

Linn-Kristin Becker
Trude Nonstad Fornes

Ikke-finansiell informasjon for predikering av konkurs

Non-financial information for bankruptcy
prediction

Masteroppgave i Økonomi og administrasjon
Veileder: Ranik Raaen Wahlstrøm
Mai 2022

Linn-Kristin Becker
Trude Nonstad Fornes

Ikke-finansiell informasjon for predikering av konkurs

Non-financial information for bankruptcy prediction

Masteroppgave i Økonomi og administrasjon
Veileder: Ranik Raaen Wahlstrøm
Mai 2022

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
NTNU Handelshøyskolen

Sammendrag

Formålet med denne masteravhandlingen er å undersøke hvordan ikke-finansiell informasjon kan brukes for å predikere konkurs hos selskaper. Dette gjør vi ved å konstruere åtte ulike modeller for konkursprediksjon med variabler fra ikke-finansiell informasjon, og vurderer disse utover allerede etablerte modeller med finansielle variabler fra eksisterende litteratur. Dette gir mulighet til å undersøke merverdien de ikke-finansielle variablene skaper for konkursprediksjonsmodeller, samt hvilke ikke-finansielle variabler og kategorier som har størst betydning for prediksjonsevnen.

Våre analyser utføres på et helt nytt og unikt datasett som vi har satt sammen av ukonsoliderte årsregnskaper fra norske aksjeselskaper mellom 2006 og 2020, samt ikke-finansiell informasjon om selskapenes styre, ledelse og aksjonærer. Tidligere studier innen konkursprediksjon fokuserer i stor grad på store og børsnoterte selskaper til tross for at små og mellomstore selskaper (SMB) utgjør en sterk overvekt av verdiskapningen i privat sektor. For å utforske dette området videre avgrensner vi derfor vår studie til SMB.

Vi konkluderer med at inkludering av ikke-finansielle variabler fører til en forbedring av eksisterende konkursprediksjonsmodeller. Videre finner vi at studiens viktigste ikke-finansielle variabler for å predikere konkurs er hvorvidt daglig leder sitter i styret, gjennomsnittsalder på styrets medlemmer, samt alder på styrets leder. Samtlige av disse variablene faller inn under kategorien *egenskaper ved lederskapet*. Våre funn indikerer at ikke-finansielle variabler alene ikke er tilstrekkelig for å predikere konkurs, men at prediksjonsevnen til eksisterende modeller øker når allerede etablerte finansielle variabler kombineres med ikke-finansielle variabler.

Veileder:

Ranik Raaen Wahlstrøm

Abstract

The objective of this master's thesis is to investigate how non-financial information can be used to predict bankruptcy. We do this by constructing eight different models for bankruptcy prediction with variables based on non-financial information. These models are evaluated against benchmark models with variable sets of financial variables established in the existing literature. This provides an opportunity to examine the added value that non-financial variables provide for bankruptcy prediction models, as well as which non-financial variables and categories have the greatest impact on the predictive power.

Our analyzes are performed on a completely new and unique data set that we have compiled from unconsolidated annual accounts from Norwegian limited companies between 2006 and 2020, as well as non-financial information about the companies' boards, management and shareholders. The vast majority of previous studies of bankruptcy prediction focus on large and listed companies, despite that small and medium-sized companies (SMEs) account for a predominance of the value creation in the private sector. To explore this area further, we therefore limit our study to SMEs.

We conclude that the inclusion of non-financial variables leads to an improvement of existing bankruptcy prediction models. Furthermore, we find that the study's most important non-financial variables for predicting bankruptcy are whether the CEO is a member of the board, the average age of the board members, and the age of the board chairman. All of these variables fall into the category *board structures*. Our findings indicate that non-financial variables are not sufficient to predict bankruptcy on their own, but that they increase the predictive power of existing models when combined with established financial variables.

Supervisor:

Ranik Raaen Wahlstrøm

Forord

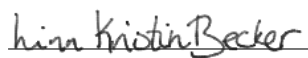
Denne avhandlingen er et avsluttende arbeid av vår mastergrad innen økonomi og administrasjon ved NTNU Handelshøyskolen. Arbeidet er utført våren 2022, og utgjør totalt 30 studiepoeng innenfor hovedprofilen Business Analytics.


Arbeidet med masteroppgaven har vært en krevende prosess, men samtidig utrolig lærerikt. Vi har fått utfordret oss selv og utviklet våre akademiske og analytiske evner. Samtidig har vi fått et dypere innblikk i et svært spennende fagfelt innen konkursprediksjon, som vår masteroppgave forhåpentligvis har gitt et godt bidrag til.

