og dermed konverteres til kontanter med minimal ændring i værdien (Hawawini & Viallet, 2011). Det er et krav, at værdipapirerne uden hindring kan omsættes til likvide midler, og at risikoen for værdiændringer er relativ lav, før de kan placeres under likvider. Derudover skal de pågældende værdipapirer have en funktion som likviditet, dvs. at de skal indgå i virksomhedens likviditetsstyring (Schack, 2009). Likvider er en del af omsætningsaktiverne i balancen.

6.2.2.3 Egenkapital

Formålet med balancen er at kunne estimere de samlede investeringer, der er foretaget af virksomhedens ejere på et givent tidspunkt, hvilket typisk er ved slutningen af regnskabsperioden. Denne investering er kendt som egenkapitalen, og den er forskellen mellem aktiverne og passiverne (Hawawini & Viallet, 2011). Denne sammenhæng kan udtrykkes med denne formel:

$$Egenkapital = Aktiver - Passiver$$

En egenkapital ved et dansk aktie- og anpartsselskab kan bestå af elementerne virksomhedskapital, overkurs ved emission, reserve for opskrivning, andre reserver og overført overskud eller underskud (Elling, 2014). Virksomhedskapitalen er den pålydende værdi af selskabets udstedte aktier, og den kan ikke udloddes. Virksomhedskapitalen skal ved stiftelse af et aktieselskab udgøre min. 500.000 kr., hvorimod den ved stiftelse af et anpartsselskab skal udgøre min. 50.000 kr. (Elling, 2014). Overkurs ved emission opstår, når virksomheden udsteder nye aktier og sælger disse til en pris, der er højere end aktiernes pålydende værdi. Det beløb, som salgssummen for de nyudstedte aktier overstiger aktiernes pålydende værdi, betegnes som overkurs ved emission (Christensen & Nielsen, 2012). Reserve for opskrivning er den post, hvor

selskabet placerer alle de beløb, som selskabets aktiver er blevet opskrevet med. Under andre reserver opføres de forskellige egenkapitalandele, der er disponeret pga. bestemmelser i lovgivningen eller i selskabets egne vedtægter. Overført overskud eller underskud betegnes også overført resultat, og det er den del af virksomhedens egenkapital, som generalforsamlingen beslutter at overføre til næste år (Christensen & Nielsen, 2012).

Generelt øges egenkapitalen, hvis virksomheden har overskud, hvorimod den mindskes, hvis virksomheden har underskud. Egenkapitalen øges ligeledes ved udstedelse af nye aktier, hvorimod den mindskes, når virksomheden opkøber sine egne aktier. Derudover mindskes egenkapitalen som følge af udbyttebetaling (Hawawini & Viallet, 2011).

6.2.2.4 Kortfristet gæld

Kortfristet gæld forfalder inden for et år, og den inkluderer fx træk på kassekreditten, og den del af den langfristede gæld, der falder til betaling inden for et år (Hawawini & Viallet, 2011). Derudover er den kortfristet gæld rentebærende og en del af passiverne i balancen.

6.2.2.5 EBITDA

EBITDA står for Earnings Before Interests, Taxes, Depreciation and Amortisation, og det er således et økonomisk udtryk for en virksomheds driftsresultat før renter, skat og afskrivninger (Hawawini & Viallet, 2011). EBITDA medtager derfor både omsætning, variable omkostninger, salgs- og distributionsomkostninger samt administrationsomkostninger.

6.2.2.6 Finans udgifter

Finans udgifter eller renteudgifter dækker over udbetalinger, der er tilknyttet virksomhedens lån. Det er en omkostningspost i resultatopgørelsen (Christensen & Nielsen, 2012).

6.2.2.7 Overført resultat

Overført resultat udgøres af årets resultat efter skat fratrukket evt. udbytte, og årets resultat efter skat fremgår direkte af resultatopgørelsen. Et overført resultat vil efterfølgende indgå som en del af egenkapitalen (Hawawini & Viallet, 2011). Det er generalforsamlingen, der beslutter, om der skal overføres et beløb til næste år. Et overført resultat kan ses som en form for udbytteudjævningspulje, da det overførte resultat skal sikre, at virksomheden kan udbetale et konstant udbytte over en årerække (Christensen & Nielsen, 2012).

6.2.3 Nøgletallenes betydning

De seks udvalgte nøgletal uddybes i dette afsnit, hvor den anvendte formel ligeledes opstilles. Der fokuseres på at uddybe, hvad hvert nøgletal betyder. Derudover præsenteres den forkortelse af hvert nøgletal, der anvendes i specialet.

6.2.3.1 Short term debt/Equity book value

Nøgletallet repræsenterer nøgletalskategorien gearing, og det er virksomhedens kortfristet gæld ift. egenkapitalens bogførte værdi. Den anvendte forkortelse er STD_EBV, og formlen for nøgletallet ser således ud:

$$\frac{Short\ term\ debt}{Equity\ book\ value}$$

Nøgletallet minder om finansiel gearing, der illustrerer forholdet mellem virksomhedens samlede gæld og egenkapitalen ved at udtrykke, hvor mange gange gælden overstiger egenkapitalen (Jensen et al., 2010). STD_EBV medtager dog kun den kortfristet gæld, hvilket medfører, at nøgletallet kun afspejler, hvorvidt virksomhedens kortfristet gæld overstiger egenkapitalen. Ud fra dette nøgletal er det således muligt at vurdere, hvorvidt virksomheden er finansieret gennem kortfristet gæld frem for egenkapital.

6.2.3.2 Cash/Total assets

Nøgletalskategorien for dette nøgletal er likviditet, og nøgletallet udgøres af virksomhedens likvider ift. aktivsummen. Nøgletallets forkortelse er C_TA, og formlen ser således ud:

$$\frac{Cash}{Total\ assets}$$

Dette nøgletal giver en vurdering af virksomhedens øjeblikkelige betalingsevne, da likvider udgør den direkte kilde til at dække de løbende driftsbetalinger og andre betalingsforpligtelser, der forfalder løbende (Christensen & Nielsen, 2012). C_TA måler således, hvor stor en andel likvider udgør af virksomhedens aktiver. En høj andel af likvider indikerer en vis grad af sikkerhed fra en kreditors synspunkt, omvendt kan et for højt beløb ses som ineffektivitet.

6.2.3.3 EBITDA/Total assets

Rentabilitet er nøgletalskategorien for dette nøgletal, der er virksomhedens EBITDA ift. virksomhedens aktivsum. Nøgletallet forkortes i indeværende speciale til EBITDA_TA, og formlen for nøgletallet er:

$$\frac{EBITDA}{Total\ assets}$$

Nøgletallet minder om afkastningsgraden, der er et centralt nøgletal til bedømmelse af virksomheder, da det viser det gennemsnitslige afkast pr. indsat kapitalkrone (Schack, 2009). Afkastningsgraden måler således, hvorvidt virksomheden formår at skabe et overskud ud fra den indskudte kapital. EBITDA_TA medtager dog ikke afskrivninger og nedskrivninger, hvilket adskiller den en smule fra afkastningsgraden. Derfor afspejler nøgletallet mængden af EBITDA-overskud, der er tjent på baggrund af selskabets aktivsum.

6.2.3.4 Retained earnings/Total assets

Nøgletallet repræsenterer nøgletalskategorien dækning, og det udgøres af virksomhedens overførte resultat ift. aktivsummen. I indeværende speciale forkortes nøgletallet til RE_TA, og formlen for nøgletallet ser således ud:

$$\frac{Retained\ earnings}{Total\ assets}$$

RE_TA måler mængden af geninvesteret indtjening eller tab, hvilket afspejler omfanget af virksomhedens gearing. Virksomheder med et lavt RE_TA finansierer kapitaludgifter gennem lån fremfor gennem overført resultat, hvorimod virksomheder med et højt RE_TA udtrykker rentabilitet og evnen til at modstå tab. Nøgletallet udtrykker således virksomhedens evne til at akkumulere indtjening vha. aktiverne, derfor er en højere værdi af nøgletallet at foretrække, idet det indikerer, at virksomheden er i stand til at fastholde mere indtjening.

6.2.3.5 EBITDA/Interest expenses

Nøgletalskategorien for dette nøgletal er aktivitet, og det er virksomhedens EBITDA ift. renteudgifter. Nøgletallet forkortes til EBITDA_IE i indeværende speciale, og formlen for nøgletallet er:

$$\frac{EBITDA}{Interest\ expenses}$$

EBITDA_IE minder om rentedækningsgraden, der er et udtryk for virksomhedens finansielle styrke og modstandskraft (Jensen et al., 2010). Der er dog ingen renteindtægter med i EBITDA_IE, og derudover skal EBIT erstattes med EBITDA i rentedækningsgraden. EBITDA_IE minder også om nøgletallet gældsdækning I, der indikerer virksomhedens evne til at imødekomme rentebetalinger (Christensen & Nielsen, 2012). Gældsdækning I tager dog udgangspunkt i resultat af primær drift, hvorimod EBITDA_IE anvender EBITDA. Nøgletallet EBITDA_IE anvendes til at analysere likviditetsrisikoen for en virksomhed, hvis nøgletallet er lig med 1,0, betyder det, at virksomheden kun akkurat genererer nok indtjening til at dække rentebetalingerne i et år. Et nøgletal under 1,0 indikerer, at der kan være et likviditetsproblem i virksomheden, hvorimod et nøgletal over 1,0 indikerer, at der er et overskud af likviditet til at dække rentebetalingerne.

6.2.3.6 Equity book value/Total assets

Det supplerende nøgletal understøtter nøgletalskategorien soliditet, og det er virksomhedens bogførte egenkapital ift. virksomhedens aktivsum. Nøgletallet forkortes til EBV_TA, og der anvendes følgende formel for indeværende speciale:

$$\frac{Equity\ book\ value}{Total\ assets}$$

Det supplerende nøgletal er, som tidligere nævnt, soliditetsgraden, der er en traditionel måde at vise en virksomheds finansielle styrke på (Jensen et al., 2010). Soliditetsgraden illustrerer, hvor stor en del af egenkapitalen, der kan mistes, inden fremmedkapitalen bliver berørt. En analyse af virksomhedens soliditet anvendes til at vurdere virksomhedens evne til at betale på lang sigt. Generelt er det således, at der opnås en højere og bedre soliditet, hvis en stor andel af egenkapitalen finansierer virksomheden. En lav soliditet indebærer ofte en høj risiko for virksomheden, hvorimod en høj soliditet indebærer en lavere økonomisk risiko (Jensen et al., 2010).

6.3 Logistisk regression

Logistisk regression er en speciel form for regression, som er formuleret til at forudsige og forklare en binær (to-gruppe) kategorisk variabel i stedet for en afhængig tal-variabel. Den logistiske regressions tilfældige variabel minder om den tilfældige variabel, der er i en multipel regression. Den tilfældige variabel repræsenterer et enkelt multivariat forhold med regressionslignende koefficienter, der indikerer den relative indflydelse af hver forudsigelig variabel. Logistisk regression bliver mindre påvirket, hvis basale forudsætninger især variablernes normalitet ikke bliver opfyldt. Dette er en fordel ved logistisk regression ift. diskriminant analyse (Hair et al., 2010). Logistisk regression er dog begrænset til kun at kunne forudsige en to-gruppe afhængig variabel. En logistisk regression kan beskrives som en estimering af forholdet mellem en enkelt ikke-tal (binær) afhængig variabel og et sæt af tal eller ikke-tal uafhængige variabler. I den generelle form vil formlen se således ud:

$$Y_1 = X_1 + X_2 + X_3 + \cdots + X_n$$

Ud fra formlen skal det understregres, at Y karakteriseres som binær og ikke-tal, hvorimod X'erne karakteriseres som både ikke-tal og tal (Hair et al., 2010). En potentiel anvendelse inkluderer at kunne forudsige hvad som helst, hvor resultatet er binær dvs. enten/eller. Eksempler på dette kan være, hvorfor/hvorfor ikke er en person kunde, eller hvorvidt vil en virksomhed få succes/gå konkurs (Hair et al., 2010). Ved anvendelse af en logistisk regression vil objektet i hvert tilfælde tilhøre en af de to grupper, og formålet er således at kunne forudsige og forklare hvert enkelt objekts gruppemedlemsskab vha. et sæt af uafhængige variabler, der er udvalgt af forskeren (Hair et al., 2010).

6.3.1 Hvorfor vælge en logistisk regression?

Ud fra en praktisk synsvinkel er der to årsager til, hvorfor der foretrækkes en logistisk regression. Den første årsag er, at logistisk regression ikke har strenge forudsætninger, og den er derfor mere robust, hvilken gør den anvendelig i mange situationer (Hair et al., 2010). Den anden årsag er, at mange forskere foretrækker logistisk regression, fordi den minder om multipel regression. Det ses ved, at den har lignende statistiske test, fremgangsmåde ift. at inkorporere tal-variabler, ikke-tal-variabler og ikke-lineære effekter samt et bredt udvalg af diagnostik. Ud fra dette og mere tekniske årsager svarer en logistisk regression til en to-gruppe diskriminant analyse, og den er derfor formentlig mere passende i mange situationer (Hair et al., 2010).

6.3.2 Sekstrinsmodellen for logistisk regression

Proceduren for at anvende en logistisk regression kan ses ud fra en sekstrinsmodel, og denne model har et perspektiv, der favner en modelkonstruktion. Gennemgangen af nedenstående sekstrinsmodel for logistisk regression er således en uddybning og specificering af den generelle sekstrinsmodel, der blev præsenteret for alle multivariate analyser ifm. analysestrategien i afsnit 5.3 *Analysestrategi*.

6.3.2.1 Trin 1 – Målsætninger for logistisk regression

Logistisk regression er identisk med diskriminant analyse, hvis man ser på, hvilke grundlæggende målsætninger, som den kan adressere. Logistisk regression er bedst egnet til følgende to forskningsmål (Hair et al., 2010):

1. Identificere de uafhængige variabler, der har indflydelse på gruppemedlemskabet af den afhængige variabel.
2. Etablere et klassifikationssystem, der er baseret på den logistiske regressionsmodel for afgørelse af gruppemedlemskabet.

Det første forskningsmål minder om den primære målsætning for diskriminant analyse og endda multipel regression, da der fokuseres på forklaringen af gruppemedlemskabet ud fra de uafhængige variabler i modellen. I klassifikationsprocessen i andet forskningsmål giver logistisk regression, ligesom diskriminant analyse, et grundlag for ikke kun at klassificere den stikprøve, der er anvendt til at estimere den logistiske regression, hvilket medfører, at der også er grundlag for at klassificere enhver anden observation, som kan have værdi for alle de uafhængige variabler. På den måde kan en logistisk regressionsanalyse klassificere andre observationer ind i de definerede grupper (Hair et al., 2010).

6.3.2.2 Trin 2 – Undersøgelsesdesign for logistisk regression

Logistisk regression har flere unikke egenskaber, der påvirker undersøgelsesdesignet. Den første er tilstedeværelsen af den binære afhængige variabel, hvilket i sidste ende påvirker modellens specifikation og estimering. Den anden problemstilling relaterer sig til stikprøvestørrelsen, og den er påvirket af forskellige faktorer, heriblandt brugen af *Maximum Likelihood Estimator* (MLE), behovet for estimering og hold-out stikprøver (Hair et al., 2010).

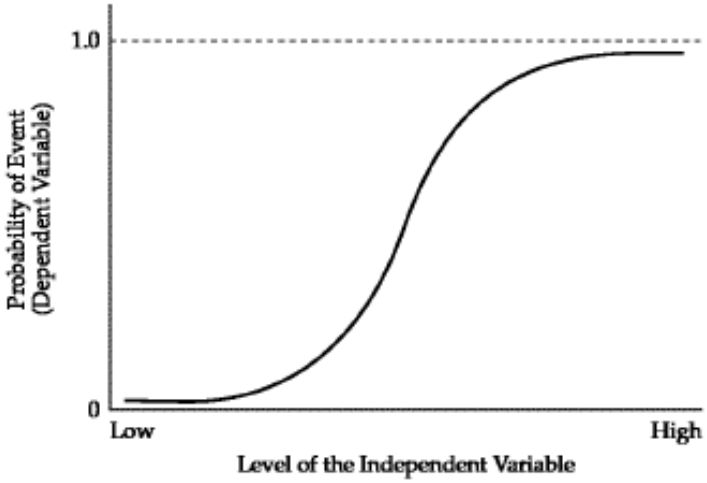
6.3.2.2.1 Den binære afhængige variabel

Den metode, der anvendes ved logistisk regression til repræsentation af den binære afhængige variabel, minder i dette tilfælde om den metode, der anvendes ved multipel regression. Logistisk regression repræsenterer de to grupper, der har interesse, som binære variabler med værdier af 0 og 1. Det har ingen betydning, hvilken gruppe der er tildelt værdierne 1 eller 0, men denne tildeling skal være bekendt ift. fortolkningen af koefficienterne (Hair et al., 2010). Hvis grupperne repræsenterer udfald eller hændelser fx succes eller fiasko, vil tildelingen af gruppekoderne have indflydelse på fortolkningen. Antages der, at gruppen med succes er kodet som 1, og gruppen med fiasko er kodet som 0, vil koefficienterne repræsentere indflydelsen på sandsynligheden for succes. Det er dog også muligt at ombytte koderne, således at succes får tildelt værdien 0, og fiasko får tildelt værdien 1.

Logistisk regression adskiller sig fra multipel regression ved, at den er specielt designet til at forudsige sandsynligheden for, at en hændelse opstår, dvs. sandsynligheden for, at en observation er i den gruppe, der er kodet med 1 (Hair et al., 2010).

6.3.2.2.1.1 Den logistiske kurve

Den binære afhængige variabel har som bekendt kun værdierne 0 og 1, derfor skal sandsynligheden ligeledes begrænses til at falde inden for samme interval. For at definere et forhold, der er begrænset af 0 og 1, anvender logistisk regression den logistiske kurve til at repræsentere forholdet mellem de uafhængige og afhængige variabler (Hair et al., 2010). Den logistiske kurve er s-formet, og den fremgår af Figur 8. Sandsynligheden vil ved meget lave niveauer af de uafhængige variabler nærme sig 0, men den vil aldrig ramme 0. Derimod vil sandsynligheden øges langs kurven, når niveauet af de uafhængige variabler øges, men derefter vil hældningen aftage, således at sandsynligheden ved ethvert niveau af de uafhængige variabler vil nærme sig 1, men den vil aldrig overstige det (Hair et al., 2010).



Figur 8 – Den logistiske kurve (Hair et al., 2010, s. 319)

6.3.2.2.1.2 Den afhængige variabels unikke karakter

Den binære karakter af den afhængige variabel (0 eller 1) har egenskaber, der bryder med forudsætningerne for multipel regression, og derfor kan multipel regression ikke anvendes. Den ene er, at fejlleddet af en særskilt variabel følger binomialfordelingen i stedet for normalfordelingen, hvilket ugyldiggør alle statistiske test, der er baseret på forudsætningen om normalfordeling. Derudover er variansen af en dikotom variabel ikke konstant, hvilket ligeledes skaber tilfælde af variansinhomogenitet. Dertil kan ingen brud på forudsætningerne blive ordnet gennem transformering af de afhængige eller uafhængige variabler (Hair et al., 2010). Logistisk regression blev udviklet specielt til at håndtere disse problemstillinger. Dens unikke forhold mellem afhængige og uafhængige variabler kræver dog en anden tilgang til at estimere den tilfældige variabel og vurdere *goodness-of-fit* samt fortolke koefficienterne, hvis der sammenlignes med multipel regression (Hair et al., 2010).

6.3.2.2.2 Stikprøvestørrelsen

Logistisk regression skal, ligesom alle andre multivariate teknikker, overveje den analyserede stikprøvestørrelse. Meget små stikprøver har utrolig mange stikprøvefejl, hvilket gør identifikationen af alle stikprøvefejlene og de største forskelle usandsynlige. Derimod vil meget store stikprøver øge den statistiske magt, således at enhver forskel, uanset om den har praktisk relevans eller ej, vil blive betragtet som statistisk signifikant. De fleste undersøgelsessituationer vil befinde sig i midten af disse ekstreme situationer, hvorfor forskeren skal overveje

stikprøvestørrelsens indflydelse på resultaterne både ift. det overordnede niveau og på gruppeniveau (Hair et al., 2010).

6.3.2.2.2.1 Den totale stikprøvestørrelse

Det første aspekt af stikprøvestørrelsen er, at den totale stikprøvestørrelse skal være tilstrækkelig for at kunne støtte estimeringen af den logistiske regressionsmodel. Logistisk regression adskiller sig fra de andre teknikker ved at anvende MLE som estimeringsteknik, og da MLE kræver store stikprøver, resultere det i, at logistisk regression alt andet lige vil kræve større stikprøvestørrelser end multipel regression (Hair et al., 2010). Derudover bør forskeren overveje at opdele stikprøven i delstikprøver og hold-out stikprøver for at validere den logistiske regressionsmodel. Hvis der foretages en opdeling af stikprøven, vil de omtalte betingelser for stikprøvestørrelsen stadig være gældende for både delstikprøverne og hold-out stikprøverne, hvilket formentlig vil fordoble den overordnede stikprøvestørrelse baseret på model-specifikationen. Det er vigtigt at sikre, at udvælgelsen af hold-out stikprøven sker tilfældigt, således at ingen ordning i observationerne påvirker estimeringsprocessen og valideringen (Hair et al., 2010).

6.3.2.2.2.2 Stikprøvestørrelse for hver kategori af den afhængige variabel

Udover vigtigheden af den totale stikprøvestørrelse er det også vigtigt at overveje stikprøvestørrelsen for hver gruppe af den afhængige variabel. Den anbefalede stikprøvestørrelse for hver gruppe er min. ti observationer for hver estimeret parameter. Det antal er meget større end ved multipel regression, der har et min. på fem observationer for hver parameter, og dette er for den totale stikprøve og ikke for stikprøvestørrelsen for hver gruppe, hvilket er tilfældet ved logistisk regression (Hair et al., 2010).

6.3.2.3 Trin 3 – Forudsætninger for logistisk regression

En logistisk regression har generel en mangel på forudsætninger ift. diskriminant analyse og multipel regression, hvilket er en kendetegnende fordel for logistisk regression. Den kræver ikke, at de uafhængige variabler følger en specifik fordeling, og problemstillinger som fx variansinhomogenintet bliver ikke aktuelle. Derudover kræver logistisk regression ikke et lineært forhold mellem uafhængige variabler og de afhængige variabler, hvilket er tilfældet ved multipel regression (Hair et al., 2010).

6.3.2.4 Trin 4 – Estimering af den logistiske regressionsmodel og vurdering af det overordnede fit

Et af de unikke karakteristika ved logistisk regression er dens brug af det logistiske forhold, der anvendes til både at estimere den logistiske regressionsmodel og etablere forholdet mellem de afhængige og uafhængige variabler. Der sker en unik transformering af den afhængige variabel, hvilket påvirker estimeringsprocessen og de resulterende koefficienter for de uafhængige variabler. Den logistiske regression anvender forskellige tilgange til at vurdere modellens overordnede fit, og disse tilgange deles med både diskriminant analyse og multipel regression (Hair et al., 2010).

6.3.2.4.1 Estimering af den logistiske regressionsmodel

Logistisk regression har en enkelt tilfældig variabel, der er sammensat af estimerede koefficienter for hver uafhængig variabel, hvilket også er tilfældet i multipel regression, men denne tilfældige variabel er estimeret på en anden måde. Logistisk regression har fået sit navn fra logit transformationen, der er brugt på den afhængige variabel, hvilket har skabt flere forskelle i estimerings- og fortolkningsprocessen (Hair et al., 2010). Logit modellen anvender, som tidligere nævnt, en speciel form for den logistiske kurve, der er s-formet, således at sandsynligheden ligger i intervallet mellem 0 og 1. For at estimere en logistisk regressionsmodel skal kurven af de forudsagte værdier tilpasses de aktuelle data, og de aktuelle dataværdier af de afhængige variabler kan kun være 0 eller 1. I logistisk regression følges der, den samme procedure som ved multipel regression, hvor den afhængige variabel forudsiges ved en tilfældig variabel, der er sammensæt af de logistiske koefficienter og de tilsvarende uafhængige variabler. Forskellen mellem de to procedurer er, at i en logistisk regression kan de forudsagte værdier ikke være uden for intervallet mellem 0 og 1 (Hair et al., 2010).

6.3.2.4.1.1 Den logistiske transformation

Logistisk regression forudsiger ligesom multipel regression en afhængig variabel, der kan karakteriseres som et tal. Sandsynlighedsværdien er i dette tilfælde begrænset til at være inden for intervallet mellem 0 og 1, og den logistiske transformation skal være med til at sikre dette (Hair et al., 2010). Den logistiske transformation består af at gentage en sandsynlighed som odds og kalkulere logit værdien.

Sandsynligheder er i deres originale form ikke begrænset til at have en værdi mellem 0 og 1. Derfor er det nødvendigt at gentage proceduren, men hvor sandsynligheden i stedet udtrykkes som odds, der er sandsynlighedsforholdet af de to udfald eller begivenheder. Odds beregnes således, hvor $Prob_i$ er den udregnede sandsynlighed (Hair et al., 2010):

$$\frac{Prob_i}{(1 - Prob_i)}$$

Ud fra denne formel vil enhver sandsynlighedsværdi blive angivet som en tal-variabel, der kan estimeres direkte, hvorefter enhver odds værdi kan konverteres tilbage til en sandsynlighed, der ligger mellem 0 og 1. Problemet med begrænsningen af den forudsagte værdi til intervallet mellem 0 og 1 er dermed løst, da en forudsigelse af odds værdien og en efterfølgende konvertering til sandsynligheden sikrer dette (Hair et al., 2010).

Hvis sandsynligheden for succes er 0,80, giver det en sandsynlighed for fiasko på 0,20 $(0{,}20 = 1{,}0 - 0{,}80)$. Disse sandsynligheder giver et odds for succes på 4,0, $\left(4{,}0 = \frac{0{,}80}{0{,}20}\right)$, hvilket betyder, at succes er fire gange mere sandsynlig end fiasko. Oddset for fiasko vil således være 0,25, $\left(0{,}25 = \frac{0{,}20}{0{,}80}\right)$, dvs. at fiasko sker en fjerdedel af gangene ift. succes. Uanset om der fokuseres på succes eller fiasko, er det muligt at angive sandsynligheden som odds (Hair et al., 2010). En sandsynlighed på 0,50 resulterer i et odds på 1,0, da begge udfald har lige store chancer for at forekomme. Et odds under 1,0 repræsenterer sandsynligheder under 0,50, og et odds over 1,0 svarer til en sandsynlighed større end 0,50.

Odds værdien har en nedre grænse på 0, hvorimod den ikke har nogen øvre grænse. For at undgå at odds værdien bliver under 0, kan logit værdien beregnes, og den beregnes ved at tage logaritmen til oddset. Odds værdier mindre end 1,0 vil have negative logit værdier, hvorimod odds værdier over 1,0 vil have positive logit værdier, og en odds værdi på 1,0 har en logit værdi på 0 (Hair et al., 2010). Dette medfører, at uanset hvor lave de negative logit værdier bliver, er det stadig muligt at transformere dem til en odds værdi, der er større end 0, og det sker ved at tage antilogaritmen til logit værdien. Anvendes logit værdien, medfører det, at man har en tal-variabel, der både kan have positive og negative værdier, og som altid kan transformeres tilbage til en sandsynlighedsværdi, der er mellem 0 og 1. Det er dog ikke muligt, at logit værdien kan ramme hverken 0 eller 1. Denne værdi bliver nu den afhængige variabel i den logistiske regressionsmodel (Hair et al., 2010).