Vi ønsker å rette en stor takk til vår veileder, førsteamanuensis Ranik Raaen Wahlstrøm, for uvurderlig støtte og veiledning gjennom hele prosessen. Gjennom gode innspill og konstruktive tilbakemeldinger har dette bidratt til å løfte kvaliteten på vårt arbeid. Vi ønsker også å takke familie og venner for god støtte underveis i arbeidet.

Innholdet i denne oppgaven står for forfatternes regning.

Trondheim, mai 2022


Linn-Kristin Becker


Trude Nonstad Fornes

Innholdsfortegnelse

1	Introduksjon	1
2	Litteraturgjennomgang	4
2.1	Finansielle variabler	4
2.1.1	SEBRA-modellen	9
2.2	Markedsbaserte variabler	13
2.3	Ikke-finansielle variabler	14
2.3.1	Egenskaper ved lederskapet	15
2.3.2	Egenskaper ved eierskapet	17
2.3.3	Andre ikke-finansielle variabler	18
3	Data	20
3.1	Beskrivelse av datasett	20
3.2	Responsvariabel	20
3.3	Forprosessering	21
3.4	Ikke-finansielle variabler	24
3.5	Manglende verdier	27
3.6	Finansielle variabler	29
4	Metode	31
4.1	Balansering og estimering	32
4.2	Håndtering av manglende observasjoner	33
4.3	Håndtering av ekstremverdier	34
4.4	Oppsett av modellene	34
4.5	LASSO	37
4.5.1	Parametervalg	38
4.6	Logistisk regresjon	39
4.7	Evaluering av modellene	40

4.7.1	AUC	40
4.7.2	Brier score	40
4.7.3	AIC	41
4.7.4	Nøyaktighetsratio	41
4.8	Repliserbarhet	41
5	Resultater	42
5.1	Modell 1	42
5.1.1	Modell 1A	42
5.1.2	Modell 1B	45
5.1.3	Modellevaluering	50
5.2	Modell 2	52
5.2.1	Modell 2A	52
5.2.2	Modell 2B	53
5.2.3	Modell 2C	56
5.2.4	Modellevaluering	59
5.3	Modell 3	60
5.3.1	Modell 3A	60
5.3.2	Modell 3B	62
5.3.3	Modell 3C	65
5.3.4	Modellevaluering	68
6	Konklusjon	70
	Referanser	81
	Appendiks	82

Figurer

1	Manglende verdier for ikke-finansielle variabler 2006 - 2020	28
2	Manglende verdier for ikke-finansielle variabler 2014 - 2020	29
3	Inndeling i treningssett og valideringssett	33
4	Valg av tuningparameter λ	38
5	Illustrasjon av ROC-kurve	40
6	Sammenligning av AUC-verdier for modell 1A og 1B	51
7	Sammenligning av AUC-verdier for modell 2A, 2B og 2C	59
8	Sammenligning av AUC-verdier for modell 3A, 3B og 3C	68

Tabeller

1	Filtrering av datasett	21
2	Observasjoner per regnskapsår fordelt på konkurs og ikke-konkurs	23
3	Beskrivelse av ikke-finansielle variabler	25
4	Deskriptiv statistikk av kategoriske ikke-finansielle variabler	26
5	Deskriptiv statistikk av kontinuerlige ikke-finansielle variabler	26
6	Beskrivelse av finansielle variabler	30
7	Oversikt over variabelsett	35
8	Oversikt over modeller	36
9	Resultater modell 1A	44
10	Resultater modell 1B	46
11	Resultater modell 2A	52
12	Resultater modell 2B	54
13	Resultater modell 2C	57
14	Resultater modell 3A	61
15	Resultater modell 3B	63
16	Resultater modell 3C	66

1 Introduksjon

Dersom et selskap går konkurs vil det skape mange negative ringvirkninger, så både interne og eksterne interessenter vil forsøke å unngå dette. Pelja & Wahlstrøm (2021) trekker frem tap av egenkapital, arbeidsplasser, utlån og skatteinntekter som konsekvenser av særlig betydning ved konkurs. I tillegg til selskapet isolert sett vil også kreditorer, revisorer, aksjeeiere og toppledelsen være interessert i å unngå konkurs siden dette også er noe som påvirker dem (Ravi Kumar & Ravi, 2007). Konkursprediksjon er altså relevant for flere aktører i samfunnet.

Banker og andre kreditorer predikerer konkurs ved hjelp av statistiske modeller, og benytter prediksjonene blant annet ved vurdering av lånesøknader (Paraschiv, Schmid & Wahlstrøm, 2021). Dersom det innvilges lån til selskaper som går konkurs vil dette kunne medføre store tap. Samtidig vil det å avslå lånesøknader fra selskaper som ikke går konkurs føre til tapte markedsandeler i dagens konkurransepregede utlånsmarked. Altman, Iwanicz-Drozdowska, Laitinen & Suvas (2014) trekker i tillegg frem investorer og forvaltere som andre aktører som ser nytten av å benytte slike modeller. Dette er aktører som nøye må vurdere hvilke selskaper de ønsker i sine porteføljer, og pålitelige verktøy som kan benyttes i denne utvelgelsen står derfor sentralt. Variablene som inngår i konkursprediksjonsmodeller kan gjerne deles inn i tre kategorier; regnskapsbaserte, markedsbaserte og ikke-finansielle (Ciampi, Giannozzi, Marzi & Altman, 2021). Regnskapsbaserte variabler er gjerne ulike nøkkeltall satt sammen basert på informasjon fra regnskapet. Markedsbaserte variabler tar utgangspunkt i verdien av selskapet gjennom eksempelvis aksjeverdi. Ikke-finansielle variabler kan blant annet være sosiodemografiske forhold som alder og kjønn, eller informasjon om styre- og eierskapsstruktur (Antulov-Fantulin, Lagravinese & Resce, 2021; Liang, Tsai, Lu & Chang, 2020).

Små og mellomstore bedrifter (SMB) utgjør den største gruppen av lånekunder for europeiske og amerikanske banker, og spiller en stor rolle i verdensøkonomien (Altman & Sabato, 2007; Altman, Sabato & Wilson, 2010; Ciampi, 2015). I 2019 var 99 % av alle bedrifter i

Norge SMB med opptil 100 ansatte (Departementene, 2019). Videre utgjorde disse rundt 70 % av all verdiskapningen i privat sektor. Det er derfor svært viktig for flere aktører å kunne predikere konkurs av slike selskaper.

De fleste tidligere studier om konkursprediksjon fokuserer på større og børsnoterte selskaper, mens svært få tar for seg SMB (Paraschiv et al., 2021). Dette kan være fordi det sjeldent er mulig å lage markedsvariabler for SMB. I tillegg kan den tilgjengelige regnskapsinformasjonen fra SMB være inkonsekvent rapportert og dermed mindre pålitelig og tilgjengelig, sammenlignet med regnskapsinformasjon fra store og børsnoterte selskaper (Ciampi, 2015). Dette taler for å utvikle egne modeller for konkursprediksjon for SMB som inkluderer ikke-finansielle variabler. Dette bekreftes av Altman et al. (2010) som finner at det å inkludere ikke-finansielle variabler fører til en signifikant forbedring av konkursprediksjonsmodellene som er utviklet spesielt for SMB. Tilsvarende funn har også Ciampi (2015) i en studie av konkursprediksjon av små bedrifter hvor prediksjonsevnen blir signifikant forbedret i modellen som inneholder både regnskapstall og ikke-finansielle variabler, sammenlignet med en modell som utelukkende benytter regnskapstall. Imidlertid er det utført få studier på utvikling av modeller for konkursprediksjon av SMB som inkluderer ikke-finansielle variabler (Altman et al., 2010). I nyere tid har det vært noe mer fokus på dette, men disse modellene har fortsatt signifikante klassifikasjonsfeil, noe som taler for behovet for videre forskning (Ciampi et al., 2021). På bakgrunn av dette er følgende problemstilling formulert:

I hvilken grad kan ikke-finansielle variabler forbedre konkursprediksjon for SMB, og hvilke ikke-finansielle variabler og kategorier har størst betydning?