6.3.2.4.1.2 Modelestimering

Estimeringsprocessen af de logistiske koefficienter minder om den metode, der anvendes i regression, bortset fra at den afhængige variabel i den logistiske regression kun anvender to aktuelle værdier, der er 0 eller 1. De estimerede koefficienter for de uafhængige variabler er estimeret ved at anvende enten logit værdien eller odds værdien som den afhængige variabel, og de to modelformuleringer illustreres her (Hair et al., 2010):

$$Logit_i = \ln\left(\frac{prob_{event}}{1 - prob_{event}}\right) = b_0 + b_1X_1 + \cdots + b_nX_n$$

$$Odds_i = \left(\frac{prob_{event}}{1 - prob_{event}}\right) = e^{b_0 + b_1X_1 + \cdots + b_nX_n}$$

Begge modelformuleringer er brugbare, men den der vælges påvirker, hvordan koefficienterne er estimeret. Denne proces kan rumme en eller flere uafhængige variabler, og den uafhængige variabel kan være enten tal eller ikke-tal (binær) (Hair et al., 2010).

Den ikke-lineære karakter af den logistiske transformation kræver, at der anvendes en anden procedure end den fra multipel regression. Den anvendte procedure er MLE, og den anvendes gentagne gange for at finde de mest egnede estimater for koefficienterne. Logistisk regression maksimerer således sandsynligheden for, at en begivenhed vil forekomme. Sandsynligheds-værdien anvendes derefter til at beregne et mål for modellens overordnede fit (Hair et al., 2010).

6.3.2.4.2 Vurdering af goodness-of-fit for den estimerede model

Goodness-of-fit for en logistisk regression kan vurderes på tre måder, hvoraf den første er statistiske mål, den anden er pseudo R^2 værdier, og den tredje metode er en undersøgelse af klassifikationsnøjagtigheden. På trods af at de tre metoder undersøger modellens fit fra forskellige perspektiver, burde de opnå ensartede konklusioner (Hair et al., 2010).

6.3.2.4.2.1 Statistiske mål

Sandsynlighedsværdien er det generelle mål for, hvor egnet modellen er. Logistisk regression måler modelestimeringens fit med værdien af -2 gange logaritmen af sandsynlighedsværdien, og det refereres til som *-2LL*. Minimumsværdien for *-2LL* er 0, hvilket svarer til, at modellen er perfekt egnet, og den situation opstår, når sandsynligheden er 1, hvorefter *-2LL* er 0. Jo lavere

værdien er af *-2LL*, jo bedre egnet er modellen (Hair et al., 2010). *-2LL* værdien kan bruges til at sammenligne ligninger ud fra ændringer i fit eller at kalkulere mål, der er sammenlignelige til R^2 i multipel regression.

Sandsynlighedsværdien kan blive sammenlignet mellem ligninger for at kunne vurdere forskellen i det forudsagte fit med statistiske test for signifikansen af disse forskelle. Fremgangsmåden følger nedenstående tre trin:

1. *Estimere en nul model*

   Det første trin er at kalkulere en nul model, der fungerer som udgangspunktet for at kunne sammenligne forbedringerne i modellens fit. Den mest almindelige nul model er en uden en eneste uafhængig variabel, da det gør det muligt at sammenligne den med enhver model, der indeholder uafhængige variabler (Hair et al., 2010).
2. *Estimere den foreslåede model*

   Denne model indeholder de uafhængige variabler, der skal inkluderes i den logistiske regressionsmodel. Forhåbentlig vil modellens fit forbedres fra nul modellen og resultere i en lavere *-2LL* værdi. Der kan estimeres en foreslået model for hver uafhængig variabel (Hair et al., 2010).
3. *Vurdere -2LL forskellen*

   Det sidste trin er at vurdere den statistiske signifikans af *-2LL* værdien mellem de to modeller, hhv. nul modellen og den foreslåede model. Hvis de statistiske test understøtter signifikante forskelle, kan der angives, at de uafhængige variabler i den foreslåede model signifikant forbedrer modelestimeringens fit (Hair et al., 2010).

Det er også muligt at sammenligne to foreslåede modeller, hvor forskellen i *-2LL* reflekterer forskellen i modellernes fit ud fra de forskellige modelspecifikationer. Det er fx muligt at sammenligne to foreslåede modeller med hhv. to og tre uafhængige variable, hvor der således undersøges den forbedring, som der opnår ved at tilføje en ekstra uafhængig variabel. I disse tilfælde vil en af modellerne blive udvalgt som nul modellen, hvorefter den sammenlignes med den anden model (Hair et al., 2010).

Chi-i-anden testen og de tilhørende test for statistisk signifikans bruges til at evaluere reduktionen af *-2LL* værdien. Det er dog vigtigt at være opmærksom på, at disse statistiske test er specielt følsomme over for stikprøvestørrelsen, og det gælder både meget små og meget store

stikprøvestørrelser. Derfor skal forskeren være forsigtig med at drage konklusioner kun ud fra signifikansniveauet af chi-i-anden testen i logistisk regression (Hair et al., 2010).

Der findes ligeledes en chi-i-anden baseret test, der er udviklet af Hosmer & Lemeshow, og den er en klassifikationstest, hvor alle tilfældene først opdeles i ti forholdsvis ens klasser. Derefter er antallet af faktiske og forudsagte begivenheder sammenlignet i hver klasse med chi-i-anden statistikken (Hair et al., 2010). Denne test giver et omfattende mål af den forudsigende nøjagtighed, der ikke er baseret på sandsynlighedsværdien, men derimod på den faktiske forudsigelse af den afhængige variabel. Anvendelsen af en chi-i-anden test kræver, som tidligere nævnt, at der er fokus på stikprøvestørrelsen (Hair et al., 2010).

6.3.2.4.2.2 Pseudo R^2 mål

Flere forskellige R^2 mål er blevet udviklet og præsenteret i diverse statistikprogrammer for at repræsentere det overordnede model fit. Disse pseudo R^2 mål fortolkes på samme måde som determinantkoefficienten i multipel regression. En pseudo R^2 værdi kan afledes af den logistiske regression på samme måde som R^2 værdien i en regressionsanalyse (Hair et al., 2010). En pseudo R^2 for en logit model (R^2_{LOGIT}) kan kalkuleres ud fra nedenstående formel:

$$R^2_{LOGIT} = \frac{-2LL_{null} - (-2LL_{model})}{-2LL_{null}}$$

Logit R^2 værdien varierer fra 0,0 til 1,0, og et perfekt fit har en *-2LL* værdi på 0,0 og en R^2_{LOGIT} på 1,0. Der findes to andre mål hhv. Cox & Snell R^2 og Nagelkerke R^2, der minder om pseudo R^2 værdien, og de er således også kategoriseret som pseudo R^2 mål. Cox & Snell R^2 målet fungerer på tilsvarende måde, hvor højere værdier indikerer et bedre model fit. Dette mål er dog begrænset således, at det ikke kan nå maksimumværdien på 1. Derfor foretog Nagelkerke en modifikation af Cox & Snell R^2, hvilket medførte en udvidelse af målets interval, således at det nu spænder fra 0 til 1, og det resulterede i Nagelkerke R^2. Ud fra de forskellige intervaller vil Nagelkerke R^2 altid være større end Cox & Snell R^2. Begge mål fortolkes ved, at de reflekterer den variationsmængde, der forklares med den logistiske regressionsmodel, hvor 1 indikerer et perfekt model fit (Hair et al., 2010).

6.3.2.4.2.3 Klassifikationsnøjagtighed

I logistisk regression anvendes en klassifikationsmatrix til at vurdere klassifikations-nøjagtigheden for den estimerede model. Klassifikationsmatrixen måler, hvor godt gruppemedlemskabet er forudsagt og udvikler en hitrate, der dækker over den procentdel, der er korrekt klassificeret (Hair et al., 2010). Hitraten kan sammenlignes med forskellige chance-relaterede mål fx *maximum chance* kriteriet og *proportional chance* kriteriet for at understøtte en vurdering af klassifikationsnøjagtigheden. Disse standarder tilhører oprindelig diskriminant analysen, men de anvendes ofte inden for logistisk regression. *Maximum chance* kriteriet placerer alle observationerne i den største gruppe, og det er det mest konservative mål, fordi det genererer den højeste standard for sammenligning. Den største gruppes procentmæssige andel af stikprøven udgør således *maximum chance* kriteriet, og klassifikationsnøjagtigheden bør således overstige dette kriterie (Hair et al., 2010). *Proportional chance* kriteriet anvendes, når gruppestørrelserne er uens, og der ønskes at klassificere alle observationer korrekt, og dermed ikke kun dem i den største gruppe. *Proportional chance* kriteriet bygger på denne formel:

$$C_{PRO} = p^2 + (1 - p^2)$$

Hvor p er forholdet af observationer i gruppe 1, og $(1 - p)$ er forholdet af observationer i gruppe 2 (Hair et al., 2010). Klassifikationsnøjagtigheden bør således også overstige dette kriterie, og generelt argumenteres der for, at begge kriterier bør overstiges med 25 %. De 25 % kan dog skabe en øvre grænse for *maximum chance* kriteriet, hvis den ene gruppe er meget større end den anden (Hair et al., 2010).

En klassifikationsmatrix afspejler også type 1 og type 2 fejl, hvor type 1 fejl er kendt som alfa (α), og type 2 fejl er kendt som beta (β). Type 1 fejl er sandsynligheden for at forkaste H_0-hypotesen, selvom den er sand, og dette betegnes generelt som falsk positiv. Ved at specificere et alfa niveau fastsætter forskeren den tilladte grænse for type 1 fejl, og indikerer sandsynligheden for at konkludere, at signifikansen eksisterer, når den faktisk ikke gør (Hair et al., 2010). Når niveauet for type 1 fejl specificeres, afgør forskeren også en associeret fejl, der betegnes som type 2 fejl. Type 2 fejl er sandsynligheden for at undlade at forkaste H_0-hypotesen, selvom den er falsk. Type 1 og type 2 fejl er omvendt relateret, derfor er det ikke muligt, at begge fejltyper kan være lave, hvorfor forskeren er nødt til at finde en passende balance. En tilføjelse til type 2 fejl er 1-β, der refereres til som power, der illustrerer styrken af den statistiske inferens test. Power er sandsynligheden for korrekt at afvise H_0-hypotesen, når den skal afvises.

Derfor er power således sandsynligheden for, at statistisk signifikans vil blive indikeret, hvis den er tilstede (Hair et al., 2010). Forholdet mellem de forskellige fejl sandsynligheder ifm. at teste for forskellen i to middelværdier ses i Figur 9.

		Reality	
		No Difference	*Difference*
Statistical Decision	H_0: *No Difference*	**1 – α**	**β** **Type II error**
	H_a: *Difference*	**α** **Type I error**	**1 – β** **Power**

Figur 9 – Type 1 og Type 2 fejl (Hair et al., 2010, s. 10)

Ofte anvendes der mere end en af de tre præsenterede metoder til at undersøge modellens fit, da der forhåbentligt vil være et sammenfald af indikatorerne fra disse metoder, der vil give forskeren den nødvendige støtte til at evaluere modellens overordnede fit (Hair et al., 2010).

6.3.2.5 Trin 5 – Fortolkning af resultaterne

Fortolkningen af koefficienterne adskiller logistisk regression fra multipel regression, da den afhængige variabel er blevet transformeret i et af de tidligere trin, og derfor skal koefficienterne evalueres på en speciel måde. Efterfølgende vil der fokuseres på, hvordan retningen og størrelsen af koefficienterne er bestemt (Hair et al., 2010).

6.3.2.5.1 Kontrol af koefficienternes signifikans

Logistisk regression tester hypoteser om individuelle koefficienter, hvilket også er tilfældet med multipel regression. I multipel regression omhandler den statistiske test, hvorvidt koefficienten er signifikant forskellig fra 0, da en koefficient på 0 indikerer, at den ikke har nogen påvirkning på den afhængige variabel (Hair et al., 2010). I logistisk regression anvendes der også en statistisk test for at se, hvorvidt den logistiske koefficient er forskellig fra 0. Her skal man dog være opmærksom på, at i logistisk regression anvendes logaritmen som den afhængige variabel. En værdi på 0 svarer således til et odds på 1,0 eller en sandsynlighed på 0,50, hvilket indikerer, at sandsynligheden er ens for hver gruppe, og derfor har den uafhængige variabel ingen effekt ift. at kunne forudsige gruppemedlemskabet (Hair et al., 2010).

I logistisk regression anvendes Wald teststørrelsen (Z^2) for at vurdere signifikansen af hver koefficient (Hair et al., 2010). Denne teststørrelse følger en chi-i-anden fordeling med en frihedsgrad på 1, $(df = 1)$, og Wald teststørrelsen bygger på formlen $Z^2 = \left(\frac{b_i}{SE(b_i)}\right)^2$, hvor b_i er estimatet af en logistisk koefficient, og SE(b_i) er standardfejlen til denne (Agresti & Finlay, 2008). Wald teststørrelsen afspejler den statistiske signifikans for hver estimerede koefficient, således at der kan foretages en hypotesetestning, hvilket også er tilfældet i multipel regression. Den logistiske koefficient er statistisk signifikant, hvis p-værdien er under signifikansniveauet, hvorefter H_0-hypotesen afvises (Agresti & Finlay, 2008). Hvis dette er tilfældet, kan man derefter fortolke den logistiske koefficient, hvorvidt den påvirker den estimerede sandsynlighed og dermed forudsigelsen af gruppemedlemskabet (Hair et al., 2010).

6.3.2.5.2 Fortolkning af koefficienterne

En fordel ved logistisk regression er, at der kun har behov for at vide, hvorvidt en begivenhed forekommer eller ej ift. at definere en dikotom værdi som den afhængige variabel. En begivenhed kan i dette tilfælde være foretagelse af køb eller ej samt en virksomheds succes eller fiasko. Når dataene analyseres ved at anvende den logistiske transformation, får den logistiske regression og dens koefficienter en anden betydning end ved en regression med en afhængig tal-variabel. Ud fra estimeringsprocessen fremgår det, at koefficienterne (B_0, B_1, B_2, …, B_n) er faktiske mål for ændringer i sandsynlighedsrationen (oddset). Logistiske koefficienter er dog svære at fortolke i deres originale form, fordi de er udtrykt i logaritmer, da der anvendes logaritmen som den afhængige variabel (Hair et al., 2010). De fleste statistikprogrammer angiver også en eksponentiel logistisk koefficient, der er en transformation, dvs. en antilogaritme, af den originale logistiske koefficient. Det er muligt at fortolke på enten de originale eller de eksponentielle logistiske koefficienter, men de to typer af logistiske koefficienter adskiller sig fra hinanden ved, at de reflekterer forholdet af den uafhængige variabel ud fra to forskellige typer af den afhængige variabel (Hair et al., 2010). Dette uddybes i Figur 10.

Logistic Coefficient	Reflects Changes in . . .
Original	Logit (log of the odds)
Exponentiated	Odds

Figur 10 – De to typer af logistiske koefficienter (Hair et al., 2010, s. 327)

Her fremgår det, at den originale koefficient reflekterer ændringer i logaritmen af odds, hvorimod den eksponentielle koefficient reflekterer ændringer i odds. Efterfølgende vil hver form for koefficient beskrives både ift. retningsbestemmelse og størrelse af de uafhængige variablers forhold, da hver koefficienttype kræver forskellige fortolkningsmetoder.

6.3.2.5.2.1 Retningsbestemmelse af forholdet

Retningen af forholdet reflekterer ændringerne i den afhængige variabel, der er forbundet med ændringer i den uafhængige variabel, og dette forhold kan være enten positivt eller negativt. Et positivt forhold betyder, at en stigning i den uafhængige variabel er forbundet med en stigning i den forudsagte sandsynlighed, og det modsatte er gældende ved et negativt forhold (Hair et al., 2010). Retningen af forholdet påvirkes dog forskelligt alt afhængigt af, hvilken type koefficient der anvendes, og derfor uddybes de to typer efterfølgende.

Fortegnet for de originale koefficienter indikerer retningen af forholdet, og dette fortegn kan være positivt eller negativt. En positiv koefficient øger sandsynligheden, hvorimod en negativ værdi formindsker den forudsagte sandsynlighed, hvilket skyldes, at de originale koefficienter er udtrykt i logit værdier. En logit værdi på 0,0 udlignes til en odds værdi på 1,0 og en sandsynlighed på 0,50, hvorimod de negative koefficienter relaterer sig til odds værdier mindre end 1,0 og sandsynligheder mindre end 0,50, og de positive koefficienter relaterer sig til odds værdier over 1,0 og sandsynligheder større end 0,50 (Hair et al., 2010).

De eksponentielle koefficienter skal fortolkes anderledes, da de er logaritmer af de originale koefficienter. Derfor angiver de eksponentielle koefficienter i realiteten odds værdier, hvilket betyder, at eksponentielle koefficienter ikke kan have en negativ værdi. Det skyldes, at logaritmen til 0, dvs. ingen effekt, er 1,0, og derfor vil en eksponentiel koefficient på 1,0 faktisk svare til et forhold med ingen retningsbestemmelse. Dette medfører, at eksponentielle koefficienter over 1,0 reflekterer et positivt forhold, hvorimod en værdi under 1,0 repræsenterer et negativt forhold (Hair et al., 2010).

6.3.2.5.2.2 Størrelsen på forholdet

Den numeriske værdi af koefficienten skal evalueres for at bestemme, hvor meget sandsynligheden vil ændre sig ved en ændring på en enhed i den uafhængige variabel. Ved tal-variabler kigges der ofte på, hvor meget den estimerede sandsynlighed vil ændre sig for hver enhedsændring i den uafhængige variabel. I logistisk regression er der som bekendt et ikke-

lineært forhold, der er begrænset mellem 0 og 1, og derudover skal der tages højde for både de originale og de eksponentielle koefficienter (Hair et al., 2010). De originale logistiske koefficienter er mest hensigtsmæssige at anvende til at afgøre retningsbestemmelsen af forholdet, men derimod er de mest uhensigtsmæssige at anvende til at afgøre størrelsen af forholdet. Det skyldes, at de reflekterer ændringen i logit værdien, der er en måleenhed, der ikke specifikt forklarer, hvor meget sandsynligheden faktisk ændrer sig. Eksponentielle logistiske koefficienter reflekterer direkte størrelsen af ændringen i odds værdien, fordi de er eksponenter, der fortolkes lidt anderledes. Deres indflydelse er multiplikativ, hvilket betyder, at koefficienternes effekt (oddset) ikke er tilføjet til den afhængige variabel, men derimod multipliceret for hver enhedsændring i den uafhængige variabel. En alternativ og formentlig nemmere metode til at bestemme den mængde som sandsynligheden ændres med er følgende formel (Hair et al., 2010):

$$Percentage\ change\ in\ odds = (Exponentiated\ coefficient_i - 1{,}0) * 100$$

Det er også muligt, at oversætte odds værdier til sandsynlighedsværdier ud fra denne simple formel (Hair et al., 2010):

$$Probability = \frac{Odds}{(1 + Odds)}$$

6.3.2.6 Trin 6 – Validering af resultaterne

Det sidste trin i en logistisk regressionsanalyse indeholder en sikring af den eksterne og interne validitet af resultaterne. På trods af at logistisk regression ikke er modtagelig ift., at der sker et overfit af resultaterne, er valideringsprocessen stadig essentiel især ved små stikprøver. Den mest almindelige metode til at etablere ekstern validitet er en vurdering af hitraterne gennem en separat stikprøve (hold-out stikprøve) eller ved anvendelse af en procedure, der gentange gange behandler estimeringsstikprøven. Ekstern validitet er understøttet, når hitraten af den udvalgte metode overstiger de sammenlignelige standarder, der repræsenterer den forudsigende nøjagtighed, der er forventet tilfældigt (Hair et al., 2010). Hvis hold-out stikprøven stammer fra den oprindelige stikprøve, etablerer denne tilgang intern validitet og en begyndende indikation af ekstern validitet. Hvis hold-out stikprøven udgøres af en anden separat stikprøve evt. fra en anden population eller segment af populationen, adresserer valideringen mere den eksterne validitet af resultaterne (Hair et al., 2010).

En af de mest almindelige valideringsformer er, som tidligere nævnt, gennem frembringelsen af en hold-out stikprøve, der også betegnes som valideringsstikprøven, og denne er adskilt fra analysestikprøven, der er anvendt til at estimere modellen. Formålet er at anvende den logistiske regressionsmodel til et helt andet sæt af respondenter for at vurdere niveauerne af den opnåede forudsigende nøjagtighed, idet disse respondenter ikke er inddraget i estimeringsprocessen, burde det give indsigt ift. generaliserbarheden af den logistiske regressionsmodel (Hair et al., 2010).

6.4 Operationalisering af teorien

De præsenterede teorier inden for konkurs, nøgletal og logistisk regression skal være behjælpelige med at besvare indeværende speciales problemformulering. Den generelle teori inden for konkurs bidrager ikke direkte til besvarelsen af problemformuleringen, i stedet giver den en generel forståelse af, hvad det betyder, at en virksomhed er opløst efter konkurs samt processen herved.

Begrundelsen for at anvende Altman & Sabatos (2007) fem nøgletal er, at de jf. afsnit 6.2.1 *Udvælgelse af nøgletal* har foretaget en grunddig udvælgelsesproces for at identificere de mest anvendelige nøgletal. Dertil er Edward I. Altman, der er den ene af forfatterne i Altman & Sabato (2007), pioner inden for konkursprognosticering, da han som den første anvendte en multipel analyse i år 1968 (Altman, 1968). Derudover var Edward I. Altman ligeledes medvirkende til at udvikle en konkursmodel, der dannede grundlag for kommerciel udnyttelse (Christensen & Nielsen, 2012). Det vidner om, at forfatteren besidder et godt kendskab og en vis portion erfaring ift. at udarbejde konkursprognosticeringsmodeller, hvilket ligeledes er et argument for at anvende de nøgletal, som han har været med til at udvælge. Desuden dækker de fem nøgletal med hver deres nøgletalskategori de vigtigste aspekter af en virksomheds finansielle profil, hvilket også har haft betydning for valget af disse nøgletal (Altman & Sabato, 2007). Der er dog suppleret med soliditetsgraden, da det er vurderet, at de fem nøgletalskategorier ikke afdækker dette aspekt tilstrækkeligt. De seks nøgletal fokuserer på virksomhedens gearing, betalingsevne og finansielle styrke samt modstandskraft over for evt. tab, og en analyse af disse karaktertræk hos en virksomhed vurderes at være dækkende til at kunne foretage en analyse af virksomhedens økonomiske situation. Derfor anvendes disse kvantitative nøgletal til udarbejdelse af en konkursprognosticeringsmodel i indeværende speciale, hvilket understreger afgrænsningen fra at anvende kvalitative nøgletal jf. afsnit 2

Problemfelt. De udvalgte seks nøgletal er hhv. STD_EBV, C_TA, EBITDA_TA, RE_TA, EBITDA_IE og EBV_TA.

Logistisk regression benyttes som den valgte multivariate teknik til udførelsen af den statistiske analyse, idet logistisk regression ikke har strenge forudsætninger, der skal være opfyldt, hvilket gør den anvendelig til indeværende speciale. Derudover minder den om multipel regression, da den har lignende statistiske test, hvilket letter arbejdet med den logistiske regression. Logistisk regression er ligeledes den mest anvendte multivariate teknik til udarbejdelse af konkursprognosticeringsmodeller, hvilket understøtter anvendelsen af denne teknik (Altman & Sabato, 2007). Derudover passer den logistiske regression godt til karakteristikaene af datagrundlaget og hensigten med analysen, idet analysen omhandler en forudsigelse af et binært resultat, der i dette tilfælde er, hvorvidt en virksomhed er i normal drift eller opløst efter konkurs. Det binære resultat understreges ved, at en virksomhed ikke kan være i normal drift, samtidig med at den er opløst efter konkurs. Formålet med at anvende en logistisk regression er således, at den skal kunne forudsige de enkelte virksomheders gruppemedlemskab ud fra de udvalgte nøgletal, da det er af afgørende betydning for at kunne besvare specialets problemformulering.

Den valgte metode til at udføre modelsøgningen af den logistiske regression i SPSS er forward stepwise (Wald), og denne metode er valgt, da Hair et al. (2010) ligeledes benytter en stepwise metode, og derudover er den en meget pædagogisk metode. En stepwise metode er en gentagende procedure, der skiftevis tilføjer og fjerner en uafhængig variabel. Beslutningen om at tilføje eller fjerne en variabel er foretaget ud fra, hvorvidt variablen forbedrer modellen (Keller, 2011). Der kan anvendes forskellige kriterier for at guide implementeringen, hhv. største reduktion i *-2LL* værdien, største Wald koefficient eller højest betingede sandsynlighed (Hair et al., 2010). Stepwise metoden starter med en tom basismodel, hvor ingen uafhængige variabler er medtaget. Derefter tilføjes den uafhængige variabel med den mindste p-værdi under en fastsat grænse for signifikansniveauet. Efterfølgende foretages der en ny analyse for at undersøge, hvorvidt der skal tilføjes en af de resterende variabler til modellen eller fjernes en variabel i modellen. En variabel fjernes, hvis dens p-værdi er over en fastsat grænse for signifikansniveauet. Denne vekslen mellem at tilføje og fjerne en variabel fortsætter indtil ingen variabler hverken tilføjes eller fjernes (Keller, 2011). I stepwise metoden er det vigtigt, at signifikansniveauet for at implementere en model er mindre end signifikansniveauet for at fjerne en variabel, da modellen ellers vil køre i ring. Den anvendte metode til indeværende analyse er

dog forward stepwise, hvorfor metoden formentlig er en blanding mellem forward og stepwise. Det kan få den betydning, at der evt. ikke fjernes variabler fra modellen, men at der blot trinvist tilføjes variabler ud fra den mindste p-værdi ift. den fastsatte grænse. Der anvendes metoden forward stepwise til modelsøgningen, da der i SPSS under logistisk regression ikke findes en metode, der kun betegnes stepwise.

Udgangspunktet for den statistiske analyse er således de seks udvalgte kvantitative nøgletal, den logistiske regression som multivariat teknikken og metoden forward stepwise (Wald) til modelsøgningen i SPSS.

7 Analyse

Analysen indledes med et afsnit om datagrundlaget for analysen, hvori bearbejdningen af de indsamlede data gennemgås, hvorefter der foretages en analyse af stikprøvens deskriptive statistik. Derefter udarbejdes der en logistisk regression, som følger sekstrinsmodellen, der tidligere er præsenteret som analysestrategien for specialet.

7.1 Datagrundlag

Afsnittet er opdelt i to underafsnit, hvor det første afsnit beskriver processen fra de rå datasæt til de udregnede nøgletal med udgangspunkt i Bilag 3, og i det andet afsnit udarbejdes der en deskriptiv statistik af stikprøven.

7.1.1 Databearbejdning

Metoden til at fremskaffe de to rå datasæt i Bilag 1 og Bilag 2 blev præsenteret i afsnit 5.2.1 *Dataindsamling*, og de udgør, som tidligere nævnt, datagrundlaget for indeværende analyse. Bilag 1 består af 223 virksomheder, der er opløst efter konkurs, og Bilag 2 består af 305 virksomheder i normal drift. Databearbejdningen af de to datasæt er foretaget i Bilag 3, hvor begge datasæt er integreret i hvert deres ark, hhv. *Input, Opløst efter konkurs* og *Input, Normal drift*, og begge input-ark er blevet organiseret hensigtsmæssigt ift. overskrifter og årstal. Bilag 1 indeholder data fra år 2010 til år 2014, hvorimod Bilag 2 indeholder data fra år 2011 til år 2015. Derfor har det i Bilag 3 været nødvendigt at fjerne år 2010 i arket *Input, Opløst efter konkurs* samt år 2015 i arket *Input, Normal drift*, da det medfører en ensartet tidsperiode for datasættene. I afsnit 5.2.1 *Dataindsamling* fremgår det, at der er forsøgt at ensarte tidsperioderne for de to datasæt, idet der i søgekriteriet for virksomheder i normal drift er tilføjet en begrænsning ift. etableringsdato samt dato for regnskabsafslutning. Begrænsningerne medførte dog i stedet, at der i Bilag 2 blev eksporteret en tom kolonne for år 2015 frem for, at perioden blev ændret til år 2010 til år 2014. Disse uoverensstemmelser mellem de to rå datasæt har medført, at udgangspunktet for valg af årstal er begrænset til perioden fra år 2011 til år 2014.

I Bilag 3 er der udarbejdet et ark, der er betegnet *List*, og dette ark har flere funktioner. Her er der bl.a. foretaget en opgørelse over antallet af virksomheder, der er opløst efter konkurs, inden for forskellige regnskabsposter for hvert år i tidsperioden. Derefter er opgørelsen af de enkelte årstal sammenlignet og vurderet, hvor det fremgår, at år 2011 har et betydeligt højere antal

virksomheder inden for de enkelte regnskabsposter end de resterende årstal. Dette er årsagen til, at der benyttes år 2011 som udgangspunkt for analysen jf. afsnit 2 *Problemfelt*. Valg af årstal er foretaget på baggrund af konkursramte virksomheder, da arket *Input, Opløst efter konkurs* indeholder færre virksomheder end arket *Input, Normal drift*. Derudover er det sjældent, at virksomhederne i arket *Input, Opløst efter konkurs* er repræsenteret med regnskabsdata for alle årene, idet langt de fleste af virksomhederne er gået konkurs i løbet af tidsperioden. I arket *List* er der ligeledes beregnet aktivsummen for konkursramte virksomheder og virksomheder i normal drift, da den ikke fremgår direkte af input-arkene i Bilag 3, samtidig med at den jf. afsnit 6.2.3 *Nøgletallenes betydning* indgår i kalkulationen af flere af nøgletallene til analysen. Aktivsummen er kun udregnet for år 2011, da det, som tidligere nævnt, er det udvalgte årstal til analysen. I afsnit 5.2.1 *Dataindsamling* blev eksportkriteriet for de to rå datasæt præsenteret, hvor det fremgik, at alle tilgængelige aktivposter for de to datasæt er eksporteret. Efterfølgende blev det klart, at mange af aktivposterne blot udspecificerede posterne anlægsaktiver og omsætningsaktiver. Derfor udregnes aktivsummen i Bilag 3 kun vha. aktivposterne anlægsaktiver og omsætningsaktiver.

I arkene *Konkursdata, 2011* og *Normal drift data, 2011* i Bilag 3 er værdierne af de anvendte regnskabsposter for år 2011 samlet. Regnskabsposterne er alle i 1.000 kr., og de udgøres af EBITDA, finans udgifter, overført resultat, aktivsum, likvider, egenkapital og kortfristet gæld, der blev præsenteret i afsnit 6.2.2 *Elementer i nøgletallene*. Derudover indeholder *Konkursdata, 2011* og *Normal drift data, 2011* generelle informationer om virksomhederne, og disse informationer udgøres af CVR-nr., antal ansatte, etableringsdato, virksomhedsform, primær branchekode og primær branchebetegnelse. Det er ligeledes i *Konkursdata, 2011* og *Normal drift data, 2011*, at der er foretaget en frasortering af de virksomheder, der ikke har en værdi i alle regnskabsposterne, dvs. hvis feltet for fx EBITDA er blankt eller indeholder et nul er virksomheden frasorteret. Denne frasortering er nødvendig ift., at der skal udregnes nøgletal ud fra disse regnskabsposter, hvorfor der bør være en værdi i både tæller og nævner. Der kan dog argumenteres for, at værdien på nul i visse tilfælde kunne være bibeholdt, hvis den kun indgår som tæller i brøken for nøgletalsudregningen. På trods af dette er virksomheder med værdien på nul frasorteret, da det har været svært at vurdere, hvorvidt værdien på nul afspejler en manglende indberetning eller ingen værdi for den pågældende regnskabspost. Den manglende indberetning af regnskabsposter kan skyldes, at der er forskellige regnskabskrav til de forskellige regnskabsklasser, som påvirker en virksomheds rapportering af årsregnskabet (Christensen &

Nielsen, 2012). Beslutningen om at fjerne virksomheder med en regnskabspost på nul kan have medført bias ift. stikprøven, derimod er det vurderet, at hvis disse virksomheder var bibeholdt i stikprøven, kunne det ligeledes have medført bias. Virksomhederne, der fremgår af arkene *Konkursdata, 2011* og *Normal drift data, 2011*, er således dem, der er fundet egnet til analysen.

Det illustreres i Figur 11, at der er frasorteret 113 virksomheder, der er opløst efter konkurs, hvorimod der kun er frasorteret 34 virksomheder i normal drift. Det medfører, at den totale stikprøve er reduceret med 28 % til 381 virksomheder, idet antallet af konkursramte virksomheder er reduceret med hele 51 %, og virksomheder i normal drift er reduceret med 11 %. Den høje frasorteringsprocent for konkursramte virksomheder skyldes, som tidligere nævnt, at der er frasorteret alle virksomheder, der i en eller flere regnskabsposter har et blankt felt eller en værdi på nul.

Status	Antal	Frasorteret	Frasorteret i pct.	Total
Opløst efter konkurs	223	113	51 %	110
Normal drift	305	34	11 %	271
I alt	528	147	28 %	381

Figur 11 – Frasortering af virksomheder (Egen tilvirkning)

Regnskabsposten finans udgifter i arkene *Konkursdata, 2011* og *Normal drift data, 2011* i Bilag 3 har oprindeligt et negativt fortegn, hvilket ses i input-arkene i Bilag 3. Af regnetekniske årsager er værdien af denne regnskabspost konverteret til et positivt tal, idet det ellers ville medføre et falsk positivt resultat af nøgletallet EBITDA_IE, da EBITDA flere steder har en negativ værdi. Konverteringen af fortegnet for regnskabsposten finans udgifter ændrer ikke, at finans udgifter stadig er en omkostningspost fra resultatopgørelsen jf. afsnit 6.2.2.6 *Finans udgifter*. En analyse af regnskabsposternes værdier foretages efterfølgende i afsnit 7.1.2.5 *Regnskabsposter*.

Konkursdekretet for de konkursramte virksomheder, der indgår i analysen, er tilføjet i arket *Konkursdata, 2011* i Bilag 3, og konkursdekretet er fundet via hjemmesiden konkurser.dk, hvor det er muligt at finde en virksomheds konkursdekret mere end tre år tilbage (Auktioner ApS, 2016). Kolonnen med konkursdekreterne i arket *Konkursdata, 2011* er markeret med blå for at understrege, at disse data ikke er en del af det rå datasæt fra Bilag 1. Konkursdekretet er jf. afsnit 6.1 *Konkurs* skifterettens erkendelse af, at en skylder tages under konkursbehandling, og

konkursen indtræder således i det øjeblik, at konkursdekretet afsiges. Derfor anvendes virksomhedernes konkursdekret i indeværende speciale som en indikation af, hvornår de pågældende virksomheder er gået konkurs.

Nøgletallene til analysen er udregnet i Bilag 3 i arkene *Konkursnøgletal, 2011* og *Normal drift nøgletal, 2011* ud fra de formler, der er præsenteret i afsnit 6.2.3 *Nøgletallenes betydning*. De udregnede nøgletal er visualiseret med fem decimaler, da nogle af nøgletallenes værdier er meget lave. Nøgletallene i arket *Konkursnøgletal, 2011* er udregnet på baggrund af data fra arket *Konkursdata, 2011* i Bilag 3, hvorimod nøgletallene i arket *Normal drift nøgletal, 2011* er udregnet på baggrund af data fra arket *Normal drift data, 2011* i Bilag 3. En mere detaljeret analyse af de udregnede nøgletal foretages i afsnit 7.1.2.6 *Nøgletal*.

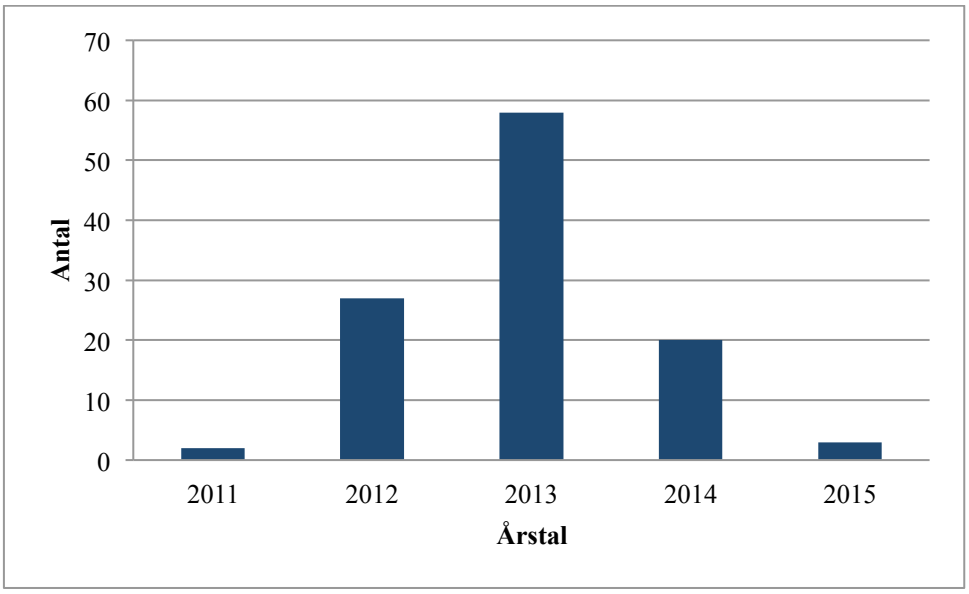
7.1.2 Deskriptiv statistik

I en deskriptiv statistik af stikprøven analyseres den ud fra optællinger, grafiske fremstillinger og beskrivende størrelser som fx gennemsnit og median (Clement & Ingemann, 2011). Denne analyse foretages for at give et overblik over stikprøvens karakteristika, herunder de to gruppers karakteristika samt for at vurdere, hvor repræsentativ stikprøven er for populationen, og hvor meget de to grupper ligner hinanden. Det er den præsenterede stikprøve fra forrige afsnit, der foretages en analyse af, og den består, som tidligere nævnt, af 381 virksomheder med regnskabsdata for år 2011. Populationen for indeværende speciale er alle danske virksomheder, der har aflagt et regnskab for år 2011. Derfor sammenlignes flere af stikprøvens karakteristika med informationer om alle danske virksomheder i år 2011, og disse informationer er erhvervet gennem Danmarks Statistik og den tilhørende Statistikbank (Danmarks Statistik, 2016). Den deskriptive statistik opdeles i konkurser, brancher, virksomhedsformer, medarbejderantal og etableringsår, regnskabsposter samt nøgletal.

7.1.2.1 Konkurser

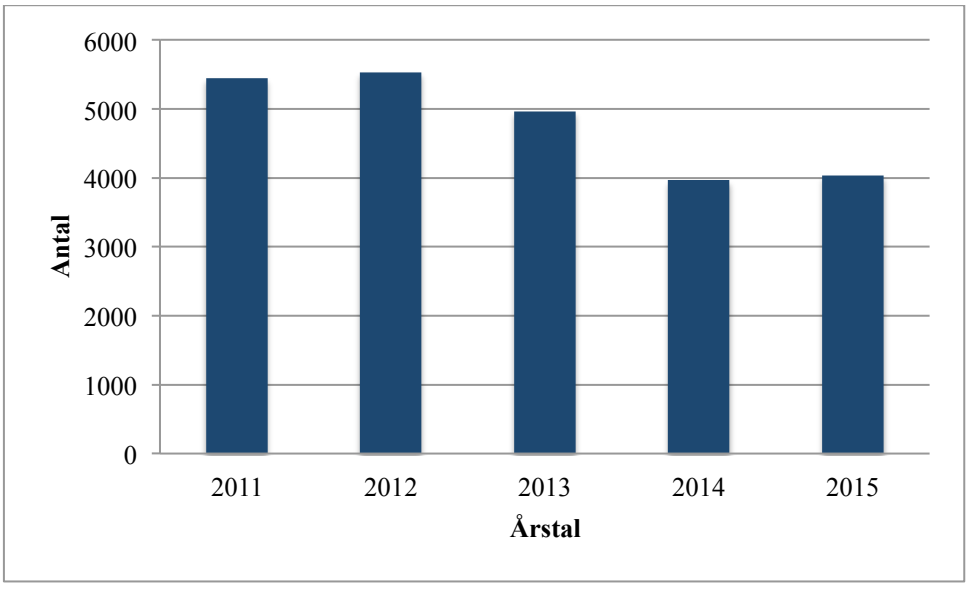
Fordelingen af stikprøvens konkurser illustreres i Figur 12, og den er udarbejdet ud fra konkursdekretet i arket *Konkursdata, 2011* i Bilag 3. Det skyldes, som tidligere nævnt, at konkursdekretet er skifterettens erkendelse af, at en virksomhed tages under konkursbehandling. Figur 12 illustrerer, at størstedelen af stikprøvens konkurser har fundet sted i år 2013, hvorimod de færreste konkurser er sket i år 2011. Ud fra Figur 12 er antallet af konkurser stigende fra år 2011 til år 2013, hvorefter antallet er faldet fra år 2013 til år 2015. Virksomhederne i stikprøven

er således gået konkurs over en periode på fem år, hvilket kan have betydning for analysens resultater, da det kan være svære at forudsige en virksomheds konkurs flere år ude i fremtiden.



Figur 12 – Konkurser i stikprøven, Data: Bilag 3 (Egen tilvirkning)

I Figur 13 illustreres det totale antal af konkurser i Danmark for den samme periode, som stikprøvens konkurser strækker sig over. På trods af en stigning i antallet af konkurser fra år 2011 til år 2012 er antallet af konkurser faldende fra år 2012 til år 2015, hvorefter antallet af konkurser er stigende fra år 2014 til år 2015. Udviklingen fra år 2014 til år 2015 stemmer overens med den udvikling, der blev beskrevet i afsnit 1 *Indledning*.



Figur 13 – Konkurser i Danmark, Data: Danmarks Statistik, 2016, KONK9 (Egen tilvirkning)

Udviklingen i Figur 13 stemmer dog ikke overens med udviklingen i Figur 12, hvilket indikerer, at udviklingen i stikprøvens konkurser ikke afspejler den generelle udvikling i antallet af konkurser i Danmark. Denne uoverensstemmelse indikerer, at stikprøvens udvikling i antallet af konkurser ikke er repræsentativ for populationens udvikling, og det kan få betydning for analysens resultater.

7.1.2.2 Brancher

Stikprøvens brancher indikeres af den primære branchekode i arkene *Konkursdata, 2011* og *Normal drift data, 2011* i Bilag 3, og brancherne udgøres af 22 hovedafdelinger, der er inddelt på baggrund af hovedgruppernes fordeling. Brancheinddelingen følger Dansk Branchekode 2007 (DB07), der er en dansk underopdeling af EU's fælles branchenomenklatur (Danmarks Statistik, 2007). Stikprøvens branchefordeling ses i Figur 14, der illustrerer hver gruppes antal og andel samt stikprøvens samlede branchefordeling. Det fremgår af Figur 14, at *Bygge- og anlægsvirksomhed*, *Transport og godshåndtering*, *Overnatningsfaciliteter og restaurations-virksomhed*, *Information og kommunikation* samt *Liberale, videnskabelige og tekniske tjenesteydelser* er overrepræsenteret i gruppen opløst efter konkurs ift. gruppen i normal drift. Derimod er *Engroshandel og detailhandel; reparation af motorkøretøjer og motorcykler* væsentligt underrepræsenteret i gruppen opløst efter konkurs, hvilket skyldes, at denne branche udgør hele 62 % af gruppen i normal drift. Generelt er ni brancher ikke repræsenteret i gruppen opløst efter konkurs, hvorimod det kun er syv brancher for gruppen i normal drift, og dette er angivet ud fra antal og ikke procentsats. *Engroshandel og detailhandel; reparation af motorkøretøjer og motorcykler* udgør over 50 % af den samlede stikprøve, og de næststørste brancher er *Fremstillingsvirksomhed* samt *Bygge- og anlægsvirksomhed*, der hver udgør 8 %, hvilket umiddelbart indikerer en skævvridning i stikprøven. I den samlede stikprøve er seks brancher ikke repræsenteret, hvis der tages udgangspunkt i antal.

Brancher	Opløst efter konkurs		Normal drift		Total	
	Antal	Pct.	Antal	Pct.	Antal	Pct.
Landbrug, jagt, skovbrug og fiskeri	1	1 %	9	3 %	10	3 %
Råstofindvinding	0	0 %	0	0 %	0	0 %
Fremstillingsvirksomhed	8	7 %	24	9 %	32	8 %
El-, gas- og fjernvarmeforsyning	0	0 %	2	1 %	2	1 %
Vandforsyning; kloakvæsen, affaldshåndtering og rensning af jord og grundvand	0	0 %	1	0 %	1	0 %
Bygge- og anlægsvirksomhed	19	17 %	10	4 %	29	8 %
Engroshandel og detailhandel; reparation af motorkøretøjer og motorcykler	34	31 %	169	62 %	203	53 %
Transport og godshåndtering	10	9 %	2	1 %	12	3 %
Overnatningsfaciliteter og restaurationsvirksomhed	9	8 %	4	1 %	13	3 %
Information og kommunikation	7	6 %	3	1 %	10	3 %
Pengeinstitut- og finansvirksomhed, forsikring	5	5 %	13	5 %	18	5 %
Fast ejendom	1	1 %	7	3 %	8	2 %
Liberale, videnskabelige og tekniske tjenesteydelser	9	8 %	10	4 %	19	5 %
Administrative tjenesteydelser og hjælpetjenester	3	3 %	9	3 %	12	3 %
Offentlig forvaltning og forsvar; socialsikring	0	0 %	0	0 %	0	0 %
Undervisning	1	1 %	0	0 %	1	0 %
Sundhedsvæsen og sociale foranstaltninger	3	3 %	5	2 %	8	2 %
Kultur, forlystelser og sport	0	0 %	0	0 %	0	0 %
Andre serviceydelser	0	0 %	3	1 %	3	1 %
Private husholdninger med ansat medhjælp; husholdningers produktion af varer og tjenesteydelser til eget brug, i.a.n.	0	0 %	0	0 %	0	0 %
Ekstraterritoriale organisationer og organer	0	0 %	0	0 %	0	0 %
Uoplyst	0	0 %	0	0 %	0	0 %
Total	110	100 %	271	100 %	381	100 %

Figur 14 – Branchefordeling i stikprøven, Data: Bilag 3 (Egen tilvirkning)

Brancheforde lingen i Danmark i år 2011 illustreres i Figur 15, hvor den største branche er *Engroshandel og detailhandel; reparation af motorkøretøjer og motorcykler*, der udgør 15 %. De næststørste brancher er *Landbrug, jagt, skovbrug og fiskeri*, *Liberale, videnskabelige og tekniske tjenesteydelser* samt *Bygge- og anlægsvirksomhed*, der udgør hhv. 11 %, 11 % og 10 %.

Brancher	**Antal**	**Pct.**
Landbrug, jagt, skovbrug og fiskeri	32705	11 %
Råstofindvinding	214	0 %
Fremstillingsvirksomhed	15715	5 %
El-, gas- og fjernvarmeforsyning	1793	1 %
Vandforsyning; kloakvæsen, affaldshåndtering og rensning af jord og grundvand	2590	1 %
Bygge- og anlægsvirksomhed	31575	10 %
Engroshandel og detailhandel; reparation af motorkøretøjer og motorcykler	44681	15 %
Transport og godshåndtering	12077	4 %
Overnatningsfaciliteter og restaurationsvirksomhed	13670	5 %
Information og kommunikation	14588	5 %
Pengeinstitut- og finansvirksomhed, forsikring	8983	3 %
Fast ejendom	27220	9 %
Liberale, videnskabelige og tekniske tjenesteydelser	31781	11 %
Administrative tjenesteydelser og hjælpetjenester	15856	5 %
Offentlig forvaltning og forsvar; socialsikring	284	0 %
Undervisning	5311	2 %
Sundhedsvæsen og sociale foranstaltninger	18677	6 %
Kultur, forlystelser og sport	6351	2 %
Andre serviceydelser	16227	5 %
Private husholdninger med ansat medhjælp; husholdningers produktion af varer og tjenesteydelser til eget brug, i.a.n.	235	0 %
Ekstraterritoriale organisationer og organer	12	0 %
Uoplyst	188	0 %
Total	300733	100 %

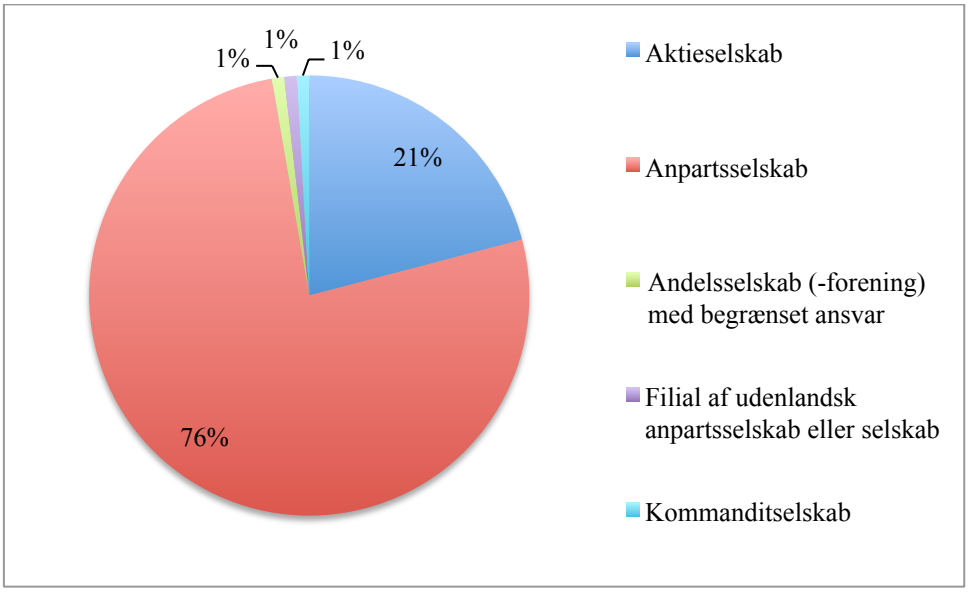
Figur 15 – Brancheforde ling i Danmark år 2011, Data: Danmarks Statistik, 2016, GF2 (Egen tilvirkning)

Brancheforde lingen i Figur 15 indikerer, at branchen *Engroshandel og detailhandel; reparation af motorkøretøjer og motorcykler* er meget overrepræsenteret i stikprøven, idet den, som tidligere nævnt, udgør over halvdelen af stikprøven. En overrepræsenteret branche i stikprøven kan få stor betydning for resultaterne af den statistiske analyse, idet en generalisering kompliceres. Derimod er bl.a. brancherne *Landbrug, jagt, skovbrug og fiskeri*, *Fast ejendom*,

Liberale, videnskabelige og tekniske tjenesteydelser samt *Administrative tjenesteydelser og hjælpetjenester* væsentligt underrepræsenteret i stikprøven ift. branchefordelingen i Danmark, hvilket ligeledes kan have betydning for en generalisering af analysens resultater. Denne skævvridning i branchefordelingen mellem stikprøven og den generelle branchefordeling i Danmark medfører, at stikprøven til analysen ikke er repræsentativ for populationen ift. branchefordelingen. Dette kan få betydning for den efterfølgende statistiske analyse samt ikke mindst anvendelsen af resultaterne.

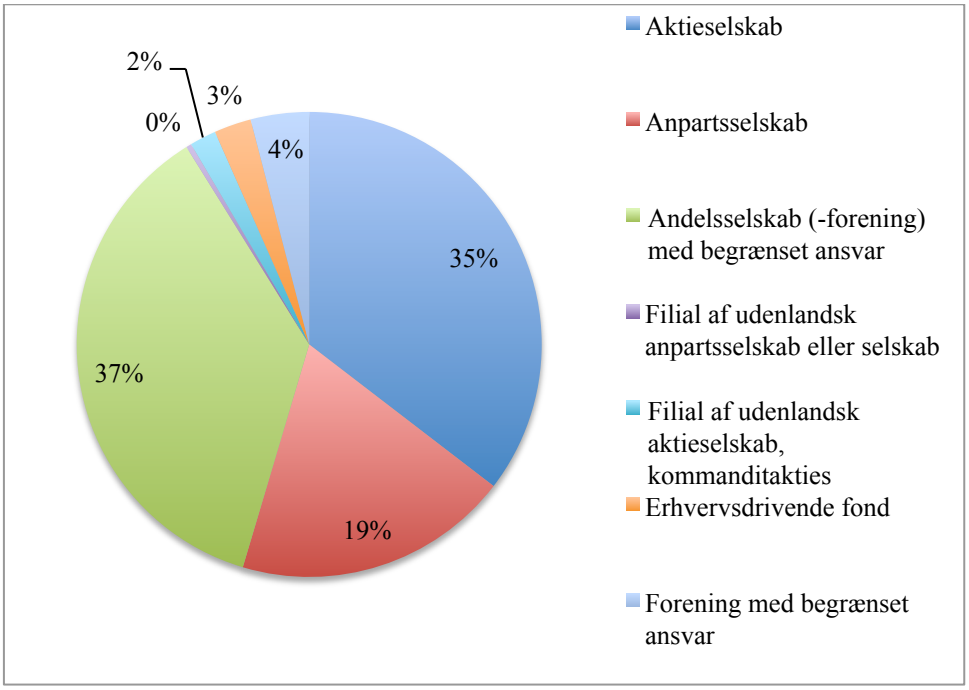
7.1.2.3 Virksomhedsformer

Stikprøvens virksomhedsformer fremgår også af arkene *Konkursdata, 2011* og *Normal drift data, 2011* i Bilag 3, hvorfor disse ligeledes analyseres ift. virksomhedsformerne i Danmark i år 2011. Figur 16 illustrerer de virksomhedsformer, der er repræsenteret i gruppen opløst efter konkurs. Hovedparten af virksomhedernes virksomhedsform er *anpartsselskab*, idet den udgør 76 %. Derudover udgør virksomhedsformen *aktieselskab* 21 %, hvorimod de resterende tre virksomhedsformer udgør i alt 3 %.