For å belyse denne problemstillingen har vi lagd et helt nytt og unikt datasett, med en sammensetning av data og informasjon fra flere datakilder, som så vidt oss bekjent ikke tidligere er benyttet i litteraturen om konkursprediksjon. Det nye datasettet gir mulighet til å studere i hvilken grad de ikke-finansielle variablene skaper verdi for modeller for konkursprediksjon, og hvilke ikke-finansielle variabler som har størst betydning for prediksjonsevnen til slike modeller. Vi tar utgangspunkt i alle ukonsoliderte årsregnskaper for norske aksje- og

allmennaksjeselskaper for perioden 2006 til 2020, levert av Brønnøysundregistrene. Dette utgjør i overkant av 4,3 millioner årsregnskaper. Disse sammenkobler vi med ikke-finansiell informasjon levert av Enin AS. Detaljert informasjon om alder, kjønn og lokasjon på daglig leder, styreleder og styremedlemmer, samt informasjon om aksjonærer, lederskaps- og eierskapsstruktur gir mulighet til å utvikle nye, ikke-finansielle variabler. Dette danner i så måte et unikt grunnlag for å studere bidraget ikke-finansiell informasjon har på predikeringen av konkurs for SMB.

Vi benytter *Least Absolute Shrinkage Selector Operator* (LASSO) og logistisk regresjon i våre analyser. Videre lar vi logistisk regresjon vurdere sannsynligheten for konkurs over flere år ved å trene modellene med årsregnskaper fra de samme selskapene over flere år. Dette er omtalt i litteraturen som *descrete hazard models* (Shumway, 2001), og sammenlignet med statistiske modeller som kun benytter én observasjon per selskap, tar *descrete hazard models* hensyn til at selskaper endrer seg over tid.

Studien konkluderer med at inkludering av ikke-finansielle variabler fører til en forbedring av eksisterende modeller for konkursprediksjon, og de viktigste ikke-finansielle variablene er hvorvidt daglig leder sitter i styret, gjennomsnittsalderen til styremedlemmene og alderen til styrets leder, som alle er i kategorien *egenskaper ved lederskapet*. Avslutningsvis konkluderes det med at ikke-finansielle variabler alene ikke fanger opp nok informasjon til å oppnå en akseptabel prediksjonsevne, men at inkludering av ikke-finansielle variabler øker forklaringskraften til modeller som kun benytter regnskapstall for konkursprediksjon.

Avhandlingen er inndelt i seks kapitler. I kapittel 2 vil det foreligge en litteraturgjennomgang av sentrale hendelser innenfor konkursprediksjon. I kapittel 3 følger en beskrivelse av datasettet som er benyttet til analysene. Her vil det bli gitt en beskrivelse av forprosessering samt deskriptiv statistikk av de ikke-finansielle variablene. Videre vil kapittel 4 omhandle en metodedel, hvor valg av metode blir begrunnet og diskutert. I kapittel 5 vil oppgavens funn og resultater bli fremstilt og diskutert. Avslutningsvis vil kapittel 6 inneholde en konklusjon, hvor resultater og analyser blir diskutert og oppsummert. Her vil det også bli diskutert eventuelle svakheter med arbeidet og forslag til videre forskning.

2 Litteraturgjennomgang

Konkursprediksjon har blitt omtalt i litteraturen siden 1930-tallet, hvor fokuset lå på analyser av selskapers regnskapstall for å predikere konkurs frem i tid (Smith & Winakor, 1930). Etter dette ble forskningsfokuset dreid over mot univariate modeller, noe som vedvarte frem mot midten av 1960-tallet før de multivariate modellene overtok (Gissel, Giacomino & Akers, 2007). Predikering av konkurs er også den dag i dag svært aktuelt.

2.1 Finansielle variabler

Den mest anerkjente univariate studien ble publisert av Beaver (1966) hvor ulike nøkkeltall ble benyttet til å predikere konkurs. Denne studien blir omtalt som en av de aller første og mest anerkjente studier innen nøkkeltallbasert konkursprediksjon (Gissel et al., 2007). En univariat modell kjennetegnes ved at én uavhengig variabel blir benyttet til å predikere den avhengige variabelen, i dette tilfellet hvorvidt et selskapet antas å gå konkurs eller ikke. For at et selskap skal kunne gå konkurs, må selskapet være insolvent. Et selskap ble av Beaver definert som insolvent dersom en av følgende hendelser oppstod:

- Manglende evne til å betale økonomiske forpliktelser
- Konkurs
- Obligasjonsmislighold
- Overtrukket bankkonto
- Manglende betaling av foretrukket aksjeutbytte