Figur 16 – Virksomhedsformer i opløst efter konkurs, Data: Bilag 3 (Egen tilvirkning)

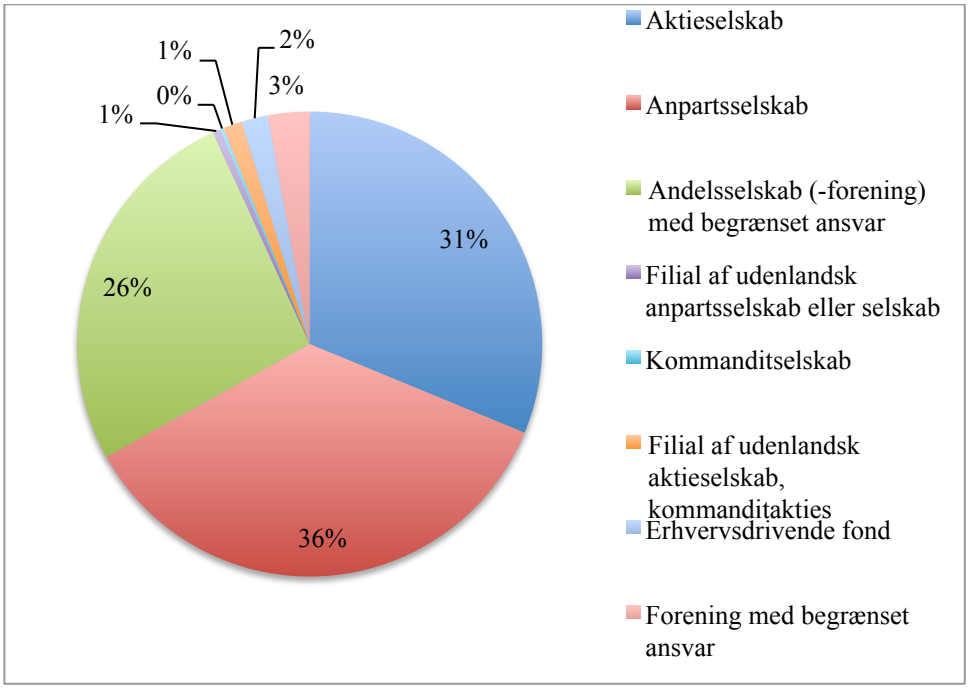
Virksomhedsformerne for gruppen i normal drift visualiseres i Figur 17, hvor *aktieselskab* udgør 35 %, hvorimod *anpartsselskab* udgør 19 %. Derimod udgør *andelsselskab (-forening) med begrænset ansvar* 37 %, hvilket dermed er den hyppigste virksomhedsform i normal drift. De resterende fire virksomhedsformer udgør til sammen 9 %.



Figur 17 – Virksomhedsformer i normal drift, Data: Bilag 3 (Egen tilvirkning)

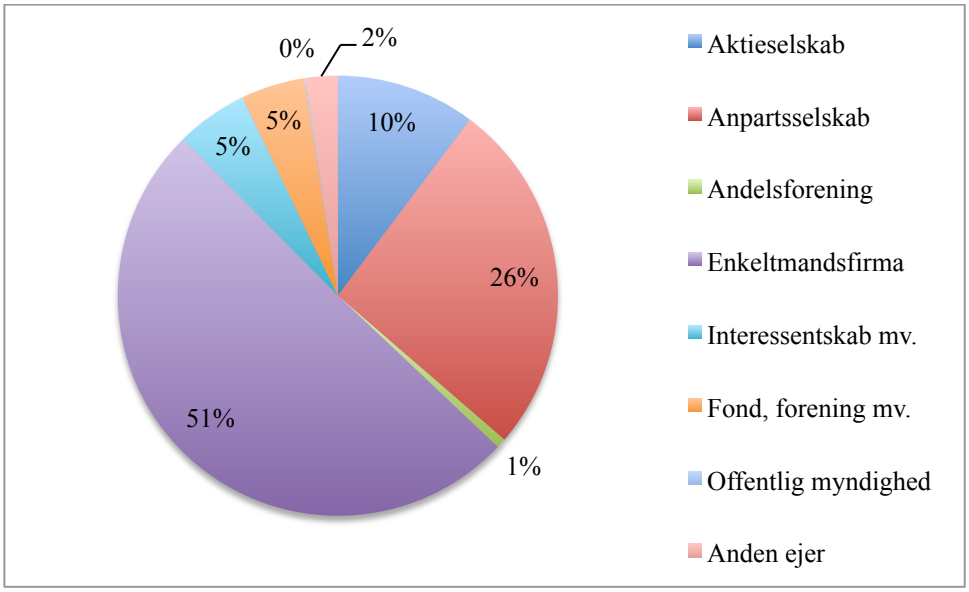
Det fremgår af arket *Normal drift data, 2011* i Bilag 3, at *andelsselskab (-forening) med begrænset ansvar* ofte er tilknyttet købmænd og døgnkiosker samt supermarkeder. Det indikerer, hvorfor det er den mest benyttede virksomhedsform, da *Engroshandel og detailhandel; reparation af motorkøretøjer og motorcykler*, som tidligere nævnt, er den største branche i gruppen for normal drift. Resultaterne af Figur 16 og Figur 17 indikerer, at der er en markant skævvridning mellem de to gruppers virksomhedsformer, idet virksomhedsformen *anpartsselskab* dominerer de konkursramte virksomheder, hvorimod *aktieselskab* og *andelsselskab (-forening) med begrænset ansvar* præger virksomhederne i normal drift.

Figur 18 illustrerer virksomhedsformerne i den samlede stikprøve, og her er det *aktieselskab*, *anpartsselskab* og *andelsselskab (-forening) med begrænset ansvar*, der er dominerende med hhv. 31 %, 36 % og 26 %. De resterende fem virksomhedsformer benyttes dermed kun i 7 % af virksomhederne i stikprøven.



Figur 18 – Virksomhedsformer i stikprøven, Data: Bilag 3 (Egen tilvirkning)

Fordelingen af virksomhedsformerne i Danmark i år 2011 fremgår af Figur 19, hvor over halvdelen af virksomhederne benytter virksomhedsformen *enkeltmandsfirma*. Derimod udgør *aktieselskab* og *anpartsselskab* kun hhv. 10 % og 26 %, og de resterende virksomhedsformer udgør til sammen 13 %.

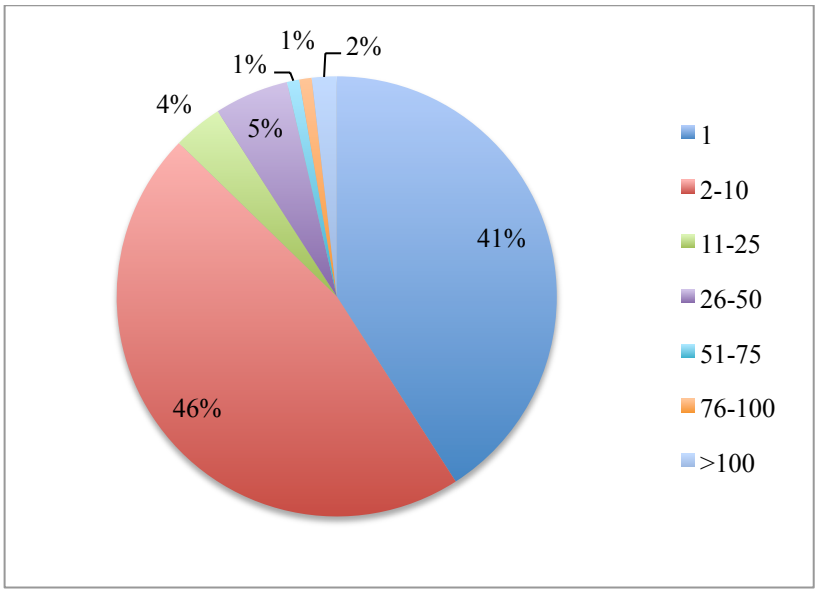


Figur 19 – Virksomhedsformer i Danmark år 2011, Data: Danmarks Statistik, 2016, GF5 (Egen tilvirkning)

Andelsforening udgør fx kun 1 % af virksomhedsformerne i Danmark i år 2011, hvorimod den udgør 26 % af virksomhedsformerne i stikprøven. Dette er en foruroligende skævvridning, der kan få betydning for generaliseringen af analysens resultater. Derudover udgør både *aktieselskab* og *anpartsselskab* væsentligt mere i stikprøven, hvilket også kan påvirke resultaterne af analysen. Den største udfordring ligger dog i, at virksomhedsformen *enkeltmandsfirma* slet ikke er repræsenteret i stikprøven. Dette skyldes formentligt, at enkeltmandsejede virksomheder ikke er forpligtet til at aflægge et årsregnskab (Christensen & Nielsen, 2012). Derfor vil de ikke indgå i stikprøven, da stikprøven er indsamlet via NN Erhverv ud fra deres tilgængelige regnskabsdata.

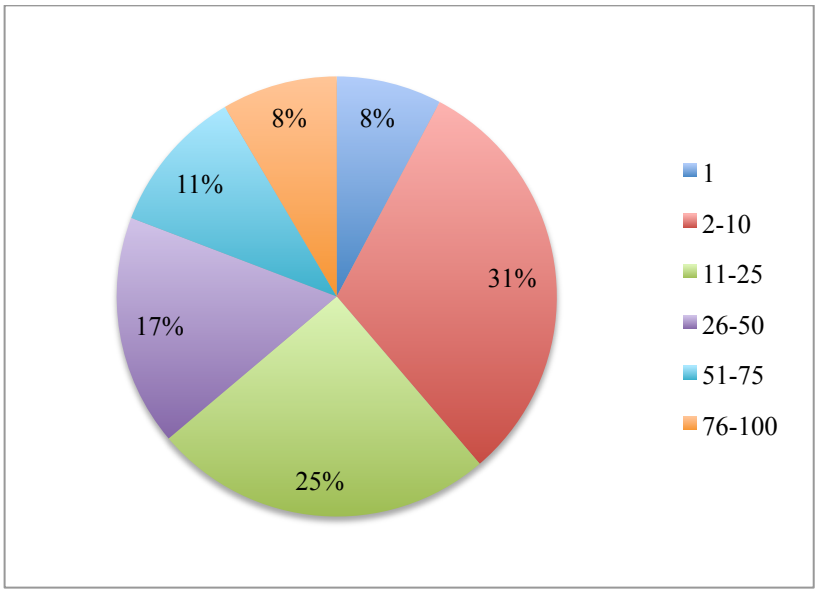
7.1.2.4 Medarbejderantal og etableringsår

Informationer vedrørende stikprøvens medarbejderantal og etableringsdato er også en del af arkene *Konkursdata, 2011* og *Normal drift data, 2011* i Bilag 3, og disse informationer analyseres i dette afsnit. Virksomhedernes medarbejderantal er inddelt i grupper, der skal afspejle den fordeling, der er i stikprøven. Figur 20 illustrerer, at 41 % af de konkursramte virksomheder kun har en medarbejder, hvorimod 46 % af virksomhederne har 2-10 medarbejdere. Det er således kun 4 % af de konkursramte virksomheder, der har et medarbejderantal, der er over 50, og derudover har kun 2 % af virksomhederne et medarbejderantal på over 100. Denne fordeling indikerer, at de konkursramte virksomheder generelt har få medarbejdere.



Figur 20 – Medarbejderantal i opløst efter konkurs, Data: Bilag 3 (Egen tilvirkning)

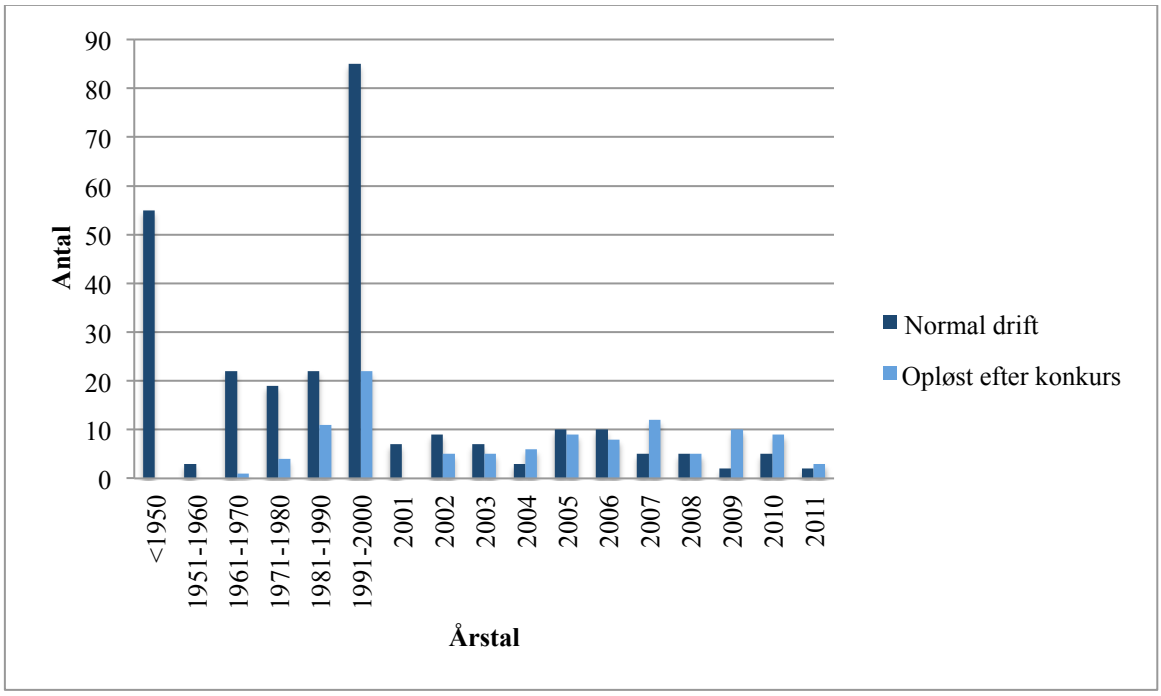
Figur 21 illustrerer medarbejderantal for virksomhederne i normal drift. Her udgør virksomheder med en medarbejder kun 8 %, hvorimod virksomheder med 2-10, 11-25 og 26-50 medarbejdere udgør hhv. 31 %, 25 % og 17 %. Andelen af virksomheder i normal drift med et medarbejderantal på over 50 er 19 %, hvilket er en del højere sammenlignet med de 4 % for de konkursramte virksomheder. Det fremgår i Figur 21, at der ikke er en gruppe, der repræsenterer et medarbejderantal på over 100, og det skyldes, at der under indsamlingen af dette datasæt er afgrænset fra at medtage virksomheder med et medarbejderantal på over 100 jf. afsnit 5.2.1 *Dataindsamling*. Under indsamlingen af datasættet for konkursramte virksomheder er der ikke afgrænset ved et medarbejderantal på 100, men idet at kun 2 % af virksomhederne overstiger et medarbejderantal på 100, understøtter det, hvorfor det er en fordel at begrænse medarbejderantallet for virksomheder i normal drift. Generelt illustrerer Figur 20 og Figur 21, at medarbejderantallet i normal drift virksomheder er højere end for de konkursramte virksomheder, idet 87 % af de konkursramte virksomheder har et medarbejderantal på maks. ti, hvorimod den andel kun er 39 % for virksomheder i normal drift.



Figur 21 – Medarbejderantal i normal drift, Data: Bilag 3 (Egen tilvirkning)

Virksomhedernes medarbejderantal for de to grupper i stikprøven er forskellig, men alligevel sikrer begrænsningen af medarbejderantallet på de 100 for virksomheder i normal drift, at denne variation er formindsket. Variationen i medarbejderantallet mellem de to grupper i stikprøven kan dog alligevel få konsekvenser for analysens resultater.

Virksomhedernes etableringsdato er, som tidligere nævnt, en del af Bilag 3, dog er der efterfølgende kun fokuseret på selve årstallet for at undgå en unødvendig høj detaljeringsgrad. Etableringsårene er derefter inddelt i forskellige perioder for at begrænse omfanget af Figur 22, der illustrerer etableringsårene for virksomhederne i stikprøven. Figur 22 er opdelt i de to grupper, således at det er muligt at se forskelle og overensstemmelser mellem dem. Der blev etableret flest virksomheder i årene 1991-2000, og det gælder både konkursramte virksomheder samt virksomheder i normal drift. Derudover blev en stor del af virksomhederne i normal drift etableret før år 1950, hvorimod ingen konkursramte virksomheder blev etableret før år 1961. Det vidner om, at de konkursramte virksomheder i stikprøven generelt er yngre end stikprøvens virksomheder, der er i normal drift. Det er dog værd at bemærke, at over 30 af de konkursramte virksomheder er etableret før år 2000 jf. Figur 22, hvilket understreger, at de konkursramte virksomheder ikke kun er nystartede virksomheder.



Figur 22 – Etableringsår i stikprøven, Data: Bilag 3 (Egen tilvirkning)

Figur 22 illustrerer, at stikprøvens virksomheder er etableret over en længere årrække, og derudover er der etableret forholdsvis lige mange virksomheder af hver gruppe fra år 2001 til år 2011. Det indikerer, at de to grupper er sammenlignelige, hvis der fokuseres på 00'erne på trods af, at virksomhederne i normal drift generelt er ældre end de konkursramte virksomheder. Forskellen mellem de to gruppers etableringsår kan dog få konsekvenser for analysens resultater.

7.1.2.5 Regnskabsposter

Figur 23 viser en deskriptiv statistik af de to gruppers regnskabsposter, og udregningerne bag statistikken fremgår af arket *Deskriptiv statistik* i Bilag 3. Regnskabsdataene der er udgangspunktet for Figur 23 er alle i 1.000 kr., og de er en del af arkene *Konkursdata, 2011* og *Normal drift data, 2011* i Bilag 3.

Regnskabsposter	N	Min.	Maks.	Median	Middelværdi	Standard-afvigelse
	Opløst efter konkurs					
EBITDA	110	-21273	15212	-15,00	-254,72	3295,25
Finans udgifter	110	1	18031	62,50	745,37	2514,02
Overført resultat	110	-77684	153627	-160,50	1564,80	19427,07
Aktivsum	110	74	602084	1475,00	15792,89	64937,35
Likvider	110	-148	55687	34,00	867,51	5546,76
Egenkapital	110	-27057	156568	38,00	3088,85	19594,70
Kortfristet gæld	110	45	162226	1252,00	7531,87	22607,13
	Normal drift					
EBITDA	271	-14423	790000	1393,00	8459,64	51121,54
Finans udgifter	271	1	287000	515,00	2977,65	17952,92
Overført resultat	271	-77480	12387000	6972,00	84143,68	760316,87
Aktivsum	271	456	13824000	30123,00	171255,35	907918,63
Likvider	271	1	150698	320,00	3520,51	12603,46
Egenkapital	271	-3536	13497000	10476,00	115021,82	848856,96
Kortfristet gæld	271	175	1183260	6083,00	31808,71	91741,08

Figur 23 – Deskriptiv statistik af regnskabsposter, Data: Bilag 3 (Egen tilvirkning)

I Figur 23 er der kun decimaler på beregningerne median, middelværdi og standardafvigelse, idet regnskabsdataene i arkene *Konkursdata, 2011* og *Normal drift data, 2011* i Bilag 3 ikke indeholder decimaler. Det fremgår af Figur 23, at alle 381 virksomheder i stikprøven har en

værdi for hver af de syv regnskabsposter, hvilket skyldes, som tidligere nævnt, at virksomheder, der manglede en eller flere af regnskabsposterne, er blevet frasorteret.

Minimumværdierne i Figur 23 illustrerer, at flere af virksomhedernes regnskabsposter har negative værdier, uanset om de er i normal drift eller opløst efter konkurs. For gruppen opløst efter konkurs har både EBITDA, overført resultat, likvider og egenkapital en negativ minimumsværdi, og for gruppen i normal drift har EBITDA, overført resultat og egenkapital en negativ minimumsværdi. Derudover er regnskabsposten finans udgifter, som tidligere nævnt, konverteret til et positivt tal, hvorfor den oprindelig også havde en negativ værdi. Et negativt EBITDA forekommer, hvis virksomhedens omsætning ikke formår at dække de variable omkostninger, salgs- og distributionsomkostninger samt administrationsomkostninger. En virksomheds likvider fremstår negative, hvis virksomhedens samlede mængde af likvider er negative fx pga. overtræk på bankkonti. Regnskabsposten overført resultat er negativ, hvis årets resultat er negativt, og i de tilfælde overføres et underskud til virksomhedens egenkapital (Christensen & Nielsen, 2012). Er dette underskud større end virksomhedens egenkapital, resulterer det i en negativ egenkapital, hvilket er tilfældet for flere af virksomhederne i stikprøven. Det overrasker en del, at gruppen i normal drift har flere regnskabsposter, der er præget af negative værdier, hvorimod det er mindre overraskende, at konkursramte virksomheder har negative regnskabsposter.

Regnskabsposternes negative værdier kan give udfordringer ift. at beregne de udvalgte nøgletal til den efterfølgende statistiske analyse, det er også derfor, at finans udgifter er konverteret til et positivt tal. Formlerne for de udvalgte nøgletal, der præsenteres særskilt i afsnit 6.2.3 *Nøgletallenes betydning*, understreger, at de negative regnskabsposter i Figur 23 kan forblive negative. Det skyldes, at de regnskabsposter, der har negative værdier ikke indgår på samme tid i udregningen af et nøgletal. Det medfører, at på trods af de negative værdier af fx egenkapitalen opstår der ikke et falsk positivt resultat af de udregnede nøgletal, hvilket er af væsentlig betydning for anvendelsen af analysens resultater.

Maksimumværdierne i Figur 23 illustrerer, at der er en stor forskel mellem de to grupper af virksomheder, derudover understreger den, at ingen regnskabspost er præget af udelukkende negative værdier. Regnskabsposterne overført resultat, aktivsum og egenkapital for virksomheder i normal drift er præget af meget høje maksimumværdier, specielt da beløbene, som tidligere nævnt, er angivet i 1.000 kr.

Medianen i Figur 23 angiver det midtpunkt, der deler observationerne i to lige store dele, og for de konkursramte virksomheder indikerer det, at min. halvdelen af observationerne har negative værdier af regnskabsposterne EBITDA og overført resultat. Dette er derimod ikke tilfældet med virksomheder i normal drift, da alle medianer er positive, og derudover er de markant højere end for de konkursramte virksomheder, hvilket indikerer en væsentlig forskel mellem de to grupper. Medianen for de konkursramte virksomheder er udregnet på baggrund af de to midterste observationer, idet denne gruppe består af 110 virksomheder, der er et lige antal.

Ud fra middelværdien i Figur 23 er der ligeledes en markant forskel mellem de to grupper, idet middelværdien af regnskabsposterne for virksomhederne i normal drift er markant højere end for de konkursramte virksomheder. Derudover fremgår det, at middelværdien for EBITDA for gruppen opløst efter konkurs er den eneste, der har en negativ middelværdi. Anvendelse af en stikprøve frem for en population medfører, at outliers får en større betydning for middelværdien, hvilket gør det interessant at kigge på stikprøvens standardafvigelse (Clement & Ingemann, 2011). Standardafvigelsen er et udtryk for, hvor tæt observationerne ligger op ad hinanden, idet standardafvigelsen indikerer, hvor meget de enkelte målinger spreder sig fra middelværdien (Bendsen, 2016). Ud fra Figur 23 illustrerer de beregnede standardafvigelser, at der er en stor spredning mellem observationerne i hver gruppe, hvor især gruppen for virksomheder i normal drift har meget store standardafvigelser.

7.1.2.6 Nøgletal

Forud for den deskriptive statistik af nøgletallene er der i Figur 24 udarbejdet en oversigt over forventningerne til middelværdien af nøgletallene i de to grupper. Forventningerne til de enkelte nøgletal er foretaget ud fra den deskriptive statistik af regnskabsposterne, der blev udarbejdet i forrige afsnit. Det fremgår af Figur 24, at middelværdien for alle seks nøgletal forventes at være højere for virksomheder i normal drift frem for konkursramte virksomheder.

Nøgletal	Nøgletalskategori	Forventning til middelværdi
STD_EBV	Gearing	Normal drift > Opløst efter konkurs
C_TA	Likviditet	Normal drift > Opløst efter konkurs
EBITDA_TA	Rentabilitet	Normal drift > Opløst efter konkurs
RE_TA	Dækning	Normal drift > Opløst efter konkurs
EBITDA_IE	Aktivitet	Normal drift > Opløst efter konkurs
EBV_TA	Soliditet	Normal drift > Opløst efter konkurs

Figur 24 – Forventninger til nøgletal (Inspireret af Beaver, 1966, s. 81)

Den deskriptive statistik af de to gruppers nøgletal visualiseres i Figur 25, og udregningerne bag statistikken er en del af arket *Deskriptiv statistik* i Bilag 3, hvor de anvendte regnskabsdata findes i arkene *Konkursnøgletal, 2011* og *Normal drift nøgletal, 2011* i Bilag 3. I Figur 25 er de fem decimaler for nøgletallene bibeholdt til minimum- og maksimumværdierne, hvorimod der kun er anvendt to decimaler til beregningerne median, middelværdi og standardafvigelse. Ud fra Figur 25 indgår alle 381 virksomheder i stikprøven, hvilket stemmer overens med den deskriptive statistik af regnskabsposterne.

Nøgletal	N	Min.	Maks.	Median	Middel-værdi	Standard-afvigelse
	Opløst efter konkurs					
STD_EBV	110	-183,33333	181,88288	0,22	-0,11	30,96
C_TA	110	-0,09737	0,72059	0,03	0,09	0,15
EBITDA_TA	110	-12,77202	1,64865	-0,01	-0,16	1,29
RE_TA	110	-8,33161	0,59799	-0,20	-0,56	1,33
EBITDA_IE	110	-369,00000	227,00000	-0,23	-8,57	61,06
EBV_TA	110	-5,74093	0,93802	0,03	-0,28	1,06
	Normal drift					
STD_EBV	271	-46,53623	87,32000	0,67	1,59	7,32
C_TA	271	0,00001	0,67445	0,01	0,05	0,08
EBITDA_TA	271	-0,28183	0,49976	0,07	0,08	0,09
RE_TA	271	-0,72188	0,91073	0,32	0,32	0,27
EBITDA_IE	271	-26,74699	1954,00000	3,96	33,27	153,40
EBV_TA	271	-0,44890	0,99924	0,41	0,43	0,25

Figur 25 – Deskriptiv statistik af nøgletal, Data: Bilag 3 (Egen tilvirkning)

Minimumsværdierne i Figur 25 illustrerer, at alle nøgletal undtagen C_TA for gruppen i normal drift har negative værdier, hvilket skyldes de negative regnskabsposter, der blev analyseret i afsnit 7.1.2.5 *Regnskabsposter*. C_TA for virksomheder i normal drift har dog en meget lav positiv værdi. Derudover adskiller minimumsværdierne af STD_EBV og EBITDA_IE for de konkursramte virksomheder sig fra de resterende nøgletal, da de er meget negative.

Maksimumværdierne for alle nøgletallene er positive, hvilket understreger at ingen af nøgletallene udelukkende har negative værdier. STD_EBV og EBITDA_IE for de konkursramte virksomheder har dog ligeledes en markant højere maksimumværdi end de andre nøgletal for konkursramte virksomheder, dog overstiges disse værdier af EBITDA_IE for virksomheder i normal drift. STD_EBV for gruppen i normal drift er også en del højere end de fire resterende nøgletal for gruppen i normal drift på trods af, at den ikke overstiger værdien af EBITDA_IE for virksomheder i normal drift. Maksimumværdierne indikerer også, at nøgletallene C_TA, EBITDA_TA, RE_TA og EBV_TA ikke overstiger et med undtagelse af EBITDA_TA for konkursramte virksomheder.

De meget lave og høje værdier af enkelte nøgletal indikerer, at datasættet kan være præget af outliers. Der er dog i indeværende analyse ikke forsøgt at fjerne eller korrigere disse evt. outliers, da det er vurderet, at denne proces ligeledes kan resultere i bias, idet grænsen mellem en outlier og en acceptabel værdi er svær at bedømme. Derudover vil en frasortering af virksomheder med outliers begrænse stikprøven yderligere, hvilket ikke er attraktivt, idet stikprøven allerede er begrænset væsentligt jf. afsnit 7.1.1 *Databearbejdning*. Ud fra Figur 25 er det primært nøgletallene STD_EBV og EBITDA_IE for begge grupper, der virker påvirket af evt. outliers, hvorfor en frasortering af virksomheder ud fra disse nøgletal potentielt vil medføre en frasortering af brugbare virksomheder, hvis der fokuseres på de resterende nøgletal. Tilstedeværelsen af outliers kan dog have indflydelse på anvendelsen af den statistiske analyses resultater.

Den udregnede median indikerer, at over halvdelen af virksomhederne som er opløst efter konkurs har negative værdier af nøgletallene EBITDA_TA, RE_TA og EBITDA_IE, hvorimod de resterende nøgletal i Figur 25 har positive medianer. Flere af nøgletallene har dog medianer omkring nul, herunder C_TA, EBITDA_TA og EBV_TA for konkursramte virksomheder samt C_TA for virksomheder i normal drift.

Ud fra den beregnede middelværdi er der forskel på nøgletallene mellem de to grupper, da fem ud af seks nøgletal er negative for virksomheder, der er opløst efter konkurs, hvorimod alle middelværdierne for virksomheder i normal drift er positive. Derfor holder næsten alle forventningerne fra Figur 24 stik, idet kun C_TA er højere for konkursramte virksomheder ift. virksomheder i normal drift, og dette på trods af at middelværdierne for både likvider og aktivsummen er højere for virksomheder i normal drift jf. Figur 23. Standardafvigelsen

indikerer, at der er en forholdsvis stor spredning i nøgletallene STD_EBV og EBITDA_IE for begge grupper, hvor spredningen dog klart er størst i nøgletallet EBITDA_IE. Den begrænsede spredning i flere af nøgletallene skyldes formentlig, at flere af nøgletallene, som tidligere nævnt, ikke overstiger værdien 1, mens dette er ikke tilfældet for nøgletallene STD_EBV og EBITDA_IE, hvorfor der er en større spredning i disse nøgletal.

Ud fra afsnit 6.2.3 *Nøgletallenes betydning* er det muligt at uddybe betydningen af de enkelte nøgletal ud fra de udregnede middelværdier i Figur 25. Middelværdierne for STD_EBV illustrerer, at den kortfristet gæld i gennemsnit overstiger egenkapitalen for virksomheder i normal drift, hvorimod dette ikke er tilfældet for de konkursramte virksomheder. Dette indikerer, at virksomhederne i normal drift er finansieret via kortfristet gæld frem for egenkapital, hvorimod dette ikke er tilfældet med konkursramte virksomheder. Middelværdien af STD_EBV for konkursramte virksomheder er dog negativ, hvilket skyldes, at flere af virksomhederne har en negativ egenkapital jf. arket *Konkursdata, 2011* i Bilag 3. Derfor er en positiv fortolkning af nøgletallet misvisende, idet egenkapitalen er negativ selvom den kortfristede gæld ikke overstiger det konkrete beløb, hvorfor et negativt nøgletal ikke indikerer et positivt gearingsforhold for virksomheden. Middelværdierne for C_TA illustrerer, at de konkursramte virksomheder i gennemsnit har flere likvider end virksomhederne i normal drift, hvilket indikerer en bedre øjeblikkelig betalingsevne for virksomhederne, der er opløst efter konkurs. Dette resultat er dog overraskende, da forventningen jf. Figur 24 var, at de konkursramte virksomheder alt andet lige havde en dårligere øjeblikkelig betalingsevne end virksomhederne i normal drift.

Middelværdien af EBITDA_TA for virksomheder i normal drift indikerer, at den gennemsnitslige virksomhed i normal drift har formået at skabe et EBITDA-overskud ud fra virksomhedens aktivsum. Det har derimod ikke været tilfældet for de konkursramte virksomheder, idet middelværdien for EBITDA_TA i denne gruppe er negativ. Nøgletallet RE_TA afspejler virksomhedens evne til at akkumulere indtjening ud fra aktiverne, derfor indikerer middelværdierne af dette nøgletal, at virksomheder i normal drift gennemsnitslig formår dette, hvorimod det ikke er tilfældet for de konkursramte virksomheder, idet de konkursramte virksomheder har en negativ middelværdi af nøgletallet. Middelværdierne af nøgletallet EBITDA_IE indikerer, at der er et overskud af likviditet i gruppen for virksomheder i normal drift, hvorimod der er et likviditetsproblem for de konkursramte virksomheder, idet nøgletallet er negativt. Middelværdierne for nøgletallet EBV_TA, der også betegnes

soliditetsgraden, illustrerer, at den gennemsnitslige virksomhed i normal drift kan miste 43 % af kapitalen inden fremmedkapitalen påvirkes. Derimod angiver den negative middelværdi af nøgletallet for konkursramte virksomheder, at det vil være misvisende at analysere på, hvor meget kapital virksomhederne kan miste før fremmedkapitalen påvirkes, idet nøgletallets negative værdi understreger, at egenkapitalen allerede er tabt.

Overordnet indikerer nøgletallene, at der er en væsentlig forskel mellem de to grupper af virksomheder, hvilket understreger, at der et grundlag for at foretage en logistisk regression ud fra dette datasæt.

7.1.2.7 Afrunding af den deskriptive statistik

Den deskriptive statistik af stikprøven illustrerer, at udviklingen i stikprøvens konkurser ikke afspejler den generelle udvikling i antallet af konkurser i Danmark. Derudover er der en markant skævvridning i stikprøvens branchefordeling ift. branchefordelingen i Danmark, herunder især vægtningen af branchen *Engroshandel og detailhandel; reparation af motorkøretøjer og motorcykler*. Virksomhedsformerne i stikprøven indeholder ikke *enkeltmandsfirma*, og derudover er virksomhedsformen *andelsselskab (-forening) med begrænset ansvar* væsentlig overrepræsenteret, hvilket indikerer en skævvridning mellem stikprøven og virksomheds-formerne i Danmark. Disse karakteristika for den totale stikprøve indikerer, at den umiddelbart ikke er repræsentativ for populationen, hvilket kan få betydning for muligheden for at generalisere analysens resultater.

En opgørelse af medarbejderantallet for de to grupper indikerer, at variationen mellem grupperne er forsøgt begrænset, idet virksomhederne i normal drift ikke kan have et medarbejderantal på over 100. Medarbejderantallet for de to grupper varierer dog en del på trods af denne begrænsning. De to gruppers etableringsår indikerer, at virksomhederne i normal drift generelt er ældre end de konkursramte virksomheder på trods af, at etableringerne af de to typer af virksomheder er foretaget forholdsvis jævnt i 00'erne. Der eksisterer også en skævvridning mellem de to gruppers branchefordeling på trods af, at den er mere markant mellem den totale stikprøve og branchefordelingen i Danmark. Tilsvarende er gældende for de to gruppers virksomhedsformer, hvor der er en skævvridning, men den er mere markant for den totale stikprøve sammenlignet med virksomhedsformerne i Danmark. De to gruppers karakteristika understreger, at de ikke er matchet ud fra branchefordeling eller virksomhedsform, hvorimod

forskellen mellem de to grupper er mere begrænset ift. etableringsår samt medarbejderantal, og dette kan få betydning for den efterfølgende statistiske analyse.

Den deskriptive statistik af nøgletallene illustrerer, at der er en væsentlig forskel mellem de to gruppers nøgletal, men modsat de andre uoverensstemmelser kan dette være positivt, da det indikerer, at der er en mulighed for, at den logistiske regression kan skelne mellem de to typer af virksomheder, og dermed klassificere virksomhederne korrekt. Tilstedeværelsen af evt. outliers kan dog få betydning for resultaterne af den efterfølgende logistiske regression.

7.2 Logistisk regression

Den logistiske regressionsanalyse følger, som tidligere nævnt, sekstrinsmodellen, der først er præsenteret i afsnit 5.3 *Analysestrategi*, hvorefter den er uddybet og specificeret ift. den logistiske regression i afsnit 6.3.2 *Sekstrinsmodellen for logistisk regression*.

7.2.1 Trin 1 – Målsætningerne for den logistiske regression

Målsætningerne for den logistiske regressionsanalyse tager udgangspunkt i begge forskningsmål, der er præsenteret i afsnit 6.3.2.1 *Trin 1 – Målsætninger for logistisk regression*. Det skyldes, at analysen skal identificere de uafhængige variabler, der har indflydelse på gruppemedlemskabet af den afhængige variabel, dvs. hvorvidt en virksomhed karakteriseres som i normal drift eller opløst efter konkurs. Derudover er det en målsætning, at analysen skal etablere et grundlag for at kunne klassificere enhver anden virksomhed, udover dem der indgår i analysen, som værende i normal drift eller opløst efter konkurs.

Analysens forskningsmål kan karakteriseres som et afhængighedsforhold, hvor virksomhedens status er det afhængige koncept, hvorimod nøgletallene er de uafhængige koncepter. Karakteristikaene for de enkelte koncepter uddybes i det efterfølgende afsnit. Dette afhængighedsforhold understreger anvendelsen af den logistiske regression som den multivariate teknik, og derudover harmonerer denne teknik med de måletekniske egenskaber, som de afhængige og uafhængige variabler besidder. Disse egenskaber omhandler bl.a. at den afhængige variabel er binær, og at stikprøverne er identificerbare. De to grupper i stikprøven er jf. afsnit 7.1.1 *Databearbejdning* ikke lige store, men dette forhindrer ikke anvendelsen af den logistiske regression. En mere generel argumentation for at anvende logistisk regression er præsenteret i afsnit 6.4 *Operationalisering af teorien*.

7.2.2 Trin 2 – Undersøgelsesdesignet for den logistiske regression

I Trin 2 fokuseres der på tre områder, og disse er udvælgelse af afhængig og uafhængige variabler, vurdering af den tilstrækkelige stikprøvestørrelse for den planlagte analyse samt opdeling af stikprøven ud fra et valideringsformål.

7.2.2.1 Udvælgelse af afhængig og uafhængige variabler

Den afhængige variabel i en logistisk regression karakteriseres som binær og ikke-tal, og de to grupper repræsenteres med værdierne 0 og 1. Umiddelbart har det, jf. afsnit 6.3.2.2.1 *Den binære afhængige variabel*, ingen betydning, hvilken gruppe, der tildeles værdierne 0 og 1, men det er vigtigt, at denne tildeling er bekendt for forskeren, da det har indflydelse på fortolkningen af koefficienterne. De to grupper, der har interesse for analysen, er virksomhederne i normal drift samt virksomhederne, der er opløst efter konkurs. Specialets fokus er at undersøge sandsynligheden for konkurs, og derfor tildeles denne gruppe af virksomheder værdien 1. Dette resulterer i, at virksomheder i normal drift tildeles værdien 0. Tildelingen medfører, at koefficienterne i analysen vil repræsentere indflydelsen på sandsynligheden for, at en virksomhed går konkurs. Det betyder således, at den logistiske regression analyserer sandsynligheden for, at en virksomhed er i den gruppe, der er tildelt værdien 1. Tildelingen af værdierne kan opstilles mere formelt på denne måde.

$$Y_1 = \begin{cases} 0 & Normal\ drift \\ 1 & Opløst\ efter\ konkurs \end{cases}$$

Denne tildeling understreges med udtrækket fra SPSS i Figur 26.

Original Value	Internal Value
Normal drift	0
Opløst efter konkurs	1

Figur 26 – Kodning af den afhængige variabel (Egen konstruktion i SPSS)

De uafhængige variabler i analysen udgøres af de seks præsenterede nøgletal i afsnit 6.2.1 *Udvælgelse af nøgletallene*, og kalkulationen af disse nøgletal er gennemgået i afsnit 7.1.1 *Databearbejdning*. De seks uafhængige variabler kan opstilles på følgende måde:

$$X_1 = STD/EBV$$
$$X_2 = C/TA$$
$$X_3 = EBITDA/TA$$
$$X_4 = RE/TA$$
$$X_5 = EBITDA/IE$$
$$X_6 = EBV/TA$$

Begrundelsen for at benytte disse seks nøgletal er præsenteret i afsnit 6.4 *Operationalisering af teorien*, hvor der er lagt vægt på, at disse nøgletal afspejler de vigtigste aspekter af en virksomheds finansielle profil. X_1 til X_5 er en del af Altman & Sabatos (2007) undersøgelse, hvorimod X_6 er soliditetsgraden, der er inddraget som et supplement til de fem andre nøgletal. De seks udvalgte nøgletal er uddybet særskilt i afsnit 6.2.3 *Nøgletallenes betydning*.

7.2.2.2 Stikprøvestørrelsen

Den totale stikprøvestørrelse er jf. afsnit 7.1.1 *Databearbejdning* på 381 virksomheder, hvoraf 271 virksomheder er i normal drift, og de resterende 110 virksomheder er opløst efter konkurs. En total stikprøvestørrelse på 381 observationer er vurderet tilstrækkelig, idet Hair et al. (2010) benytter sig af en total stikprøvestørrelse på kun 100 observationer til at foretage en logistisk regression. Den totale stikprøve til indeværende speciale overstiger dette antal, hvorfor den totale stikprøvestørrelse vurderes at være acceptabel. Den totale stikprøve bør dog, jf. afsnit 6.3.2.2.2.1 *Den totale stikprøvestørrelse*, opledes i to, således at den ene vedrører den faktiske analyse, hvorimod den anden, der betegnes hold-out stikprøven, skal anvendes til validering af den logistiske regressionsmodel. Det gælder dog de samme betingelser for både den totale stikprøve og hold-out stikprøven, at der for hver estimeret parameter bør være min. ti observationer for hver gruppe jf. afsnit 6.3.2.2.2.2 *Stikprøvestørrelse for hver kategori af den afhængige variabel*. Dette 10:1 forhold af observationer ift. antallet af uafhængige variabler er opfyldt for begge grupper i den totale stikprøve, da der for hver af de seks uafhængige variabler er over 45 observationer $\left(\frac{271}{6} = 45{,}2\right)$ for virksomheder i normal drift og over 18 observationer $\left(\frac{110}{6} = 18{,}3\right)$ for konkursramte virksomheder. Opdelingen af den totale stikprøve har resulteret i en stikprøve til analysen på 243 observationer og en hold-out stikprøve på 138 observationer, og denne opdeling kan ses i Figur 27. Baggrunden for at foretage netop denne opdeling bearbejdes yderligere i afsnit 7.2.2.3 *Opdeling af stikprøven*.

Stikprøve	Normal drift	Opløst efter konkurs	Antal i alt
Totale stikprøve	271	110	381
Analysens stikprøve	173	70	243
Hold-out stikprøve	98	40	138

Figur 27 – Opdeling af den totale stikprøve (Egen tilvirkning)

Opdelingen indebærer, at stikprøven til analysen består af 173 observationer for virksomheder i normal drift og 70 observationer for konkursramte virksomheder, og denne stikprøve imødekommer således stadig kravet om et 10:1 forhold for begge grupper. Hold-out stikprøven består af 98 observationer for virksomheder i normal drift og 40 observationer for virksomheder, der er opløst efter konkurs. Dette medfører, at hold-out stikprøven kun overholder 10:1 forholdet for virksomheder i normal drift, da der kun er lidt over seks observationer $\left(\frac{40}{6} = 6{,}7\right)$for hver af de seks uafhængige variabler for konkursramte virksomheder. Denne opdeling er dog alligevel foretaget, da det jf. Hair et al. (2010) er vigtigere at kunne validere resultaterne frem for at undgå en opdeling, der ikke opfylder kravet for kun en af grupperne. Derudover overholder hold-out stikprøven i Hair et al. (2010) og Altman & Sabato (2007) heller ikke kravet om et 10:1 forhold for begge grupper, hvilket understreger, at det er forsvarligt at foretage denne opdeling af den totale stikprøve.

7.2.2.3 Opdeling af stikprøven

En opdeling af stikprøven kan, som tidligere nævnt, anvendes til at validere den logistiske regression. Hver gang der foretages en opdeling af stikprøven, er det dog vigtigt at sikre, at de endelige stikprøvestørrelser er tilstrækkelige til at understøtte antallet af uafhængige variabler, der inkluderes i analysen (Hair et al., 2010). Jf. afsnit 7.2.2.2 *Stikprøvestørrelsen* medfører opdelingen af stikprøven, at den ene af grupperne i hold-out stikprøven ikke er tilstrækkelig. Der er dog tidligere argumenteret for, at en opdeling af den totale stikprøve i dette tilfælde er det mest hensigtsmæssige for indeværende speciale.

I udvælgelsen af hold-out stikprøven er det ligeledes vigtigt at sikre tilfældighed, således at enhver ordning af observationerne ikke påvirker estimeringsprocessen og valideringen (Hair et al., 2010). Til udvælgelsen af hold-out stikprøven er der anvendt en tilfældig tal generator fra

Intemodino Group, da den gjorde det muligt at definere et bestemt område samt antallet af tilfældige tal, der skulle genereres (Intemodino Group, 2016). Den totale stikprøve består, som tidligere nævnt, af 271 virksomheder i normal drift og 110 konkursramte virksomheder, hvilket indebærer, at de konkursramte virksomheder udgør 29 % af den totale stikprøve jf. Figur 28.

Status	Antal	Pct.
Opløst efter konkurs	110	29 %
Normal drift	271	71 %
I alt	381	100 %

Figur 28 – Gruppernes forhold i den totale stikprøve (Egen tilvirkning)

Forholdet mellem de to grupper i den totale stikprøve bevares således ift. hold-out stikprøven, da det giver et bedre grundlag for at kunne sammenligne resultaterne af hold-out stikprøven med resultaterne af den resterende stikprøve i analysen. I den totale stikprøve indgår der 110 konkursramte virksomheder, hvilket indebærer, at det ikke vil være muligt at opdele de konkursramte virksomheder i en hold-out stikprøve og en stikprøve til analysen uden at en af dem ikke overholder 10:1 forholdet, idet der er seks uafhængige variabler. Der skulle have været min. 120 konkursramte virksomheder i den totale stikprøve, hvis det skulle være muligt at overholde 10:1 forholdet ved en opdeling. Hair et al. (2010) opdelte deres totale stikprøve på 100 observationer til en hold-out stikprøve på 40 observationer og en stikprøve til analysen på 60 observationer. Derudfra er det vurderet passende, at hold-out stikprøven for indeværende analyse indeholder 40 konkursramte virksomheder, hvilket efterlader 70 konkursramte virksomheder i stikprøven til analysen. Grunden til at den ikke opdeles mere lige fx 55/55 eller 50/60 skyldes, at der i både Hair er al. (2010) samt Altman & Sabato (2007) fokuseres på, at hold-out stikprøven kun bør udgøre en lille stikprøve af den totale stikprøve. Derudover vil en opdeling på 55 i hold-out stikprøven og 55 i den resterende stikprøve medføre, at ingen at de to stikprøver overholder 10:1 forholdet for de konkursramte virksomheder. Med udgangspunkt i de 40 konkursramte virksomheder i hold-out stikprøven er der beregnet det antal af virksomheder i normal drift, der skal indgå i hold-out stikprøven, før den opnår det samme forhold mellem de to grupper, som der er tilfældet i den totale stikprøve jf. Figur 28. Denne kalkulation har resulteret i, at hold-out stikprøven skal indeholde 98 virksomheder i normal drift for at sikre et identisk forhold mellem grupperne i hold-out stikprøven og den totale stikprøve. Dette medfører, at den resterende stikprøve til analysen ligeledes opnår samme forhold mellem grupperne, som der er

tilfældet i både hold-out stikprøven og den totale stikprøve. Efter fastlæggelsen af antallet af de to grupper i hold-out stikprøven er der indtastet de informationer, der fremgår af Figur 29, i den tilfældige tal-generator.

Antal	Tal mellem	
40	2	111
98	2	272

Figur 29 – Tilfældig tal-generator (Egen tilvirkning)

Afgrænsningen af områderne fra 2 til 111 og fra 2 til 272 skyldes, at det svarer til de rækkenumre, som observationerne har i Bilag 3 i arkene *Konkursnøgletal, 2011* og *Normal drift nøgletal, 2011*. Resultatet af indtastningerne fra Figur 29 kan ses i Figur 30, hvor de udtrukne tilfældige tal fremgår. Disse tilfældige tal er derefter markeret med en blå farve i arkene *Konkursnøgletal, 2011* og *Normal drift nøgletal, 2011* i Bilag 3. Det er ud fra markeringerne i disse ark, at den totale stikprøve er opdelt i arkene *Stikprøve, SPSS* og *Hold-out*, der er udgangspunktet for analysen. Efter opdelingen af den totale stikprøve i arkene *Stikprøve, SPSS* og *Hold-out* er den binær afhængige variabel efterfølgende tilføjet til disse ark, hvor de konkursramte virksomheder tildeles værdien 1, hvorimod virksomhederne i normal drift tildeles værdien 0, jf. afsnit 7.2.2.1 *Udvælgelse af afhængig og uafhængige variabler*.

Status	De tilfældige tal
Opløst efter konkurs	106 4 31 49 20 76 59 58 79 8 75 26 38 77 21 3 46 53 103 15 35 71 24 47 40 110 67 48 108 52 56 97 94 66 55 81 62 25 43 27
Normal drift	158 19 17 87 238 78 145 147 261 229 237 258 41 39 33 270 143 187 91 262 108 6 179 4 54 138 141 170 230 134 183 113 222 193 75 204 42 256 191 64 223 15 178 44 188 148 161 57 257 43 250 46 142 228 95 192 206 244 111 241 5 198 159 122 125 253 14 124 197 127 123 154 213 175 37 51 79 72 224 71 169 215 164 47 90 50 266 3 84 83 103 106 88 105 210 80 272 128

Figur 30 – De tilfældige tal (Egen tilvirkning)

Figur 31 visualiserer, at analysen bygger på en stikprøve, der består af 243 observationer, hvilket stemmer overens med den tidligere omtalte stikprøve, der er tilegnet analysen. Derudover illustrerer Figur 31, at der ikke er nogen manglende cases, dvs. ufuldendte observationer, hvorfor alle 243 er en del af den logistiske regressionsanalyse.

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	243	100,0
	Missing Cases	0	,0
	Total	243	100,0
Unselected Cases		0	,0
Total		243	100,0

Figur 31 – Oversigt over stikprøven (Egen konstruktion i SPSS)

7.2.3 Trin 3 – Forudsætningerne for den logistiske regression

De manglende forudsætninger for logistisk regression, der blev beskrevet i afsnit 6.3.2.3 *Trin 3 – Forudsætninger for logistisk regression*, medfører, at der ikke er forudsætninger, der skal tages højde for inden estimeringsprocessen påbegyndes. Det er, som tidligere nævnt, en stor fordel for logistisk regression, da det ikke sætter begrænsninger for analysen.

7.2.4 Trin 4 – Estimering af den logistiske regressionsmodel og vurdering af modellens overordnede fit

Estimeringen af den logistiske regressionsmodel tager udgangspunkt i nedenstående modelformulering, der blev præsenteret i afsnit 6.3.2.4.1.2 *Modelestimering*.

$$Logit_i = \ln\left(\frac{prob_{event}}{1 - prob_{event}}\right) = b_0 + b_1X_1 + \cdots + b_nX_n$$

Til estimeringsprocessen af de logistiske koefficienter er der opstillet følgende hypoteser, der hver er tilknyttet et af de seks nøgletal, der blev præsenteret som de uafhængige variabler i afsnit 7.2.2.1 *Udvælgelse af afhængig og uafhængige variabler*.

$$H_0: b_1 = 0; b_2 = 0; b_3 = 0; b_4 = 0; b_5 = 0; b_6 = 0 \quad ingen\ logit - lineær\ sammenhæng$$

$$H_1: b_1 \neq 0; b_2 \neq 0; b_3 \neq 0; b_4 \neq 0; b_5 \neq 0; b_6 \neq 0 \quad logit - lineær\ sammenhæng$$

H_0-hypoteserne illustrerer, at der ikke er en logit-lineær sammenhæng mellem hver af koefficienterne og sandsynligheden for, at en virksomhed går konkurs. H_1-hypoteserne illustrerer derimod, at der er en logit-lineær sammenhæng mellem hver af koefficienterne og sandsynligheden for, at en virksomhed går konkurs. Der er opstillet separate hypoteser til hver af de seks uafhængige variabler, da det er vurderet, at det efterfølgende arbejde med hver af koefficienterne bliver mere overskueligt, hvis hypoteserne er opdelt fra starten af. Det er ofte set, at notationen i hypoteserne er foretaget med et beta, men da Hair et al. (2010) benytter et *b* i modelformuleringen, er der holdt fast i denne notation til hypoteserne.

7.2.4.1 Modelestimering

I afsnit 6.4 *Operationalisering af teorien* fremgår det, at den valgte metode til modelestimeringen er forward stepwise (Wald), og derudover blev der specificeret, hvordan grænserne for at tilføje og fjerne en variabel skulle defineres. De fastsatte grænser til indeværende analyse er 0,05 for at tilføje en variabel og 0,10 for at fjerne en variabel, da disse værdier er de mest almindelige signifikansniveauer for en stepwise metode, og det er derfor også de værdier, som SPSS selv foreslår. Et signifikansniveau på 0,05 for at tilføje en variabel og et signifikansniveau på 0,10 for at fjerne en variabel overholder således de kriterier, der er fremsat til grænserne i afsnit 6.4 *Operationalisering af teorien*. Der anvendes proceduren MLE for at finde de mest egnede koefficienter i den logistiske regression, og denne procedure er uddybet i afsnit 6.3.2.4.1.2 *Modelestimering*.

7.2.4.1.1 Basismodellen

Indledningsvis producerer SPSS en klassifikationsmatrix af basismodellen, der kan ses i Figur 32. Der er i denne model ingen variabler med, idet der kun er implementeret en konstant. Til udarbejdelsen af denne klassifikationsmatrix er der anvendt en skæringsværdi på 0,500.

Observed			Predicted		
			Status		
			Normal drift	Opløst efter konkurs	Percentage Correct
Step 0	Status	Normal drift	173	0	100,0
		Opløst efter konkurs	70	0	,0
	Overall Percentage				71,2

Figur 32 – Klassifikationsmatrix (Egen konstruktion i SPSS)

Det fremgår af Figur 32, at klassifikationsnøjagtigheden er 71,2 %, og denne nøjagtighed opnås, hvis alle virksomheder forudsiges til at være i normal drift. En klassifikationsnøjagtighed på 71,2 % er således resultatet af en logistisk model, der ikke formår at forudsige en eneste af de konkursramte virksomheder til at være opløst efter konkurs.

Figur 33 illustrerer, at basismodellen kun indeholder en konstant, men alligevel giver den lidt interessant information. Den tester en hypotese (H_1-hypotese) om, at hyppigheden af de 173 virksomheder i normal drift og hyppigheden af de 70 konkursramte virksomheder er statistisk signifikant forskellige fra hinanden. Den modsvarende hypotese (H_0-hypotese) er, at hyppigheden af de to grupper er ens. Hypotesen testes af Wald teststørrelsen i Figur 33, hvor H_0-hypotesen afvises, da p-værdien er 0. Det medfører, at der afvises, at der er et lige antal af virksomheder i normal drift og virksomheder opløst efter konkurs. Dette bekræfter således de to gruppers fordeling i den stikprøve, der er udgangspunktet for analysen, og som er præsenteret i afsnit 7.2.2.2 *Stikprøvestørrelsen*. Basismodellen bidrager således ikke med noget nyt, idet den blot bekræfter udgangspunktet for analysen, hvorefter der er baggrund for at påbegynde den reelle modelestimering.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-,905	,142	40,798	1	,000	,405

Figur 33 – Variabler i ligningen (Egen konstruktion i SPSS)

Alle de uafhængige variabler, der ikke er med i basismodellen, fremgår af Figur 34. For hver af variablerne er der udregnet en score statistik, der er et mål for association, der er brugt i logistisk regression, og det anvendes til udvælgelsen af variablerne i stepwise proceduren (Hair et al., 2010).