For sine analyser benyttet Beaver (1966) regnskapstall fra totalt 79 insolvente selskaper fordelt på 38 ulike bransjer hentet fra Moodys Industrial Manual. Selskapene hadde eierdel mellom USD 0,6 og 45 millioner, med et gjennomsnitt på omtrent USD 6 millioner. Selskapene ble videre sammenstilt og parett med 79 selskaper som ikke ble definert som insolvent, slik at hvert insolvente selskap ble sammenlignet med ett solvent selskap. For å skille ut bransje- og størrelseseffektene ble parene satt sammen av selskaper med tilnær-

met lik størrelse og bransje. Beaver (1966) tok utgangspunkt i 30 finansielle variabler som det i litteraturen var kjent at hadde en påvirkning på sannsynligheten for konkurs. Disse nøkkeltallene ble valgt ut med hensyn til at variablene i tidligere studier hadde prestert bra og at variablene var definert i en kontantstrøm. De 30 variablene ble videre redusert til følgende seks nøkkeltall:

$$\frac{\textit{Kontantstrøm}}{\textit{Samlet gjeld}}$$

$$\frac{\textit{Netto resultat}}{\textit{Sum eiendeler}}$$

$$\frac{\textit{Samlet gjeld}}{\textit{Sum eiendeler}}$$

$$\frac{\textit{Arbeidskapital}}{\textit{Sum eiendeler}}$$

$$\frac{\textit{Omløpsmidler}}{\textit{Kortsiktig gjeld}}$$

$$\textit{Likvide eiendeler} - \textit{Kortsiktig gjeld}$$

Et av hovedfunnene fra studien var at konkurs kunne predikeres i inntil fem år før konkurs fant sted. I tillegg ble forholdet mellom kontantstrøm og samlet gjeld ansett for å være nøkkeltallet med best prediksjonsevne.

Den første multivariate studien ble publisert av Altman (1968), og er en studie som fortsatt er høyt aktet (Gissel et al., 2007). I denne studien ble det utviklet en modell for konkursprediksjon omtalt som Z-score modellen. Denne benytter metoden diskriminantanalyse og er gitt ved:

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

hvor:

$$X_1 = \frac{\textit{Arbeidskapital}}{\textit{Sum eiendeler}}$$

$$X_2 = \frac{\textit{Opptjent egenkapital}}{\textit{Sum eiendeler}}$$

$$X_3 = \frac{\textit{EBIT}}{\textit{Sum eiendeler}}$$

$$X_4 = \frac{\textit{Markedsverdi egenkapital}}{\textit{Sum gjeld}}$$

$$X_5 = \frac{\textit{Salgsinntekter}}{\textit{Sum eiendeler}}$$

Altman (1968) argumenterte for at disse fem variablene dekte hovedaspektene av et selskap, og at de derfor var godt egnet for å si noe om selskapets økonomiske stilling. Variabelen X_1 , arbeidskapital som andel av sum eiendeler, var likviditetsmålet som ble funnet å være mest signifikant. X_2 , opptjent egenkapital som andel av sum eiendeler, var et lønnsomhetsmål som målte den kumulative lønnsomheten over tid, og Altman (1968) viste til at selskapets alder derfor implisitt ville bli fanget opp av denne variabelen. X_3 , inntjening før renter og skatt som andel av sum eiendeler, var et mål på verdiskapningen til selskapets eiendeler, og han viste til at dette var et særlig viktig mål for å predikere konkurs. Variabelen X_4 , markedsverdi egenkapital som andel av sum gjeld, viste hvor mye selskapets eiendeler kan synke i verdi før selskapet ble insolvent ved at gjelden oversteg verdien av eiendelene, og Altman (1968) fremhevet at denne variabelen inneholdt en markedsdimensjon som tidligere studier ikke har hensyntatt. Variabelen X_5 , salgsinntekter som andel av sum eiendeler, var et lønnsomhetsmål som viste selskapets evne til å håndtere konkurranse i markedet. Altman (1968) viste til at dette målet var minst signifikant isolert sett, men at samspillet med de andre variablene førte til at dette likevel var en av de viktigste variablene. For alle disse variablene vil lave verdier være forbundet med økt risiko for konkurs.

Til sine analyser benyttet Altman (1968) informasjon fra 66 selskaper som drev med vare-

produksjon, hvor 33 var konkurs og 33 ikke var konkurs. Selskapene som utgjorde gruppen som var konkurs varierte i størrelse. Totalkapitalen varierte fra USD 0,7 millioner for det minste, og opp til USD 25,9 millioner for det største selskapet. Type industri varierte også blant de ulike selskapene, og det var dermed ikke et homogent utvalg. Det ble derfor forsøkt å utføre et nøye utvalg for gruppen av selskaper som ikke var konkurs. Denne gruppen bestod av et paret utvalg valgt ved en stratifisert tilfeldig metode etter bransjetype og størrelse. Størrelsen for selskapene som kunne inkluderes i utvalget ble derfor begrenset til selskaper med total kapital mellom USD 1 - 25 millioner. Z-score modellen oppnådde gode resultater for prediksjon av konkurs ett år frem i tid, med en treffprosent på 95 %.