			Score	df	Sig.
Step 0	Variables	STD_EBV	3,224	1	,073
		C_TA	13,610	1	,000
		EBITDA_TA	8,178	1	,004
		RE_TA	64,380	1	,000
		EBITDA_IE	3,738	1	,053
		EBV_TA	62,823	1	,000
	Overall Statistics		79,552	6	,000

Figur 34 – Variabler uden for ligningen (Egen konstruktion i SPSS)

Ud fra score statistikkerne i Figur 34 kan det ses, at fire ud af de seks uafhængige variabler er statistisk signifikante, da de alle har en p-værdi, der er under signifikansniveauet på 0,05. Det er kun STD_EBV (X_1) og EBITDA_IE (X_5), der ikke er signifikante, da de har p-værdier, der er over signifikantniveauet på hhv. 0,073 og 0,053. Resultatet af Figur 34 indikerer således, at det formentligt er muligt at forudsige noget i indeværende analyse. Implementeringsrækkefølgen af variablerne afgøres af den højeste score statistik i stepwise proceduren, derfor er RE_TA (X_4) med en score statistik på 64,380 den første variabel, der skal implementeres i modellen.

7.2.4.1.2 Stepwise modelestimering

Figur 35 indeholder de samme elementer som Figur 33, og den illustrerer, at indeværende stepwise modelestimering består af to step, hvoraf RE_TA (X_4) tilføjes i step 1, og C_TA (X_2) tilføjes i step 2. Implementeringen af RE_TA (X_4) i step 1 stemmer overens med den vurdering, der blev foretaget ud fra Figur 34 i forrige afsnit. Betydningen af elementernes indhold i Figur 35 uddybes i afsnit 7.2.5 *Trin 5 – Fortolkning af resultaterne*.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	RE_TA	-5,587	,806	48,019	1	,000	,004
	Constant	-,529	,202	6,881	1	,009	,589
Step 2[b]	C_TA	6,389	1,693	14,244	1	,000	595,011
	RE_TA	-6,195	,909	46,407	1	,000	,002
	Constant	-,851	,230	13,657	1	,000	,427

a. Variable(s) entered on step 1: RE_TA.
b. Variable(s) entered on step 2: C_TA.

Figur 35 – Variabler i ligningen (Egen konstruktion i SPSS)

De variabler, der ikke er en del af modellen i step 1 eller step 2, fremgår af Figur 36. I step 1 i Figur 36 adskiller C_TA (X_2) sig fra de andre variabler, da det er den eneste variabel, hvor score statistikken er signifikant med en p-værdi på 0,000. Dette indikerer, at en implementering af denne variabel kan forbedre modellens overordnede fit. Derudover besidder C_TA (X_2) den højeste score statistik blandt de fem variabler, der ikke er implementeret i modellens step 1. Disse faktorer har været afgørende for, at C_TA (X_2) er tilføjet til modellen i step 2 i Figur 35.

			Score	df	Sig.
Step 1	Variables	STD_EBV	,563	1	,453
		C_TA	17,074	1	,000
		EBITDA_TA	,389	1	,533
		EBITDA_IE	1,982	1	,159
		EBV_TA	,000	1	,999
	Overall Statistics		22,808	5	,000
Step 2	Variables	STD_EBV	,588	1	,443
		EBITDA_TA	,031	1	,860
		EBITDA_IE	1,886	1	,170
		EBV_TA	,584	1	,445
	Overall Statistics		3,648	4	,456

Figur 36 – Variabler uden for ligningen (Egen konstruktion i SPSS)

De fire tilbageværende variabler i step 2 i Figur 36 er ikke signifikante, og de har alle forholdsvis lave score statistikker. Derfor implementeres der ikke yderligere variabler i modellen, selvom der i Figur 34 ifm. basismodellen er yderligere to variabler, der er signifikante. Disse variabler er EBITDA_TA (X_3) og EBV_TA (X_6), men efter at RE_TA (X_4) og C_TA (X_2) er tilføjet til modellen, er de ikke længere signifikante, hvorfor de ikke implementeres. Den endelige model er således den logistiske regressionsmodel, der inkluderer variablerne C_TA (X_2) og RE_TA (X_4). Det medfører, at H_0-hypotesen afvises for b_2 og b_4, da p-værdien er under signifikansniveauet, og dermed kan de ikke fjernes fra ligningen. Derimod bekræftes H_0-hypotesen for b_1, b_3, b_5 og b_6. En afvisning af H_0-hypotesen betyder, at H_1-hypotesen accepteres for b_2 og b_4, hvorimod den afvises for b_1, b_3, b_5 og b_6. Det resulterer i, at der er en logit lineær sammenhæng mellem RE_TA (X_4) og sandsynligheden for, at en virksomhed går konkurs, samt mellem C_TA (X_2) og sandsynligheden for, at en virksomhed går konkurs. Derimod er der ingen logit lineær sammenhæng mellem de resterende koefficienter og sandsynligheden for, at en virksomhed går konkurs. Modellen kan opstilles som en formel ud fra den modelformulering, der blev præsenteret i afsnit 7.2.4 *Trin 4 – Estimering af den logistiske regressionsmodel og vurdering af modellens overordnede fit*. Ud fra indholdet i step 2 i Figur 35 ser formlen således ud:

$$Logit = -0{,}851 + 6{,}389X_2 - 6{,}195X_4$$

Den endelige model er således også udgangspunktet for vurderingen af modellens overordnede fit og fortolkningen af koefficienterne, der foretages i de kommende afsnit.

7.2.4.2 Vurdering af modellens overordnede fit

Der anvendes tre forskellige metoder til at vurdere modellens overordnede *goodness-of-fit*, og disse er statistiske mål, pseudo R^2 mål og klassifikationsnøjagtighed. Alle tre metoder er præsenteret i afsnit 6.3.2.4.2 *Vurdering af goodness-of-fit for den estimerede model*. Der foretages dog ikke en decideret opdeling mellem analysen af de tre metoder, da udtrækkene fra SPSS indeholder elementer, der indgår i flere af metoderne.

Et af de statistiske mål er en chi-i-anden test, der undersøger ændringer i *-2LL* værdien, hvor den tager udgangspunkt i *-2LL* værdien af basismodellen, og sammenligner den med *-2LL* værdien af modellerne i step 1 og step 2, hvor der er tilføjet variabler. Resultatet af chi-i-anden testen ses i Figur 37, og *-2LL* værdierne, der er udgangspunktet for testen, er i Figur 38. Chi-i-anden testen anvendes til at undersøge, hvorvidt den stepwise model med de tilføjede variabler er en forbedring af basismodellen.

		Chi-square	df	Sig.
Step 1	Step	123,158	1	,000
	Block	123,158	1	,000
	Model	123,158	1	,000
Step 2	Step	14,359	1	,000
	Block	137,517	2	,000
	Model	137,517	2	,000

Figur 37 – Omnibus test af modelkoefficienter (Egen konstruktion i SPSS)

Figur 37 er opdelt i step 1 og step 2, og for hvert step er der tre niveauer, der betegnes *step, block* og *model*. I indeværende analyse er det kun *step* og *model*, der er relevante, idet *step* sammenligner *-2LL* værdien af den nyeste model med den forrige version af modellen for at vurdere, hvorvidt hver ny variabel skaber en forbedring, og *model* sammenligner altid den nye model med basismodellen (ReStore, 2011). Det medfører, at *block* udelades af indeværende analyse, da den vedrører en hierarkisk tilgang (ReStore, 2011). I step 1 i Figur 37 er der ikke forskel på *step* og *model*, og chi-i-anden testen illustrerer, at der er en signifikant forskel mellem *-2LL* værdien af basismodellen og modellen i step 1, hvilket indikerer, at modellen i step 1 er en forbedring af basismodellen. I step 2 i Figur 37 er der en forskel mellem *step* og *model*, hvor *step* illustrerer den signifikante reducering på 14,359, der er i *-2LL* værdien fra step 1 til step 2 i Figur 38. Derudover er værdien af *step* i step 2 også forskellen mellem værdien af *model* i step 1 og step 2. Chi-i-anden testen for både *step* og *model* i step 2 er signifikante, hvilket indikerer, at

modellen i step 2 er en forbedring af modellen i step 1 og basismodellen. Ud fra afsnit 6.3.2.4.2.1 *Statistiske mål* skal man dog være påpasselig med kun at drage konklusioner ud fra en chi-i-anden test, da den er følsom over for stikprøvestørrelsen. Det er således også årsagen til, at modellens fit vurderes ud fra flere forskellige metoder.

Reduceringen af *-2LL* værdien mellem step 1 og step 2 fra 168,642 til 154,283 fremgår af Figur 38, og da en *-2LL* værdi på 0 er tilsvarende et perfekt model fit, jf. afsnit 6.3.2.4.2.1 *Statistiske mål*, er denne reducering en forbedring af modellen. Værdien er dog stadig høj, hvilket vidner om, at modellen langt fra besidder et perfekt fit ud fra *-2LL* værdien.

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	168,642[a]	,398	,569
2	154,283[a]	,432	,618

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

Figur 38 – Modeloversigt (Egen konstruktion i SPSS)

Udover *-2LL* værdien indeholder Figur 38 også pseudo R^2 værdier, der er en anden metode til at vurdere modellens overordnede fit, da de, jf. afsnit 6.3.2.4.2.2 *Pseudo R^2 mål*, reflekterer den variationsmængde, der forklares med den logistiske regressionsmodel. Tilføjelsen af RE_TA (X_4) i step 1 giver et model fit med pseudo R^2 værdier rangerende fra 0,398 (Cox & Snell R^2) til 0,569 (Nagelkerke R^2). Det udtrykker, at modellen i step 1 forklarer ca. 57 % af variationsmængden i den afhængige variabel vha. den uafhængige variabel RE_TA (X_4), hvis der benyttes pseudo R^2 værdien fra Nagelkerke. Det betyder i praksis, at nøgletallet RE_TA i step 1 forklarer 57 % af den afhængige variabel i den logistiske regression, der afspejler sandsynligheden for, at en virksomhed går konkurs. Ofte anvendes Cox & Snell R^2 værdien ikke, da dens skala ikke er mellem 0 og 1, jf. afsnit 6.3.2.4.2.2 *Pseudo R^2 mål*, hvilket gør det svært at fastsætte forklaringsprocenten af variationsmængden i den afhængige variabel. Implementeringen af C_TA (X_2) i step 2 medfører en stigning i pseudo R^2 værdierne, idet Cox & Snell R^2 og Nagelkerke R^2 nu udgør hhv. 0,432 og 0,618. Stigningen i pseudo R^2 værdierne fra step 1 til step 2 understreger, at modellens forklaringsevne er forbedret. Hvis der fokuseres på Nagelkerke R^2 resulterer det i, at ca. 62 % af variationsmængden i den afhængige variabel kan forklares vha. de uafhængige variabler C_TA (X_2) og RE_TA (X_4). Det betyder i praksis, at nøgletallene C_TA og RE_TA i step 2 forklarer 62 % af den afhængige variabel i den logistiske

regression, der afspejler sandsynligheden for, at en virksomhed går konkurs. Den generelle variation mellem de to versioner af pseudo R^2 værdierne understreger, hvorfor man bør være påpasselig med at anvende dem, derudover er det vigtigt at understrege, at de er approksimationer. Logit R^2 værdien, der er præsenteret i afsnit 6.3.2.4.2.2 *Pseudo R^2 mål*, er ikke en del af udtrækket fra SPSS, derudover har det ikke været muligt at beregne den, da basismodellens *-2LL* værdi ligeledes ikke er en del af dataudtrækket fra SPSS.

Et andet mål, der er en del af metoden med de statistiske mål, er Hosmer & Lemeshow testen, der uddybes i afsnit 6.3.2.4.2.1 *Statistiske mål*, og denne test kan ligeledes anvendes til at vurdere modellens overordnede fit. Hosmer & Lemeshow testen måler overensstemmelsen mellem de faktiske og forudsagte værdier af den afhængige variabel. Ud fra denne test er et godt model fit indikeret ved, at der er en lille forskel mellem den observerede og den forventede klassifikation. Figur 39 illustrerer, hvorvidt der er en signifikant forskel mellem de observerede og de forventede værdier, og værdierne for hvert step fremgår af Figur 40. Hvis Hosmer & Lemeshow testen er signifikant, dvs. at p-værdien er under signifikansniveauet på 0,05, betyder det, at der er en forskel mellem de observerede og de forventede værdier, hvilket karakteriserer et forringet model fit. Derimod er modellens fit acceptabelt, hvis Hosmer & Lemeshow testen ikke er signifikant.

Step	Chi-square	df	Sig.
1	19,803	8	,011
2	4,940	8	,764

Figur 39 – Hosmer & Lemeshow test (Egen konstruktion i SPSS)

I kontingenstabellen i Figur 40 er stikprøven opdelt i ti klasser, hvorefter der for hver klasse er sammenlignet det observerede antal med det forventede antal. Hvis forskellen mellem antallet af de observerede og antallet af de forventede bliver større, jo mindre nøjagtigt forudsiger modellen, hvilket resulterer i en højere chi-i-anden værdi. Med udgangspunkt i kontingenstabellen i Figur 40 illustrerer Figur 39, at Hosmer & Lemeshow testen for step 1 er signifikant med en p-værdi på 0,011. Dette udtrykker, at der er en forskel mellem det observerede antal og det forventede antal i modellen i step 1. I step 2 i Figur 39 er Hosmer & Lemeshow testen derimod ikke signifikant, da p-værdien er 0,764, hvilket understreger en forbedring i modellens fit fra step 1 til step 2. Ud fra Hosmer & Lemeshow testen vurderes modellens fit til at være acceptabel. Hosmer & Lemeshow er dog baseret på en chi-i-anden test

jf. afsnit 6.3.2.4.2.1 *Statistiske mål*, hvorfor den kræver, at der er fokus på stikprøvestørrelsen, og derfor kan en vurdering af modellens overordnede fit ikke alene baseres på denne test.

		Status = Normal drift		Status = Opløst efter konkurs		
		Observed	Expected	Observed	Expected	Total
Step 1	1	24	23,717	0	,283	24
	2	21	23,266	3	,734	24
	3	24	22,758	0	1,242	24
	4	20	22,113	4	1,887	24
	5	22	21,049	2	2,951	24
	6	22	19,453	2	4,547	24
	7	20	17,467	4	6,533	24
	8	15	14,777	9	9,223	24
	9	3	7,431	21	16,569	24
	10	2	,971	25	26,029	27
Step 2	1	24	23,821	0	,179	24
	2	23	23,447	1	,553	24
	3	22	23,020	2	,980	24
	4	23	22,428	1	1,572	24
	5	22	21,391	2	2,609	24
	6	21	20,046	3	3,954	24
	7	20	17,879	4	6,121	24
	8	13	14,238	11	9,762	24
	9	4	6,169	20	17,831	24
	10	1	,563	26	26,437	27

Figur 40 – Kontingenstabel for Hosmer & Lemeshow test (Egen konstruktion i SPSS)

Den sidste metode der anvendes til at vurdere modellens overordnede fit er klassifikationsnøjagtigheden, der blev præsenteret i afsnit 6.3.2.4.2.3 *Klassifikationsnøjagtighed*. Klassifikationsmatrixen, der ses i Figur 41, repræsenterer de niveauer af den forudsagte nøjagtighed, der er opnået ved den logistiske regressionsmodel, hvor skæringsværdien er 0,500. Målet for den forudsagte nøjagtighed er hitraten, der er den procentdel af observationer, der er korrekt klassificeret.

			Predicted		
			Status		
	Observed		Normal drift	Opløst efter konkurs	Percentage Correct
Step 1	Status	Normal drift	168	5	97,1
		Opløst efter konkurs	25	45	64,3
	Overall Percentage				87,7
Step 2	Status	Normal drift	165	8	95,4
		Opløst efter konkurs	22	48	68,6
	Overall Percentage				87,7

Figur 41 – Klassifikationsmatrix (Egen konstruktion i SPSS)

Hitraten kan som nævnt i afsnit 6.3.2.4.2.3 *Klassifikationsnøjagtighed* sammenlignes med *maximum chance* kriteriet og *proportional chance* kriteriet. *Maximum chance* kriteriet er de 71,2 %, der fremgår af Figur 32, med et tillæg på 25 %, hvilket resulterer i et *maximum chance* kriterie på 89 % $(71{,}2\,\% * 1{,}25 = 89\,\%)$. *Proportional chance* kriteriet udregnes ud fra formlen i afsnit 6.3.2.4.2.3 *Klassifikationsnøjagtighed* på baggrund af stikprøven til indeværende analyse, hvilket resulterer i et forhold af observationer i gruppe 1 på 0,71 $\left(\frac{173}{243} = 0{,}71\right)$ og et forhold af observationer i gruppe 2 på 0,29 $\left(\frac{70}{243} = 0{,}29\right)$. Ud fra disse forhold og med et tillæg på 25 % giver det et *proportional chance* kriterie på 74 % $\left((0{,}71^2 + 0{,}29^2) * 1{,}25 = 0{,}74\right)$.

Den overordnede hitrate i Figur 41 for step 1 er 87,7 %, hvilket overstiger *proportional chance* kriteriet på 74 %, hvorimod den næsten er på højde med de 89 %, der udgør *maximum chance* kriteriet. Den gruppespecifikke hitrate for opløst efter konkurs i step 1 er dog kun 64,3 %, hvilket ikke overstiger hverken *proportional chance* kriteriet eller *maximum chance* kriteriet, derfor er der ud fra resultaterne i step 1 baggrund for at videreudvikle modellen. Den overordnede hitrate for step 2 i Figur 41 er ligeledes 87,7 %, hvilket er identisk med hitraten i step 1. Derimod er den gruppespecifikke hitrate for opløst efter konkurs øget til 68,6 % i step 2, men på trods af denne forbedring er der stadig et stykke op til de to kriterier. I indeværende analyse er gruppen for virksomheder i normal drift en del større end gruppen for virksomheder, der er opløst efter konkurs, og det har betydning for anvendelsen af *maximum chance* kriteriet, idet stikprøvens sammensætning gør det svært at overstige dette kriterie. Det skyldes, at virksomheder i normal drift, jf. Figur 28, udgør 71 % af stikprøven, hvilket resulterer i et højt *maximum chance* kriterie på 89 %, hvilket indebærer, at få virksomheder må kategoriseres forkert før modellens fit ikke accepteres. Derfor kan der argumenteres for, at tillægget på de 25 % i indeværende analyse skaber en øvre grænse for *maximum chance* kriteriet, jf. afsnit 6.3.2.4.2.3 *Klassifikationsnøjagtighed*.

Ud fra afsnit 6.3.2.4.2.3 *Klassifikationsnøjagtighed* refererer klassifikationsmatrixen også til type 1 og type 2 fejl, hvor type 1 fejl er de konkursramte virksomheder, der er forudsagt til at være i normal drift, hvorimod type 2 fejl er de virksomheder i normal drift, der er forudsagt til at være gået konkurs. En perfekt model vil have nul type 1 og nul type 2 fejl. De omkostningsmæssige konsekvenser ved type 1 og type 2 fejl er forskellige, men generelt er omkostningerne ved type 1 fejl større end ved type 2 fejl. For en långiver eller en aktionær kan en type 1 fejl resultere i, at hele investeringen går tabt, hvorimod en type 2 fejl kan resultere i, at

der ikke foretages en investering, hvoraf en evt. gevinst ikke opnås. Det fremgår af Figur 41, at der i både step 1 og step 2 er 30 observationer, der er klassificeret forkert. I step 1 er der 25 type 1 fejl, hvorimod der er fem type 2 fejl, og i step 2 er der 22 type 1 fejl, hvorimod der er otte type 2 fejl. Udviklingen fra step 1 til step 2 viser, at mængden af type 1 fejl er reduceret, hvilket er forbedring af modellens fit, da denne fejltype har de største omkostningsmæssige konsekvenser for en långiver eller en aktionær. Ud fra afsnit 6.3.2.4.2.3 *Klassifikationsnøjagtighed* kan den tilladte grænse for type 1 fejl fastsættes ud fra signifikansniveauet, og det er i indeværende analyse fastsæt til 5 % jf. afsnit 7.2.4.1 *Modelestimering*, hvilket resulterer i, at der maks. bør være 12 $(243 * 0{,}05 = 12)$ type 1 fejl. Klassifikationsmatrixen i Figur 41 illustrerer dog, at tilstedeværelsen af type 1 fejl overstiger den tilladte grænse, hvilket er kritisabelt for modellens overordnede fit, især fordi type 1 fejl, som tidligere nævnt, er forbundet med de største omkostningsmæssige konsekvenser.

Ovenstående metoder er alle med til at vurdere det overordnede fit af den logistiske regressionsmodel med variablerne C_TA (X_2) og RE_TA (X_4), da de hver især bidrager med interessante vinkler til vurderingen af modellens fit. Generelt er modellens fit acceptabelt ud fra en signifikant chi-i-anden test af ændringer i *-2LL* værdien samt en reducering af *-2LL* værdien, derudover er Nagelkerke R^2 over 60 %, samtidig med at Hosmer & Lemeshow testen ikke er signifikant. Derudover overstiger hitraten i klassifikationsmatrixen *proportional chance* kriteriet, og den generelle klassifikationsnøjagtighed er øget betydelig fra basismodellen. Modellens fit kan dog kritiseres ved, at *-2LL* værdien stadig er forholdsvis høj, derudover er Cox & Snell R^2 under 50 %, samtidig med at hitraten i klassifikationsmatrixen ikke overstiger *maximum chance* kriteriet. Derudover er mængden af type 1 fejl foruroligende, hvilket hænger sammen med den lave gruppespecifikke hitrate for virksomheder, der er opløst efter konkurs. Alle de anvendte metoder vidner dog om, at der er sket en forbedring i modellens overordnede fit fra step 1 til step 2, og denne forbedring er især tydelig i Hosmer & Lemeshow testen, hvorimod forbedringen er mere begrænset ud fra de andre parametre.

7.2.5 Trin 5 – Fortolkning af resultaterne

Fortolkningen af resultaterne bygger på den logistiske regressionsmodel, der blev estimeret i afsnit 7.2.4.1.2 *Stepwise modelestimering*, og derfor er Figur 35 også udgangspunktet for selve fortolkningen, hvorfor den er gengivet neden for som Figur 42.

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	RE_TA	-5,587	,806	48,019	1	,000	,004
	Constant	-,529	,202	6,881	1	,009	,589
Step 2[b]	C_TA	6,389	1,693	14,244	1	,000	595,011
	RE_TA	-6,195	,909	46,407	1	,000	,002
	Constant	-,851	,230	13,657	1	,000	,427

a. Variable(s) entered on step 1: RE_TA.
b. Variable(s) entered on step 2: C_TA.

Figur 42 – Variabler i ligningen (Egen konstruktion i SPSS)

Ud fra Figur 42 er det muligt at foretage en kontrol af koefficienternes signifikans, da Wald teststørrelsen (Wald) i Figur 42 afspejler den statistiske signifikans for hver koefficient ud fra, hvorvidt p-værdien (Sig.) er under signifikansniveauet. Dette fremgår af afsnit 6.3.2.5.1 *Kontrol af koefficienternes signifikans*, hvor formlen for Wald teststørrelsen ligeledes blev beskrevet, hvori både den logistiske koefficient (B) samt den tilhørende standardfejl (S.E.) indgår. Frihedsgraden på 1 (df) i Figur 42 er blot et udtryk for, at Wald teststørrelsen følger en chi-i-anden fordeling med en frihedsgrad på 1.

I step 2 i Figur 42 fremstår begge logistiske koefficienter samt konstanten som signifikante, da de alle har p-værdier under signifikansniveauet på 0,05. Ved anvendelsen af metoden forward stepwise (Wald) er det forudsigeligt, at koefficienterne i step 2 er signifikante, da koefficienterne ellers ikke ville blive implementeret i modellen, idet de således ikke opfylder adgangskriteriet.

Den endelige logistiske regressionsmodel i step 2 inkluderer de to variabler C_TA (X_2) og RE_TA (X_4), der jf. Figur 42 har logistiske koefficienter på hhv. 6,389 og -6,195 samt en konstant på -0,851. Konstanten på -0,851 illustrerer modellens skæringspunkt. Fortolkningen af den logistiske regressionsmodel opdeles i to, hvor der først fortolkes på retningsbestemmelsen, hvorefter størrelsen på forholdet fortolkes. Retningsbestemmelsen af forholdet kan jf. afsnit 6.3.2.5.2.1 *Retningsbestemmelse af forholdet* foretages ud fra de originale logistiske koefficienter (B) eller de eksponentielle koefficienter (Exp(B)), der begge er en del af Figur 42. De logistiske koefficienter kan aflæses til 6,389 for C_TA (X_2) og -6,195 for RE_TA (X_4), og de eksponentielle koefficienter kan aflæses til 595,011 for C_TA (X_2) og 0,002 for RE_TA (X_4) jf. Figur 42. Ud fra de originale logistiske koefficienter kan retningen på forholdet fortolkes ud fra fortegnet, og jf. step 2 i Figur 42 har C_TA (X_2) et positivt fortegn, hvorimod RE_TA (X_4) har et negativt fortegn. Det medfører, at der er et positivt forhold mellem den uafhængige variabel C_TA (X_2) og den forudsagte sandsynlighed for, at en virksomhed går konkurs. Derimod er der

et negativt forhold mellem den uafhængige variabel RE_TA (X_4) og den forudsagte sandsynlighed for, at en virksomhed går konkurs. En stigning i værdien af C_TA (X_2) vil således øge sandsynligheden for, at en virksomhed går konkurs, hvorimod en stigning i værdien af RE_TA (X_4) vil formindske sandsynligheden for, at en virksomhed går konkurs. Ud fra de eksponentielle koefficienter kan retningen på forholdet bestemmes ud fra, hvorvidt værdierne er over eller under 1,0, jf. afsnit 6.3.2.5.2.1 *Retningsbestemmelse af forholdet*, hvor en værdi over 1,0 indikerer et positivt forhold, hvorimod en værdi under 1,0 indikerer et negativt forhold. Den eksponentielle koefficient for C_TA (X_2) er 595,011, hvilket angiver et positivt forhold mellem den uafhængige variabel C_TA (X_2) og sandsynligheden for, at en virksomhed går konkurs. Derimod er den eksponentielle koefficient for RE_TA (X_4) 0,002, hvilket indikerer et negativt forhold mellem den uafhængige variabel RE_TA (X_4) og sandsynligheden for, at en virksomhed går konkurs. Fortolkningen af retningsbestemmelsen ud fra de eksponentielle koefficienter stemmer således overens med fortolkningen af forholdets retningsbestemmelse ud fra de originale logistiske koefficienter.

Fortolkningen af forholdets størrelse bør, jf. afsnit 6.3.2.5.2.2 *Størrelsen på forholdet*, foretages ud fra de eksponentielle koefficienter, da det er uhensigtsmæssigt at foretage denne fortolkning ud fra de originale logistiske koefficienter, hvorfor de ikke anvendes til fortolkningen af forholdets størrelse. Anvendelsen af de eksponentielle koefficienter er desuden den mest direkte metode til at vurdere størrelsen af ændringen i sandsynligheden for hver uafhængig variabel. Det skyldes, at værdien af den eksponentielle koefficient minus en er lig med den procentvise ændring i oddset, hvor oddset jf. afsnit 6.3.2.4.1.1 *Den logistiske transformation* er sandsynlighedsforholdet af de to udfald, og i indeværende analyse indebærer det, hvorvidt virksomheden er i normal drift eller opløst efter konkurs. Ud fra den eksponentielle koefficient på 595,011 for C_TA (X_2) i Figur 42, resulterer det i, at en stigning på en enhed i nøgletallet C_TA (X_2) øger oddset for, at en virksomhed går konkurs med 59401,1 % $\big((595{,}011 - 1{,}0) * 100 = 59401{,}1\,\%\big)$. Hvorimod den eksponentielle koefficient i Figur 42 på 0,002 for RE_TA (X_4) resulterer i, at en stigning på en enhed i nøgletallet RE_TA (X_4) formindsker oddset for, at en virksomhed går konkurs med 99,8 % $\big((0{,}002 - 1{,}0) * 100 = -99{,}8\,\%\big)$. Disse procenttal kan overstige 100 %, fordi det er oddset, der ændres dvs. sandsynlighedsforholdet af de to udfald, og dermed ikke selve sandsynligheden for, at en virksomhed går konkurs. De store ændringer i oddset skyldes bl.a. den negative konstant på -0,851, da den medfører et lavt udgangspunkt for sandsynligheden, hvorfor store ændringer i oddset er nødvendige for at opnå

markante ændringer. Hvis der sker en stigning i både C_TA (X_2) og RE_TA (X_4), har C_TA (X_2) størst indflydelse på ændringen af sandsynligheden for, at en virksomhed går konkurs. Det afspejles ved, at en ændring i C_TA (X_2) på en enhed resulterer i, at oddset forøges med 59401,1 %, hvorimod en ændring i RE_TA (X_4) på en enhed resulterer i, at oddset formindskes med 99,8 %.

Nøgletallet C_TA, der indgår i den logistiske regressionsmodel, analyserer jf. afsnit 6.2.3.2 *Cash/Total assets* virksomhedens øjeblikkelige betalingsevne, idet den sammenholder virksomhedens likvider med aktivsummen. En stigning i dette nøgletal kan realiseres ved, at virksomheden enten øger mængden af likvider eller minimerer aktivsummen, og ud fra ovenstående fortolkning vil en stigning i dette nøgletal øge sandsynligheden for, at en virksomhed går konkurs. Dette virker en anelse absurd, at en forbedring af en virksomheds øjeblikkelige betalingsevne samtidig øger sandsynligheden for, at virksomheden går konkurs, hvilket diskuteres i afsnit 8.1 *Giver resultaterne mening?*. Dette skal dog sammenholdes med analysen af nøgletallet i afsnit 7.1.2.6 *Nøgletal*, hvor det fremgår, at de konkursramte virksomheder imod alt forventning i gennemsnit har en højere værdi af nøgletallet C_TA end virksomhederne i normal drift. Ovenstående fortolkning indikerer, at hvis C_TA (X_2) stiger med en enhed øges oddset for, at en virksomhed går konkurs med 59401,1 %. Her skal det dog understreges, at en stigning på en enhed af nøgletallets nuværende værdi, således at nøgletallet i de fleste tilfælde vil overstige 1,0, ikke er realistisk, hvilket understreges af maksimumværdien af nøgletallet i afsnit 7.1.2.6 *Nøgletal*, hvor den for ingen af grupperne overstiger 1,0. Det skyldes, at likvider aldrig kan overstige aktivsummen, idet likvider indgår som en del af aktivsummen, hvorfor den vil øges tilsvarende stigningen i likvider. RE_TA, der er det andet nøgletal i den logistiske regressionsmodel, analyserer jf. afsnit 6.2.3.4 *Retained earnings/Total assets* mængden af geninvesteret indtjening eller tab, der udtrykker evnen til at modstå evt. tab. En stigning i dette nøgletal kan realiseres ved at øge det overførte resultat eller mindske aktivsummen, og ud fra ovenstående fortolkning af den logistiske regressionsmodel, vil en stigning i RE_TA (X_4) mindske sandsynligheden for, at en virksomhed går konkurs. Det virker fornuftigt, at en forbedring af virksomhedens evne til at modstå evt. tab formindsker sandsynligheden for, at virksomheden går konkurs, hvilket uddybes i afsnit 8.1 *Giver resultaterne mening?*. Dette stemmer ligeledes overens med analysen af nøgletallet i afsnit 7.1.2.6 *Nøgletal*, hvor det fremgår, at virksomheder i normal drift i gennemsnit har en højere værdi af nøgletallet RE_TA end konkursramte virksomheder. Ud fra ovenstående fortolkning vil

en stigning i RE_TA (X_4) med en enhed formindske oddset for, at en virksomhed går konkurs med 99,8 %. En stigning i RE_TA (X_4) med en enhed, således at nøgletallet i de fleste tilfælde vil overstige 1,0, er dog ikke realistisk, idet overført resultat sjældent er større end aktivsummen, hvilket understreges i afsnit 7.1.2.6 *Nøgletal*, hvor ingen af grupperne har en maksimumværdi af nøgletallet RE_TA, der overstiger 1,0.

En anden metode til at forstå, hvordan de logistiske koefficienter definerer sandsynligheden, er at kalkulere den forudsagte sandsynlighed for hver observations værdier af de uafhængige variabler. Denne metode kan ligeledes uddybe sammenhængen mellem oddset og sandsynligheden for, at en virksomhed går konkurs. I en stepwise modelestimering for logistisk regression anvendes middelværdien for observationerne til at kalkulere logaritmen af sandsynlighedsværdien ud fra den estimerede model (Hair et al., 2010). Middelværdien for de to grupper, hhv. normal drift og opløst efter konkurs, er beregnet ud fra den totale stikprøve. Dette er gjort for overskuelighedens skyld, da det er de samme middelværdier, som der er analyseret tidligere i afsnit 7.1.2.6 *Nøgletal*. Anvendelsen af middelværdierne for de to grupper indikerer, hvad den forudsagte sandsynlighed ville være for en typisk virksomhed, der er i normal drift eller opløst efter konkurs. De anvendte middelværdier for normal drift er 0,04858 for C_TA (X_2) og 0,31758 for RE_TA (X_4), og middelværdierne for opløst efter konkurs er 0,09447 for C_TA (X_2) og -0,56177 for RE_TA (X_4). Middelværdierne og kalkulationen af de forudsagte sandsynligheder for de to grupper fremgår af Figur 43, hvor de anvendte formler ligeledes visualiseres.

	Gruppe 0: Normal drift	**Gruppe 1: Opløst efter konkurs**
Middelværdi: X_2	0,04858	0,09447
Middelværdi: X_4	0,31758	-0,56177
Logit værdi[a]	-2,508	3,233
Odds[b]	0,081	25,356
Sandsynlighed[c]	0,075	0,962

[a] Kalkuleret som: Logit = $-0{,}851 + 6{,}389X_2 - 6{,}195X_4$

[b] Kalkuleret som: Odds = e^{Logit}

[c] Kalkuleret som: Sandsynlighed = Odds/(1 + Odds)

Figur 43 – Udregning af sandsynlighedsværdier (Inspireret af Hair et al., 2010, s. 338)

Først kalkuleres logit værdien for hver gruppe ved at indsætte gruppens middelværdier for de to uafhængige variabler i den logistiske regressionsmodel. De andre komponenter i den logistiske regressionsmodel er konstanten og de originale logistiske koefficienter, der er en del af Figur 42. Kalkulationen af de forudsagte sandsynligheder resulterer i en logit værdi på -2,508 for gruppe 0 og en logit værdi på 3,233 for gruppe 1 jf. Figur 43. For at kalkulere odds værdien benyttes antilogaritmen til logit værdien, hvilket resulterer i odds værdier på hhv. 0,081 for gruppe 0 og 25,356 for gruppe 1, og disse værdier kan ligeledes ses i Figur 43. Derefter anvendes den kalkulerede odds værdi til at beregne sandsynligheden for at gå konkurs, og det sker ud fra formlen i afsnit 6.3.2.5.2.2 *Størrelsen på forholdet*. Det resulterer i, at en typisk virksomhed i normal drift (gruppe 0) jf. Figur 43 har en sandsynlighed på 0,075 for at blive klassificeret forkert som en virksomhed, der er opløst efter konkurs. Derimod har en typisk konkursramt virksomhed (gruppe 1) en sandsynlighed på 0,962 jf. Figur 43 for at blive klassificeret korrekt som en virksomhed, der er opløst efter konkurs.

Udregningerne i Figur 43 vidner om, at den logistiske regressionsmodel formår at adskille de to gruppers middelværdier i form af den forudsagte sandsynlighed, hvilket er en medvirkende årsag til den forholdsvis høje klassifikationsnøjagtighed, der er opnået i Figur 41. Det skyldes, at en sandsynlighed på 0,075 vil klassificere en virksomhed til at være i normal drift (gruppe 0), og en sandsynlighed på 0,962 vil klassificere en virksomhed til at være opløst efter konkurs (gruppe 1). Denne klassifikation ud fra de kalkulerede sandsynligheder stemmer således overens med de grupper, der er udgangspunktet for de anvendte middelværdier. Dvs. at middelværdierne for gruppe 0 resulterer i en sandsynlighed, der klassificeres som gruppe 0, og middelværdierne for gruppe 1 resulterer i en sandsynlighed, der klassificeres som gruppe 1. Fortolkningen af resultaterne af den logistiske regressionsmodel er, at de logistiske koefficienter definerer et positivt forhold for C_TA (X_2) og et negativt forhold for RE_TA (X_4). Derudover bidrager den med en metode til at vurdere virkningerne på oddset ud fra ændringer i en eller begge variabler, hvor ændringer i oddset afspejler ændringer i sandsynligheden for, at en virksomhed går konkurs.

7.2.6 Trin 6 – Validering af resultaterne

Det sidste trin omhandler den interne og eksterne validering af den logistiske regressionsmodel, og det sker ud fra en hold-out stikprøve, hvoraf der foretages en vurdering af dens forudsagte nøjagtighed. Validering er opnået, hvis den logistiske regressionsmodel opnår et acceptabelt

niveau ift. at klassificere observationer, der ikke har været en del af estimeringsprocessen. Hold-out stikprøven, der blev præsenteret i afsnit 7.2.2.2 *Stikprøvestørrelsen*, er en del af den oprindelige stikprøve, hvilket indebærer, at valideringen af resultaterne primært omhandler intern validitet og kun en indikation af ekstern validitet jf. afsnit 6.3.2.6 *Trin 6 – Validering af resultaterne*. Hold-out stikprøven består af 138 virksomheder, hvoraf 98 virksomheder er i normal drift, og 40 virksomheder er opløst efter konkurs. Antallet af konkursramte virksomheder opfylder dermed ikke 10:1 forholdet af observationer ift. antallet af uafhængige variabler, hvilket er debatteret i afsnit 7.2.2.2 *Stikprøvestørrelsen*. Udvælgelsen af disse virksomheder er, jf. afsnit 7.2.2.3 *Opdeling af stikprøven*, foretaget tilfældigt, således at selve opdelingen af stikprøven ikke får betydning for valideringsprocessen.

Klassifikationsmatrixen i Figur 44 er udarbejdet ud fra den logistiske regressionsmodel, der blev estimeret i afsnit 7.2.4.1.2 *Stepwise modelestimering* med data fra hold-out stikprøven, der kan ses i arket *Hold-out* i Bilag 3. Udregningerne forud for udarbejdelsen af klassifikationsmatrixen er også en del af arket *Hold-out* i Bilag 3, hvor det fremgår, at der først er udregnet en logit værdi, hvorefter odds værdien er kalkuleret for til sidst at anvende odds værdien til at udregne sandsynlighedsværdien. Klassifikationen af de enkelte virksomheder er foretaget ud fra, hvorvidt sandsynlighedsværdien er over eller under 0,500, idet denne skæringsværdi også er anvendt i klassifikationsmatrixen for stikprøven jf. afsnit 7.2.4.2 *Vurdering af modellens overordnede fit*. Hvis sandsynlighedsværdien er over 0,500 klassificeres virksomheden som opløst efter konkurs, hvorimod en sandsynlighedsværdi under 0,500 klassificerer virksomheden til at være i normal drift. Dette skyldes den værditildeling, der blev foretaget i afsnit 7.2.2.1 *Udvælgelse af afhængig og uafhængige variabler*, hvor gruppen i normal drift blev tildelt værdien 0, hvorimod gruppen opløst efter konkurs blev tildelt værdien 1.

		Predicted		
		Status		
Observed		Normal drift	Opløst efter konkurs	**Percentage Correct**
Status	Normal drift	88	10	89,8
	Opløst efter konkurs	20	20	50,0
Overall Percentage				78,3

Figur 44 – Klassifikationsmatrix, Hold-out (Egen tilvirkning)

Udseendet på Figur 44 er tilpasset således, at den minder om klassifikationsmatrixen i Figur 41, hvilket gør det nemmere at sammenligne resultaterne af disse klassifikationsmatrixer. Den overordnede hitrate i Figur 44 er 78,3 %, hvilket er en del lavere end den overordnede hitrate fra step 2 i Figur 41 på 87,7 %. Den gruppespecifikke hitrate for normal drift er 89,8 % i Figur 44, idet 88 virksomheder klassificeres korrekt til at være i normal drift, hvorimod ti virksomheder klassificeres forkert til at være opløst efter konkurs. Denne hitrate er forholdsvis høj, men på trods af dette er den en smule lavere end den gruppespecifikke hitrate for normal drift i step 2 i Figur 41. Den gruppespecifikke hitrate for opløst efter konkurs er 50,0 % i Figur 44, da halvdelen af virksomhederne klassificeres forkert til at være i normal drift, og den resterende halvdel klassificeres korrekt til at være opløst efter konkurs. En gruppespecifik hitrate på 50,0 % er kritisk og foruroligende lav, og den er da også en del under den gruppespecifikke hitrate for opløst efter konkurs i step 2 i Figur 41.

Den overordnede hitrate på de 78,3 % i Figur 44 skal sammenlignes med *maximum chance* kriteriet og *proportional chance* kriteriet, der begge er udregnet i afsnit 7.2.4.2 *Vurdering af modellens overordnede fit* til hhv. 89 % og 74 %. Ud fra hitraten i Figur 44 overstiger den *proportional chance* kriteriet, hvorimod den er en del under *maximum chance* kriteriet. Fokuseres der på den gruppespecifikke hitrate, er den for normal drift 89,8 %, hvilket overstiger både *maximum chance* kriteriet og *proportional chance* kriteriet. Derimod er den gruppespecifikke hitrate for opløst efter konkurs kun 50,0 %, hvilket er langt under begge de udregnede standarder. Den overordnede hitrate og den lave gruppespecifikke hitrate for opløst efter konkurs indikerer, at der er et stort potentiale for at tilpasse og videreudvikle modellen, hvilket uddybes i afsnit 8.4 *Forbedringsforslag*.

Den meget lave gruppespecifikke hitrate for opløst efter konkurs illustreres også ved mængden af type 1 fejl, idet der i Figur 44 er 20 type 1 fejl, hvor konkursramte virksomheder er klassificeret til at være i normal drift. Figur 44 indeholder også ti type 2 fejl, hvor virksomheder i normal drift er klassificeret til at være gået konkurs. Type 1 fejl er, som nævnt i afsnit 7.2.4.2 *Vurdering af modellens overordnede fit*, den fejltype, der er forbundet med de største økonomiske konsekvenser, hvilket understreger, hvorfor den lave gruppespecifikke hitrate for opløst efter konkurs er problematisk. Ud fra Figur 44 er der i alt 30 type 1 og type 2 fejl, hvilket er det samme antal fejl som i Figur 41, forskellen er blot, at der til udarbejdelsen af Figur 44 er anvendt en stikprøve på 138 virksomheder, hvorimod der til udarbejdelsen af Figur 41 er anvendt en stikprøve på 243 virksomheder. I afsnit 7.2.4.1 *Modelestimering* blev

signifikansniveauet fastlagt til 0,05, hvilket afspejler den tilladte grænse for type 1 fejl, og det medfører, at der maks. bør være syv $(138 * 0{,}05 = 7)$ type 1 fejl for hold-out stikprøven. Antallet af type 1 fejl i Figur 44 er dog mere end en fordobling af det tilladte antal, hvilket illustrerer, at den estimerede logistiske regressionsmodel ikke opnår et acceptabelt fit for hold-out stikprøven.

Klassifikationsnøjagtigheden for hold-out stikprøven er, som tidligere nævnt, over *proportional chance* kriteriet, hvorimod den er under *maximum chance* kriteriet, og derudover indeholder klassifikationsmatrixen en stor mængde type 1 fejl. Dette medfører, at der ikke etableres en høj intern validitet for den logistiske regressionsmodel, hvilket ligeledes indikerer en tilsvarende lav ekstern validitet. For en dybere undersøgelse af den eksterne validitet kan der suppleres med ekstra stikprøver fra relevante populationer, hvorefter klassifikationsnøjagtigheden kan vurderes i flere forskellige senarier. Den begrænsede interne validitet understreges især ved den gruppespecifikke hitrate for opløst efter konkurs, da den kun er på 50,0 % for hold-out stikprøven, hvilket er foruroligende lav. Dette resulterer i, at den logistiske regressionsmodel ikke indikerer en tilstrækkelig intern validitet, hvilket medfører, at resultaterne, der er estimeret ud fra den logistiske regressionsmodel, umiddelbart ikke kan accepteres. Det medfører desuden, at den logistiske regressionsmodel ikke kan generaliseres, hvorfor anvendelsen af modellen uden for stikprøven begrænses.

7.3 Afrunding af analysen

Ud fra den deskriptive statistik er stikprøven til indeværende analyse ikke repræsentativ for populationen. Dette vurderes ud fra en skævvridning mellem stikprøven og populationens udvikling i antal af konkurser, brancheforde ling og virksomhedsformer. Derudover er de to grupper i stikprøven hhv. virksomheder i normal drift og virksomheder, der er opløst efter konkurs heller ikke matchet op imod hinanden. Dette vurderes ud fra en væsentlig forskel i brancheforde ling og virksomhedsformer, hvorimod etableringsåret for de to grupper og medarbejderantallet er mere sammenlignelig.

Den statistiske analyse af stikprøven med den multivariate teknik logistisk regression resulterede i, at der ud fra en forward stepwise (Wald) blev estimeret en logistisk regressionsmodel, der indeholdte to variabler. I step 1 blev variablen RE_TA (X_4) implementeret, hvorefter variablen C_TA (X_2) blev tilføjet i step 2. Derefter blev den logistiske regressionsmodel med de to

variabler vurderet til at have et acceptabelt *goodness-of-fit* ud fra en signifikant chi-i-anden test af ændringer i *-2LL* værdien, en Nagelkerke R^2 på over 60 % samt en Hosmer & Lemeshow test, der ikke er signifikant. Derudover besidder den logistiske regressionsmodel en forholdsvis høj klassifikationsnøjagtighed på 87,7 %, der er øget betydelig fra basismodellen. Fortolkningen af den logistiske regressionsmodel illustrerer, at nøgletallet C_TA, der udtrykker virksomhedens øjeblikkelige betalingsevne, har en positiv sammenhæng med sandsynligheden for, at en virksomhed går konkurs. Det betyder således, at en forbedring af virksomhedens øjeblikkelige betalingsevne øger sandsynligheden for, at en virksomhed går konkurs. Denne sammenhæng er, som tidligere nævnt, yderst mærkværdig, idet en forbedring af virksomhedens øjeblikkelige betalingsevne ikke burde indikere en øget risiko for konkurs. Fortolkningen af det andet nøgletal, der indgår i den logistiske regressionsmodel indebærer, at nøgletallet RE_TA, der omhandler virksomhedens evne til at modstå tab, har en negativ sammenhæng med sandsynligheden for, at en virksomhed går konkurs. Denne sammenhæng indikerer, at en forbedring af virksomhedens evne til at modstå tab minimerer sandsynligheden for, at en virksomhed går konkurs. Denne sammenhæng virker mere reel, hvis der sammenlignes med den førhen nævnte sammenhæng ift. C_TA. Ud fra en hold-out stikprøve valideres den estimerede logistiske regressionsmodel ikke, idet klassifikationsnøjagtigheden ikke er tilstrækkelig høj, samtidig med at den gruppespecifikke hitrate for opløst efter konkurs er foruroligende lav med kun 50 %, hvilket illustreres med mængden af type 1 fejl. Modellens manglende validering resulterer i, at den ikke kan generaliseres til populationen, hvilket medfører, at den kun er anvendelig ift. den stikprøve, der er benyttet ifm. at udarbejde den logistiske regressionsmodel. Årsagen til at den logistiske regressionsmodel ikke valideres kan skyldes, at den anvendte stikprøve til udarbejdelsen af modellen, som tidligere påpeget, ikke er repræsentativ for populationen, hvorfor den derfor ikke kan generaliseres til populationen.

Ovenstående analyse besvarer således specialets undersøgelsesproblem, der blev præsenteret i afsnit 2 *Problemfelt*, idet den logistiske regressionsmodel illustrerer, at C_TA bidrager med en positiv sammenhæng, hvorimod RE_TA bidrager med en negativ sammenhæng med forudsigelsen af konkurser. Derimod indeholder den logistiske regressionsmodel kun to ud af de seks udvalgte nøgletal, hvilket indikerer, at de fire resterende udvalgte nøgletal ikke bidrager til forudsigelsen af konkurser. Ud fra den manglende validering af modellen bidrager C_TA og RE_TA dog ikke til en generel forudsigelse af danske virksomheders konkurs, hvorimod de i stedet kun bidrager til at forudsige konkurser af de danske virksomheder, der indgår i stikprøven.

8 Diskussion

Afsnittet indebærer en diskussion af de statistiske resultater, der er opnået igennem analysen, herunder hvordan disse resultater kan anvendes, samt hvad der ligger til grund for analysens resultater. Derefter diskuteres det, hvordan den logistiske regressionsmodel kan videreudvikles for at styrke modellen. Derudover sammenholdes resultaterne af indeværende speciale med resultaterne af de tidligere undersøgelser inden for konkursprognosticering, hvoraf evt. forskelle diskuteres. Afslutningsvis diskuteres generelle kritikpunkter ift. konkursprognosticerings-modeller.

8.1 Giver resultaterne mening?

Ud fra ovenstående statistiske analyse kan der diskuteres, hvorvidt resultaterne af den logistiske regressionsmodel giver mening. Nøgletallet RE_TA omhandler, som tidligere nævnt, virksomhedens evne til at akkumulere indtjening for derigennem at kunne modstå evt. tab. Derfor giver det mening, at RE_TA (X_4) har en negativ sammenhæng med sandsynligheden for at gå konkurs. Dette afspejles også i den stikprøve, der er udgangspunktet for estimeringen af den logistiske regressionsmodel, idet gruppen med virksomheder i normal drift har en højere middelværdi af dette nøgletal ift. de konkursramte virksomheder jf. afsnit 7.1.2.6 *Nøgletal*. Den negative sammenhæng understreges ved, at en forbedret evne til at akkumulere overskud og dermed modstå evt. tab gør virksomhederne mere modstandsdygtige overfor dårlige perioder, der i sidste ende kan resultere i en konkurs. En forbedring af denne modstandsdygtighed er speciel vigtig, såfremt virksomhederne skal undgå en konkurs, da Stabell jf. afsnit 1 *Indledning* understreger, at der for flere virksomheder ikke skal mere end én dårlig måned til at vende en positiv udvikling til en negativ, og en negativ udvikling for en virksomhed kan i sidste ende resultere i en konkurs. Et fokus på at forbedre nøgletallet RE_TA giver således mening ift. at mindske virksomhedens sandsynlighed for at gå konkurs.

Nøgletallet C_TA, der blev implementeret i den logistiske regression i step 2, omhandler, som tidligere nævnt, virksomhedens øjeblikkelige betalingsevne, idet nøgletallet udtrykker, hvor stor en andel likvider udgør af aktivsummen. Derfor giver det umiddelbart ikke mening, at C_TA (X_2) har en positiv sammenhæng med sandsynligheden for at gå konkurs. Den positive sammenhæng indikeres dog i den anvendte stikprøve, idet konkursramte virksomheder har en højere middelværdi for nøgletallet C_TA ift. virksomheder i normal drift. Men på trods af dette

giver det stadigvæk ikke mening, at en forbedring af virksomhedens øjeblikkelige betalingsevne maksimerer sandsynligheden for, at en virksomhed går konkurs. Det skyldes, at en højere andel af likvider ift. aktivsummen indikerer en vis grad af sikkerhed fra en kreditors synspunkt, hvorfor det ikke burde øge risikoen for, at den pågældende virksomhed går konkurs. Den omtalte sammenhæng kan således motivere virksomheder til at begrænse mængden af likvider, og dermed forringe den øjeblikkelige betalingsevne, hvilket må betragtes som uhensigtsmæssigt, idet en virksomhed aldrig bør forringe sin betalingsevne for at undgå en konkurs.

8.2 Resultaternes anvendelse

Den mislykkede validering af den logistiske regressionsmodel har, som tidligere nævnt, betydning for den videre anvendelse af modellen samt brugbarheden af de opnåede resultater. Forud for dette er det dog væsentligt at understrege, at idet stikprøven i afsnit 7.1.2 *Deskriptiv statistik* er vurderet til ikke at være repræsentativ for populationen, er det ikke overraskende at modellen ikke valideres. Den manglende validering resulterer i, at modellen ikke kan generaliseres til virksomheder uden for stikprøven, hvilket begrænser anvendeligheden af den logistiske regressionsmodel. Derfor er det ud fra modellen ikke muligt at generalisere, hvilke nøgletal der bidrager til at forudsige konkurser i danske virksomheder. Det medfører, at de to nøgletal, der indgår i den estimerede model, kun bidrager til at forudsige konkurser i de danske virksomheder, der indgår i stikprøven. Den begrænsede anvendelighed af modellen ændrer dog ikke på, at den statistiske analyse indikerer, at det er muligt at opstille en konkurs-prognosticeringsmodel, der tager udgangspunkt i danske virksomheder.

De to nøgletal, der indgår i den logistiske regressionsmodel, er forholdsvis lettilgængelige, hvilket ligeledes understreger, at hvis modellen var blevet valideret, er det realistisk, at virksomheder kan arbejde videre med disse nøgletal. Det er dog vigtigt at understrege, at der findes ingen samlet økonomisk teori, der forklarer, hvad der egentlig forårsager en konkurs, og hvordan den afspejles i virksomhedernes regnskaber (Christensen & Nielsen, 2012). Derfor bør en virksomhed være påpasselig med kun at fokusere på at forbedre de parametre, der ud fra en konkursprognosticeringsmodel har indflydelse på sandsynligheden for at gå konkurs, idet det er muligt at andre parametre kan få betydning. Ændringer i fx markedsvilkår, kundepræferencer, lovgivning eller teknologi kan ligeledes have indflydelse på, hvorvidt en virksomhed går konkurs. Derudover kan interne forhold såsom ændringer i ledelsen eller investorer bag virksomheden også have en væsentlig betydning for, hvordan virksomheder klarer sig

økonomisk. Hvis det vurderes, at disse parametre ikke har indflydelse på, hvorvidt en virksomhed går konkurs, kan det undre en, hvorfor virksomheder overhovedet går konkurs. Det skyldes, at idet der ofte kun er relativt få nøgletal i en konkursprognosticeringsmodel, vil det medføre, at virksomheder, der tager højde for disse nøgletal, har mulighed for at undgå en konkurs. Dette taler for, at sandsynligheden for at gå konkurs påvirkes af mere end blot enkelte nøgletal.

8.3 Årsager til resultaterne

Årsagerne til analysens resultater eller mangel på samme kan skyldes forskellige aspekter, der efterfølgende diskuteres. Stikprøven er, som tidligere nævnt, ikke repræsentativ, hvilket formentlig har haft betydning for den mislykkede validering. Stikprøvens manglende repræsentativitet kan skyldes, at der under dataindsamlingen i afsnit 5.2.1 *Dataindsamling* er fastsat for mange søgekriterier, som de enkelte virksomheder skulle opfylde for at være en del af stikprøven. Her kan især nævnes søgekriterierne antal medarbejdere, etableringsdato og dato for regnskabsafslutning, hvor der kan diskuteres, hvorvidt disse søgekriterier er nødvendige og relevante. Derudover er der ift. indsamlingen af datasættet for virksomheder i normal drift fastsat en del søgekriterier til regnskabsposter, der ikke indgår i eksportkriteriet, og derfor ikke har relevans for indeværende speciale. Hvis disse søgekriterier var undladt havde det resulteret i en større stikprøve, der formentlig havde været mere repræsentativ for populationen, idet en større stikprøve altid vil minde mere om populationen frem for en mindre stikprøve. Det skal dog nævnes, at et af argumenterne for at anvende ovenstående søgekriterier er, at det begrænsede stikprøven, hvilket var hensigtsmæssigt ud fra et tidsmæssigt perspektiv.