Edmister (1972) publiserte en studie for konkursprediksjon av små selskaper, hvor flere ulike metoder for analyse av nøkkeltall ble testet for sin evne til å predikere konkurs for små selskaper. Han viste til at multipl diskriminantanalyse kunne være egnet for utvelgelse av optimale nøkkeltall og metoder, og at det derfor kunne være en mer effektiv teknikk enn de subjektive prosessene som ble benyttet på den tiden. Edmister (1972) pekte på at tidligere studier slik som Altman (1968) og Beaver (1966) kan være nyttige for konkursprediksjon av store og mellomstore selskaper, men hevdet de i stor grad overså små selskaper. Dette var i hovedsak forårsaket av vanskeligheter med å anskaffe data for små selskaper.

Altman et al. (1977) argumenterte for at det var nødvendig å utvikle en ny modell blant annet fordi størrelsen på selskaper som gikk konkurs hadde økt, og de fleste tidligere studier hadde benyttet relativt små selskaper i sine analyser. De presenterte derfor ZETA-modellen i 1977, som en videreutvikling av Z-score modellen. Det ble også i denne modellen benyttet diskriminantanalyse, og datasettet besto av 53 selskaper som var konkurs og 58 som ikke var konkurs. Datasettet inneholdt observasjoner fra perioden 1969 til 1975, og gruppen med ikke-konkurs ble paret opp mot selskapene som var konkurs med hensyn på industritype og størrelse, i likhet med metoden benyttet for utviklingen av Z-score modellen. Videre var den nye ZETA-modellen tilpasset større selskaper ved at størrelsen til selskapene i utvalget som hadde gått konkurs, gitt i total kapital, var på minst USD 20 millioner og med et gjennomsnitt på USD 100 millioner (Altman et al., 1977). ZETA-modellen hadde

en signifikant forbedret treffsikkerhet for konkursprediksjon sammenlignet med tidligere modeller.

Ohlson (1980) introduserte en modell som benyttet logistisk regresjon, og som ble utviklet på et datasett fra perioden 1970 til 1976. Datasettet inkluderte 105 selskaper som var konkurs og 2058 selskaper som ikke var konkurs. Ohlson (1980) viste til at hans studie skilte seg fra andre studier som benyttet datasett fra 70-tallet ved at andre studier hadde mindre datasett og at Altman (1968) og Altman et al. (1977) benyttet en tilnærmet 50-50 fordeling av henholdsvis konkurs og ikke-konkurs. Ohlson (1980) trakk også frem at prediksjonsevnen til en modell vil avhenge av når årsregnskapet er forventet å være tilgjengelig, og at håndteringen av dette var en svakhet ved flere tidligere studier.

Altman (2000) reviderte Z-score modellen, og de nye modellene ble omtalt som Z'- og Z''-score modellene. Z'-score modellen var spesielt tilpasset selskaper som ikke var registrert på børs. Den opprinnelige Z-score modellen benyttet markedsverdi i likningen for X_4 (markedsverdi egenkapital/sum gjeld). Markedsverdien var basert på selskapets aksjekurs, og i den reviderte Z'-score modellen ble markedsverdi derfor byttet ut med bokført verdi av egenkapitalen. Z''-score modellen var i tillegg til å være tilpasset selskaper som ikke var på børs, også tilpasset selskaper som ikke drev med vareproduksjon. Dette ble gjort ved at både den nye X_4 -variabelen fra Z'-score modellen ble benyttet og at variabelen X_5 (salgsinntekter/sum eiendeler) ble tatt ut av modellen. Altman (2000) argumenterte for at variabelen X_5 var sensitiv for bransjetype, da kapitalens omløpshastighet (salgsinntekter/sum eiendeler) var inkludert, og at dette var noe som varierte mye fra bransje til bransje.