Indeværende analyse er som bekendt kun udarbejdet på baggrund af regnskabsdata fra år 2011, derfor har det ikke været muligt at foretage en analyse af den historiske udvikling, hvilket også kan have påvirket resultaterne. Derudover fremgik det i afsnit 7.1.2.1 *Konkurser*, at konkurserne i stikprøven er sket over en periode på fem år fra år 2011 til år 2015, hvilket kan være en medvirkende årsag til den klassifikationsnøjagtighed, der er opnået i analysen. Beaver (1966), der var den første til at udarbejde en konkursprognosticeringsmodel, argumenterer dog for, at det er muligt at forudsige konkurser på baggrund af nøgletal min. fem år ude i fremtiden, hvilket taler for, at det kun har haft begrænset betydning, at virksomhederne er gået konkurs over en periode på fem år.

Udover at dataindsamlingsmetoden kan have medført stikprøvens manglende repræsentativitet, er det også muligt at den frasortering, der er foretaget i afsnit 7.1.1 *Databearbejdning* kan have påvirket, hvor repræsentativ stikprøven er for populationen. Det skyldes, at en frasortering på 51 % af de konkursramte virksomheder ikke kan undgå at påvirke den stikprøve, der benyttes i analysen. På trods af denne frasortering indikerede den deskriptive statistik i afsnit 7.1.2.6 *Nøgletal*, at stikprøven formentlig indeholder outliers, hvilket ligeledes kan have påvirket analysens resultater, da der i indeværende speciale ikke er forsøgt at fjerne eller korrigere dem jf. afsnit 7.1.2.6 *Nøgletal*.

De udvalgte nøgletal til indeværende speciale har formentlig også haft stor betydning for de resultater, der er opnået i analysen, idet det er disse nøgletal, der er analyseret i SPSS forud for opstillingen af den logistiske regressionsmodel. Derfor har valget om at inddrage disse seks nøgletal frem for andre fx kvalitative nøgletal haft betydning for resultaterne. Dette skal ses ud fra, at valget af nøgletal i de fleste tilfælde enten er subjektivt bestemt ud fra forfatterens egen overbevisning, ad hoc bestemt ud fra hvad andre forfattere har haft succes med eller rent statistisk bestemt ud fra et meget stort antal mulige nøgletal (Christensen & Nielsen, 2012). I indeværende speciale er valget af nøgletal en blanding af, hvad andre forfattere har haft succes med, hvor der i dette tilfælde benyttes Altman & Sabatos (2007) nøgletal, samt forfatterens subjektive vurdering ift. at inddrage soliditetsgraden som supplement. Valget af nøgletal til indeværende speciale adskiller sig således en del fra, hvordan der generelt udvælges nøgletal til udarbejdelsen af en konkursprognosticeringsmodel. Fælles for dem alle er dog, at der anvendes nøgletal på trods af, at der er risiko for, at disse er manipuleret i en konkurstruet virksomhed (Schack, 2009). Tilstedeværelsen af manipulerede nøgletal er svær at tage højde for, og derfor kan det heller ikke udelukkes, at flere af virksomhederne i stikprøven har manipuleret med nøgletallene, hvilket også kan have påvirket analysens resultater.

8.4 Forbedringsforslag

Ud fra ovenstående er det interessant at diskutere konkrete forbedringsforslag, der kunne have medvirket til, at den logistiske regressionsmodel blev valideret i Trin 6 i den statistiske analyse. En videreudvikling af modellen er interessant, fordi den jf. afsnit 7.2.4.2 *Vurdering af modellens overordnede fit* havde et acceptabelt fit på trods af, at den ikke blev valideret. Den primære årsag til, at den logistiske regressionsmodel ikke blev valideret er stikprøven, der har været anvendt til estimeringen af modellen. Dette understreges ved, at fem af de seks udvalgte nøgletal samt den

multivariate teknik, logistisk regression, blev anvendt i Altman & Sabatos (2007) undersøgelse, hvor det lykkedes dem at validere deres model. Stikprøven bør derfor forbedres, således at den er repræsentativ for populationen. Dette sker primært ved at anvende en langt større stikprøve, hvilket også vil forhindre, at de evt. outliers får betydning for resultatet. En større stikprøve vil også gøre det muligt at opdele den således, at hold-out stikprøven overholder 10:1 forholdet for begge grupper, hvilket ligeledes vil forbedre udgangspunktet for analysen. Derudover skal mængden af de omfattende søgekriterier begrænses, således at virksomheder fx ikke frasorteres pga. regnskabsposter, der i realiteten ikke har indflydelse på estimeringen af den logistiske regressionsmodel.

Det fremgår i afsnit 7.1.2.2 *Brancher*, at der er stor forskel på branchefordelingen i de to grupper i den anvendte stikprøve til indeværende speciale. Derfor kunne det være interessant at benytte en stikprøve, der kun omfatter en specifik branche. Det skyldes, at det vil begrænse indflydelsen af de førnævnte eksterne parametre, da alle virksomhederne i stikprøven således ville blive påvirket nogenlunde ens af disse parametre. Det fremgår ligeledes i afsnit 7.1.2 *Deskriptiv statistik*, at de to grupper i stikprøven ikke er matchet, således at deres karakteristika minder om hinanden, og denne variation ses også i antallet af virksomheder i hver gruppe. De tidligere præsenterede undersøgelser af Beaver (1966) og Altman (1968) i afsnit 1 *Indledning* illustrerede, at deres konkursprognosticeringsmodeller blev foretaget på baggrund af en stikprøve, der bestod af et lige antal af virksomheder i hver gruppe, og derudover var grupperne i Altmans (1968) undersøgelse udvalgt på baggrund af størrelse og branche. Det skal dog understreges, at Ohlson (1980) og Altman & Sabato (2007) ikke benyttede en stikprøve, hvor grupperne var lige store. Derfor er det muligt at udarbejde en konkursprognosticeringsmodel på baggrund af en stikprøve, hvor virksomhederne, der ikke er gået konkurs, er i overtal. Dette indikerer, at forholdet mellem de to grupper i den anvendte stikprøve til indeværende speciale ikke nødvendigvis bør ændres for at forbedre modellen, selvom gruppernes karakteristika stadig bør forsøges at matches.

Udover en optimering af stikprøven er et andet konkret forbedringsforslag at inddrage mere end blot seks nøgletal til udarbejdelsen af en konkursprognosticeringsmodel, selvom Altman & Sabato (2007) formåede at opstille en valideret model med kun fem nøgletal. Den estimerede logistiske regressionsmodel indeholdte, som tidligere nævnt, kun to nøgletal, hvorfor der bør undersøges, hvorvidt andre nøgletal ville implementeres i modellen. Udvælgelsen af nøgletal til konkursprognosticeringsmodeller er påvirket af det manglende teoretiske grundlag for årsager til konkurser, hvilket medfører, at der ikke er en specifik metode, der benyttes til dette formål

(Christensen & Nielsen, 2012). Til indeværende analyse er der som bekendt anvendt en blanding af to af de hyppigste anvendte metoder, hvilket resulterede i, at kun to nøgletal blev implementeret i den logistiske regressionsmodel. Derfor foretrækkes den tredje metode som en forbedringsmulighed ift. udvælgelse af nøgletal, idet der foretages en statistisk bestemmelse af, hvilke nøgletal, der bør anvendes ud fra en stor mængde af nøgletal. Derfor indebærer denne metode ikke en subjektiv vurdering, hvilket anses for at være en fordel. Derudover har nyere litteratur konkluderet, at kvantitative variabler ikke er tilstrækkelige til at kunne forudsige virksomheders konkurs, hvorfor der bør inkluderes kvalitative variabler. De kvalitative variabler kan fx være medarbejderantal, virksomhedstype og branchetype (Altman & Sabato, 2007). Ud fra diskussionen om, at en konkurs kan skyldes flere forskellige parametre, kan de kvalitative nøgletal inddrages som en metode til at afdække disse parametre, idet de kvalitative nøgletal ikke knytter sig til det regnskabstekniske aspekt.

Den statistiske analyse i indeværende speciale kan også forbedres ved at inddrage supplerende statistiske undersøgelser, der kan styrke modellens opbygning og forklaringskraft. I den statistiske analyse undersøges der, hvorvidt hver af de seks nøgletal har en logit-lineær sammenhæng med sandsynligheden for at gå konkurs, hvilket fremgik af afsnit 7.2.4 *Trin 4 – Estimering af den logistiske regressionsmodel og vurdering af modellens overordnede fit*, hvor hypoteserne blev opstillet. Dette medfører, at den statistiske analyse ikke tager højde for, om en interaktion mellem flere af variablerne kan have en sammenhæng med sandsynligheden for at gå konkurs. En virksomheds konkurs skyldes sjældent en enkelt parameter eller tilstanden på et enkelt nøgletal, hvorfor det er interessant at analysere, hvorvidt en interaktion mellem flere af nøgletallene er statistisk signifikant ift. sandsynligheden for at gå konkurs. En anden statistisk undersøgelse, der er relevant at udføre for at forbedre modellen, er en korrelationsanalyse, idet alle de uafhængige variabler stammer fra virksomhedens regnskab, hvilket indikerer, at der kan være en forholdsvis høj korrelation mellem de anvendte nøgletal. Dette understreges yderligere ved, at begge variabler i den logistiske regressionsmodel indeholder regnskabsposten aktivsummen. Et evt. korrelationsforhold mellem variablerne er væsentlig at få analyseret, da det kan resultere i potentielle problemstillinger vedrørende multikollinearitet. Der omhandler, hvorvidt en uafhængig variabel kan forklares ud fra andre variabler i analysen, og hvis multikollineariteten øges, komplicerer det fortolkningen af den tilfældige variabel, idet det bliver svære at konstatere effekten af hver enkelt variabel (Hair et al., 2010). Det fremgår i Hair et al. (2010), at logistisk regression kan påvirkes af multikollinearitet, og derudover er stepwise

metoden især følsom over for dette, hvilket understreger, hvorfor denne statistiske undersøgelse bør prioriteres, hvis der videreudvikles på modellen. Det er også muligt, at der kan inddrages en helt anden multivariat teknik end logistisk regression, der gør det muligt at foretage en sammenligning mellem de to teknikkers modeller. En alternativ multivariat teknik er diskriminant analyse, idet den tidligere har været benyttet til udarbejdelse af konkursprognosticeringsmodeller. Inddragelsen af en ny multivariat teknik ændrer dog ikke, at den logistiske regression stadig anbefales som den mest brugbare multivariate teknik til konkursprognosticering.

Det fremgår i afsnit 7.2.6 *Trin 6 – Validering af resultaterne*, at valideringen primært omhandler den interne validering, fordi hold-out stikprøven oprindelig er en del af den totale stikprøve. Derfor er udarbejdelsen af en ekstern validering ligeledes interessant ifm. en videreudvikling af den logistiske regressionsmodel, hvor der inddrages eksterne populationer for at teste modellen. En af årsagerne til, at der i indeværende speciale ikke er foretaget en ekstern validering, er, at idet den logistiske regressionsmodel ikke valideres ift. intern validitet, er der umiddelbart ikke baggrund for efterfølgende at teste den eksterne validitet. Det er dog relevant at foretage en ekstern validering, såfremt den logistiske regressionsmodel videreudvikles ud fra ovenstående forbedringsforslag.

8.5 Sammenligning med tidligere undersøgelser

Efterfølgende sammenlignes resultaterne af specialet med resultaterne i Beaver (1966), Altman (1968) samt Altman & Sabato (2007), hvorefter forskellene diskuteres. Beaver (1966) anvendte, som tidligere nævnt, en matchet stikprøve og 30 nøgletal, hvilket resulterede i en klassifikationsnøjagtighed på 90 %. Altman (1968) anvendte ligeledes en matchet stikprøve og 22 nøgletal, og det resulterede i en klassifikationsnøjagtighed på 95 %. Altman & Sabato (2007), der på flere måder har været en inspiration til indeværende speciale, benyttede, som tidligere nævnt, en stikprøve, der ikke var matchet, og derudover anvendte de kun fem nøgletal, hvilket resulterede i en klassifikationsnøjagtighed på 75 %. Klassifikationsnøjagtigheden af analysens logistiske regressionsmodel er på 87,7 %, hvilket er en smule lavere end Beavers (1966) og Altmans (1968) på trods af, at deres undersøgelser er foretaget på baggrund af en matchet stikprøve. Anvendelsen af en matchet stikprøve med to lige store grupper burde minimere klassifikationsnøjagtigheden, da det er svære at forudsige et stort antal af de konkursramte virksomheder korrekt, hvis de udgør halvdelen af stikprøven. Stikprøven til indeværende analyse

er, som tidligere nævnt, ikke matchet, hvorfor modellen burde opnå en forholdsvis høj klassifikationsnøjagtighed, idet de konkursramte virksomheder jf. afsnit 7.2.2.3 *Opdeling af stikprøven* kun udgør 29 % af stikprøven. Altman & Sabato (2007) benyttede heller ikke en matchet stikprøve, idet deres andel af konkursramte virksomheder kun udgjorde 6 % af den samlede stikprøve. Umiddelbart virker det som en risikabel metode, at de konkursramte virksomheder kun udgør 6 %, idet det er muligt at fejlklassificere alle de konkursramte virksomheder og opnå en klassifikationsnøjagtighed på 94 %. Altman & Sabatos (2007) konkursprognosticeringsmodel indeholder dog alle de fem nøgletal, hvorimod kun to nøgletal implementeres i den logistiske regressionsmodel i indeværende speciale. Forskellen mellem de to modeller kan skyldes, at Altman & Sabato (2007) fokuserer på små og mellemstore virksomheder i deres undersøgelse, hvilket der ikke tages højde for i dette speciale. Ud fra ovenstående er klassifikationsnøjagtigheden for indeværende speciale bedre end Altman & Sabato (2007), hvorimod den er dårligere end Beaver (1966) og Altman (1968). Derudover er en væsentlig forskel, at den logistiske regressionsmodel i specialet ikke valideres, hvilket er tilfældet med de tre andre konkursprognosticeringsmodeller.

8.6 Kritik af konkursprognosticeringsmodeller

En generel kritik af konkursprognosticeringsmodeller er deres klassifikationsnøjagtighed, idet en klassifikationsnøjagtighed på 95 % lyder imponerende på trods af, at konkurstilstanden, der undersøges, forekommer forholdsvis sjældent. Normal er det kun ca. 1-2 % af alle aktieselskaber, der går konkurs årligt (Christensen & Nielsen, 2012). Derfor vil en konkursprognosticeringsmodel, der forudsiger, at alle virksomheder ikke går konkurs, have en mindre fejlmargin end de fleste udarbejdede modeller. Denne kritik sætter spørgsmålstegn ved, om der overhovedet bør udarbejdes konkursprognosticeringsmodeller, idet de sjældent vil have en lavere fejlmargin, end en model der antager, at ingen virksomheder går konkurs. Et andet kritikpunkt er, at hvis markedet tror på værdien af konkursprognosticeringsmodeller, må det formodes, at de anvendes til at analysere virksomheders sandsynlighed for at gå konkurs (Christensen & Nielsen, 2012). Ud fra disse analyser vil aktionærer og långivere således være tilbageholdende med at investere i virksomheder, der af modellen forudsiges til at gå konkurs. Dette medfører, at det bliver vanskeligere for disse virksomheder at skaffe kapital til overlevelse, og derfor vil risikoen for at gå konkurs stige (Christensen & Nielsen, 2012). Ud fra dette kan konkursprognosticeringsmodeller have en tendens til at blive selvopfyldende, idet virksomheder med sandsynlighed for at gå konkurs formentlig går konkurs, hvis aktionærer og långivere tager

afstand fra disse virksomheder. Dette kritikpunkt afspejler en negativ konsekvens ved udarbejdelse og anvendelse af konkursprognosticeringsmodeller, idet modellerne således ikke er med til at undgå en konkurs, men derimod fremskynder en konkurs.

I ovenstående diskussion er der fremført forskellige problemstillinger ved den udarbejdede logistiske regressionsmodel og ved at adressere disse problemstiller med de omdiskuterede forbedringsforslag, er det muligt at styrke modellen og danne baggrund for en validering. En validering af den logistiske regressionsmodel vil medføre, at modellen kan generaliseres og afprøves i praksis, hvorefter det er interessant at analysere eksterne virksomheders meninger og holdninger ift. modellen.

9 Konklusion

Specialet tog afsæt i forfatterens interesse inden for konkursprognosticering samt den begrænsede tilstedeværelse af konkursprognosticeringsmodeller udarbejdet på baggrund af danske virksomheder og betydningen heraf. Konkursprognosticering er dog ikke et nyt fænomen, da det har eksisteret i flere årtier, hvor der er blevet udviklet konkursprognosticerings-modeller med udgangspunkt i kvantitative nøgletal og ved brug af en logistisk regression, der har været tendensen i den akademiske litteratur siden 1980'erne. Dette medførte, at specialet blev funderet i et undersøgelsesproblem, idet specialet omhandlede et konceptuelt problem, hvor forfatterens intention var at øge sin viden omkring konkursprognosticering af danske virksomheder. Specialets problemformulering udgjorde dermed også undersøgelsesproblemet, og det medførte nedenstående problemformulering:

Hvordan bidrager udvalgte nøgletal via en logistisk regression til forudsigelse af konkurser i danske virksomheder?

Besvarelsen af problemformuleringen tog udgangspunkt i seks udvalgte kvantitative nøgletal, hvoraf de fem nøgletal stammede fra Altman & Sabato (2007), hvortil soliditetsgraden blev inddraget som supplement. Den statistiske analyse blev foretaget via en logistisk regression, der fulgte sekstrinsmodellen fra Hair et al. (2010). Indledningsvis blev der udarbejdet en deskriptiv statistik af stikprøven, der illustrerede, at stikprøven ikke var repræsentativ for populationen inden for flere parametre, herunder konkurser, brancher og virksomhedsformer. Derudover fremgik det af den deskriptive statistik, at de to grupper i stikprøven heller ikke var matchet.

Den statistiske analyse af stikprøven ud fra den logistiske regression med metoden forward stepwise (Wald) resulterede i, at der blev estimeret en logistisk regressionsmodel, der indeholdte to variabler. I step 1 blev variablen RE_TA (X_4) implementeret, hvorefter variablen C_TA (X_2) blev tilføjet i step 2. Derefter blev den logistiske regressionsmodel vurderet til at have et acceptabelt *goodness-of-fit* bl.a. pga. en klassifikationsnøjagtighed på 87,7 %, en Nagelkerke R^2 på over 60 % samt en Hosmer & Lemeshow test, der ikke var signifikant. Fortolkningen af den logistiske regressionsmodel illustrerede, at nøgletallet C_TA, der omhandler virksomhedens øjeblikkelige betalingsevne, havde en positiv sammenhæng med sandsynligheden for, at en virksomhed går konkurs. Derimod havde det andet nøgletal RE_TA, der omhandler virksomhedens evne til at modstå tab, en negativ sammenhæng med sandsynligheden for, at en

virksomhed går konkurs. Disse sammenhænge medførte, at en forbedring af virksomhedens øjeblikkelige betalingsevne øger sandsynligheden for at gå konkurs, hvorimod en forbedring af virksomhedens evne til at modstå tab minimerer sandsynligheden for at gå konkurs. I Trin 6 i analysen blev den logistiske regressionsmodel ikke valideret ud fra en hold-out stikprøve, da klassifikationsnøjagtigheden ikke var tilstrækkelig høj, hvilket skyldtes en foruroligende lav gruppespecifik hitrate for de konkursramte virksomheder.

Den manglende validering af den logistiske regressionsmodel medførte, at konkurs-prognosticeringsmodellen ikke kunne generaliseres til populationen, og desuden bekræftede den manglende validering, at stikprøven ikke var repræsentativ for populationen. Dette indikerede ligeledes, at der er et stort potentiale for at videreudvikle den logistiske regressionsmodel. I diskussionen blev konkrete forbedringsforslag fremført, og disse vedrørte en optimering af stikprøven, en statistisk udvælgelsesproces ift. nøgletallene, inddragelse af kvalitative nøgletal, udarbejdelse af en korrelationsanalyse samt en analyse af interaktionen mellem de udvalgte nøgletal.

Der konkluderes således, at nøgletallet C_TA bidrager med en positiv sammenhæng, hvorimod nøgletallet RE_TA bidrager med en negativ sammenhæng til forudsigelsen af konkurser. Derfor konkluderes der ligeledes, at det kun er to ud af seks udvalgte nøgletal, der bidrager til forudsigelsen af konkurser, idet de resterende fire nøgletal ikke var en del af den logistiske regressionsmodel. Den manglende validering af den logistiske regressionsmodel medfører, at der ikke kan konkluderes generelt for danske virksomheder. Svaret på undersøgelsesproblemet og problemformuleringen er dermed, at C_TA bidrager med en positiv sammenhæng og RE_TA bidrager med en negativ sammenhæng til forudsigelsen af konkurser for de virksomheder, der er en del af stikprøven.

10 Figurliste

11 Litteraturliste

Aalborg Universitet, 2016 *"SPSS software til studerende og ansatte på AAU".* Lokaliseret d. 05.04.16 på: http://www.software.aau.dk/spss/

Agresti, Alan & Finlay, Barbara, 2008 *"Statistical Methods for the Social Sciences".* Pearson, International Edition of 4. Edition

Altman, Edward I., 1968 *"Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy".* The Journal of Finance, Vol. 23, No. 4, pp. 589-609

Altman, Edward I. & Sabato, Gabriele, 2007 *" Modelling Credit Risk for SMEs: Evidence from the U.S. Market".* ABACUS, Vol. 43, No. 3, pp. 332-357

Auktioner ApS, 2016 *"Konkurser".* Lokaliseret d. 17.05.16 på: http://www.konkurser.dk/konkurser/

Bang-Pedersen, Ulrik R.; Christensen, Lasse H. & Jensen, Kim S., 2014 *"Konkurs".* Jurist- og Økonomiforbundets Forlag, 1, Udgave, 1. Oplag

Beaver, William H., 1966 *"Financial Ratios as Predictors of Failure".* Journal of Accounting Research, Vol. 4, Issue Empirical Research in Accounting: Selected Studies, pp. 71-111

Bendsen, Thomas, 2016 *"Noter i Statistik – Beregninger".* Lokaliseret d. 21.05.16 på: http://statnoter.dk/index.php?pageID=6

Booth, Wayne C.; Colomb, Gregory G. & Williams, Joseph M., 2008 *"The Craft of Research".* The University of Chicago Press, 3. Edition

Burrell, Gibson & Morgan, Gareth, 1979 *"Sociological Paradigms and Organisational Analysis – Elements of the Sociology of Corporate Life".* Ashgate Publishing Company, Part 1, pp. 1-37

Christensen, John & Nielsen, Mogens, 2012 *"Virksomhedens årsregnskab".* Syddansk Universitetsforlag, 6. Udgave, 2. Oplag

Clement, Sanne L. & Ingemann, Jan H., 2011 *"Introduktion til praktisk statistik".* Syddansk Universitetsforlag

Danmarks Domstole, 2016 *"Ordbog"*. Lokaliseret d. 10.05.16 på: http://www.domstol.dk/saadangoerdu/ordforklaring/Pages/Kon.aspx

Danmarks Statistik, 2007 *"Dansk Branchekode 2007 – DB07"*. Lokaliseret d. 20.05.16 på: http://www.dst.dk/Site/Dst/Udgivelser/GetPubFile.aspx?id=11119&sid=helepubl

Danmarks Statistik, 2016 *"Statistikbanken"*. Lokaliseret d. 19.05.16 på: http://www.statistikbanken.dk/statbank5a/default.asp?w=1280

Elling, Jens O., 2014 *" Finansiel rapportering – teori og regulering"*. Gjellerup / Gads Forlag, 3. Udgave, 2. Oplag

Erhardtsen, Birgitte; Kongskov, Jesper & Friis, Lasse, 2014 *"OW-konkursbo begravet i henvendelser fra interessenter"*. Lokaliseret d. 14.04.16 på: http://www.business.dk/transport/ow-konkursbo-begravet-i-henvendelser-fra-interessenter

Hair, Joseph F.; Black, William C.; Babin, Barry J. & Anderson, Rolph E., 2010 *"Multivariate data analysis"*. Prentice Hall, 7. Edition

Hawawini, Gabriel & Viallet, Claude, 2011 *"Finance for Executives – Managing for value creation*". South-Western Cengage Learning, 4. Edition

Heldbjerg, Grethe, 2011 *"Grøftegravning i metodisk perspektiv"*. Samfundslitteratur, 1. Udgave, 9. Oplag

IBM, 2016 *"SPSS Statistics"*. Lokaliseret d. 05.04.16 på: http://www-01.ibm.com/software/analytics/spss/products/statistics/

Intemodino Group, 2016 *"Tilfældige tal generator"*. Lokaliseret d. 19.04.16 på: http://randomnumbergenerator.intemodino.com/dk/

Jensen, Henning S.; Rimstad, Kjetil & Riikonen, Matti, 2010 *"Anbefalinger & Nøgletal 2010"*. Den Danske Finansanalytikerforening & Norges Finansanalytikers Forening & Suomen Sijoitusanalyytikot ry, Jessen & Co.

Justitsministeriet, 2014 *"Bekendtgørelse af konkursloven"*. Lokaliseret d. 13.05.16 på: https://www.retsinformation.dk/forms/r0710.aspx?id=158134

Keller, Gerald, 2011 *"Managerial Statistics"*. South-Western College Publishing, International Edition of 9. Edition

Ledernes Hovedorganisation, 1999 *"Information vedr. Betalingsstandsning og konkurs"*. K. Larsen & Søn

NN Erhverv, 2016 *"Navne & Numre Erhverv"*. Lokaliseret d. 16.02.16 på: http://erhverv.nnmarkedsdata.dk.zorac.aub.aau.dk/Content/Search/CommonSearch.aspx

Ohlson, James A., 1980 *"Financial Ratios and the Probabilistic Prediction of Bankruptcy"*. Journal of Accounting Research, Vol. 18, No. 1, pp. 109-131

ReStore, National Centre for Research Methods, 2011 *"The SPSS Logistic Regression Output"*. Lokaliseret d. 01.05.16 på: http://www.restore.ac.uk/srme/www/fac/soc/wie/research-new/srme/modules/mod4/12/index.html

Schack, Bent, 2009 *"Regnskabsanalyse og virksomhedsbedømmelse"*. Jurist- og Økonomforbundets Forlag, 4. Udgave, 2. Oplag

Schiøtz, Peter & Jakobsen, Susanne, 2008 *"Konkursprocessen"*. Jurist- og Økonomforbundets Forlag, 1. Udgave, 1. Oplag

Simonsen, Peter, 2016 *"Antallet af konkurser stiger: Disse brancher og landsdele er hårdest ramt"*. Lokaliseret d. 14.04.16 på: http://finans.dk/live/erhverv/ECE8412708/antallet-af-konkurser-stiger-disse-brancher-og-landsdele-er-haardest-ramt/?ctxref=ext

12 Bilagsliste

Bilag 1 – Opløst efter konkurs, datasættet er hentet d. 16.02.16 fra NN Erhverv:
http://erhverv.nnmarkedsdata.dk.zorac.aub.aau.dk/Content/Search/CommonSearch.aspx

Bilag 2 – Normal drift, datasættet er hentet d. 16.02.16 fra NN Erhverv:
http://erhverv.nnmarkedsdata.dk.zorac.aub.aau.dk/Content/Search/CommonSearch.aspx

Bilag 3 – Databearbejdning med nøgletalsberegninger