Altman & Sabato (2007) publiserte en studie som bygde videre på studien til Edmister (1972), hvor de utvidet og forbedret hans tidligere arbeid. I studien sammenlignet de en modell som var utviklet spesifikt for å predikere konkurs for SMB med Z''-score modellen, som var en generell modell på tvers av selskapsstørrelse. Det ble benyttet paneldata fra perioden 1994 til 2002 for 2010 amerikanske selskaper, hvor 120 var konkurs og 1890 var

ikke-konkurs. Modellen benyttet følgende fem variabler for konkursprediksjon:

$$\frac{\textit{Kortsiktig gjeld}}{\textit{Bokført verdi egenkapital}}$$

$$\frac{\textit{Kontantbeholdning}}{\textit{Totale eiendeler}}$$

$$\frac{\textit{EBITDA}}{\textit{Totale eiendeler}}$$

$$\frac{\textit{Opptjent egenkapital}}{\textit{Totale eiendeler}}$$

$$\frac{\textit{EBITDA}}{\textit{Rentekostnader}}$$

Det ble benyttet logistisk regresjon, og det ble estimert en modell med variablene som beskrevet over, i tillegg til en modell hvor det ble benyttet de logaritmiske transformasjonene av variablene. Resultatene ble sammenlignet med resultater fra estimering med Z'' -score modellen. De fant at modellen som var spesielt utviklet for SMB som benyttet de logaritmisk transformerte variablene hadde den beste prediksjonsevnen av de tre modellene. Funnene fra studien talte for at det derfor burde skilles mellom SMB og store selskaper ved konkursprediksjon (Altman & Sabato, 2007).

2.1.1 SEBRA-modellen

Også i Norge har man sett verdien av å utvikle modeller for å predikere mislighold og konkurs. I 1987 opprettet Norges Bank en database med årsregnskaper for selskaper som hadde lån i Norges Bank. Databasen fikk navnet System for EDB-Basert Regnskapsanalyse (SEBRA) og inneholdt regnskapsopplysninger fra og med år 1981. Formålet med denne innsamlingen var å få en oversikt over tilgjengelige data om norske selskaper og om kvaliteten var av tilstrekkelig grad til å senere kunne bli benyttet til statistikk og analyse (Smogeli, 1987). Basert på denne databasen ble det senere utviklet en modell til bruk i analyser

av kredittrisiko i foretakssektoren (Eklund, Larsen & Bernhardsen, 2001). Modellen ble utviklet for Norges Bank, og omtalt som SEBRA-modellen. Modellen baserte seg på regnskapstall fra SEBRA-databasen til å anslå sannsynligheten for konkurs basert på utvalgte nøkkeltall (Eklund et al., 2001). Ved implementering av modellen besto databasen av omtrent 400 000 selskaper med tilhørende regnskapstall for perioden 1990 - 1996. Selskapenes totale risiko ble estimert gjennom aggregering av de individuelle konkurssannsynlighetene basert på inntjening, likviditet og soliditet.

Den opprinnelige SEBRA-modellen har blitt videreutviklet flere ganger til versjoner med alternative forklaringsvariabler. I 2007 ble den opprinnelige SEBRA-modellen utvidet til to nye versjoner, SEBRA-basis og SEBRA-utvidet (Bernhardsen & Larsen, 2007). Årsaken til disse utvidelsene var flere svakheter i den opprinnelige modellen. Det faktum at det var kommet nye regnskapsregler, endringer i konkursregistreringen samt tilgang på ny og mer data var andre momenter som ble trukket frem og underbygget behovet for reviderte modeller. I SEBRA-basis ble variablene for inntjening, soliditet og likviditet tatt med videre fra den opprinnelige SEBRA-modellen. I tillegg ble nye bransjevariabler basert på variablene for inntjening og soliditet inkludert. SEBRA-utvidet inneholdt de samme variablene som SEBRA-basis, men fikk i tillegg inkludert variabler for størrelse, leverandørgjeld og ubetalte offentlige avgifter (Bernhardsen & Larsen, 2007), som vist i appendiks A. Variablene som inngår i SEBRA-utvidet vil bli presentert i det følgende (Bernhardsen & Larsen, 2007).

Inntjening

At selskapet har en inntjening av vesentlig grad er helt essensielt. Det er viktig at inntektene står i et rimelig forhold til selskapets betalingsforpliktelser (Eklund et al., 2001). Dersom inntjeningen ikke dekker gjeld og forpliktelser vil dette gå ut over selskapets likviditet, og risikoen for konkurs øker. Inntjeningsvariabelen blir i SEBRA-utvidet definert som:

- *Ordinært resultat før av- og nedskrivninger etter skatt i prosent av langsiktig gjeld*

Likviditet

Mangel på likviditet er ofte en utløsende faktor for konkurs, og er følgelig en sentral variabel i en konkursprediksjonsmodell. Eklund et al. (2001) viser til at reduserte betalingsmidler er et tegn på manglende likviditet. Videre kan det være et tegn på svak likviditet dersom andelen leverandørgjeld eller skyldige offentlige avgifter i forhold til total kapital øker, og er derfor forbundet med økt risiko (Eklund et al., 2001). Den utvidede SEBRA-modellen inkluderer følgende tre variabler som beskriver et selskaps likviditet:

- *Likvider minus kortsiktig gjeld i prosent av omsetning*
- *Skyldige offentlige avgifter i prosent av total kapital*
- *Leverandørgjeld i prosent av total kapital*

Soliditet

Soliditet er et mål på selskapets evne til å tåle tap, og viser hvor stor del av selskapets eiendeler som er finansiert ved hjelp av egenkapital (Eklund et al., 2001). Selskaper som innehar en høyere egenkapitalandel vil være bedre rustet og risikoen for konkurs blir lavere. Nyetablerte selskaper med lav tilgang på kapital vil således være mer sårbare og utsatt for konkurs (Hoff & Helbæk, 2021, s. 318). Eklund et al. (2001) viser til at dersom et selskap har bokført egenkapital som er mindre enn innskutt egenkapital er dette et tegn på at selskapet har et bokført tap og indikerer dermed økt risiko. De trekker frem følgende variabler som kan gi et inntrykk av selskapets soliditet:

- *Egenkapital som andel av total kapital*
- *Bokført egenkapital mindre enn innskutt egenkapital (dummyvariabel)*

Alder

Denne variabelen angir alderen til selskapet, oppgitt i antall år siden etablering. Sannsynligheten for at et selskap overlever øker med selskapets alder (Evans, 1987). Dette samsvarer med resultatene til Eklund et al. (2001) som finner at konkurshyppigheten er større blant

nyetablerte selskap enn selskap som er godt etablert i markedet. Dette begrunnes i studien med at unge selskaper kan ha utfordringer med å skaffe nødvendig kapital, samt at det gjerne krever tid å bygge opp relevant og nødvendig kompetanse. I SEBRA-utvidet blir aldersvariabelen behandlet som en dummyvariabel, som vist under:

- *Alder (år) = 1, 2, 3, ..., 8 (dummyvariabel)*

Størrelse

Generelt er hyppigheten av konkurs større blant små selskap enn store selskap (Eklund et al., 2001). Dette kan underbygges av Evans (1987) som i sin studie finner at sannsynligheten for å overleve øker med selskapets størrelse. Eklund et al. (2001) begrunner dette med at små selskap gjerne opererer innenfor et mer begrenset geografisk område og at produktdiversifisiteten er lav. I tillegg er gjerne små selskaper relativt unge, og momentene under aldersvariabelen vil også være relevant her. Størrelsesvariabelen er definert som følgende i SEBRA-utvidet:

- *Sum eiendeler i faste kroner*

SEBRA-modellen har lenge vært den foretrukne modellen for å anslå sannsynligheter for konkurs i ikke-finansielle selskap av blant annet Norges Bank og Finanstilsynet (Bernhardsen & Larsen, 2007). Norges Bank har imidlertid siden 2018 benyttet KOSMO konkurs-sannsynlighetsmodell som tar inn både makroøkonomiske indikatorer og kredittvurderinger av selskapene i tillegg til regnskapsinformasjon (Hjelseth, Arbatli, Solheim & Vatne, 2020). Denne modellen er fortsatt under utvikling, og er foreløpig kun benyttet for et utvalg store banker i Norge. Fra og med 2021 har Finanstilsynet benyttet en ny misligholds- og konkurs-sannsynlighetsmodell for å vurdere sannsynligheten for konkurs, kalt SEMKO-modellen (Finanstilsynet, 2021). Utover at modellen er en videreføring av SEBRA-modellen som predikerer sannsynlighet basert på årsregnskap og annen selskapsspesifikk informasjon, er det foreløpig ikke offentlig kjent hvilke variabler som inngår i denne modellen.

2.2 Markedsbaserte variabler

Inkludering av markedsvariabler i modeller har vist seg å øke prediksjonsevnen for konkurs betraktelig, noe som har ført til at de tradisjonelle modellene som kun benytter seg av regnskapsvariabler har fått konkurranse (Beaver, McNichols & Rhie, 2005). Hillegneis et al. (2004) sammenlignet i sin studie rene markedsbaserte modeller med regnskapsbaserte modeller, og fant at markedsbaserte modeller er overlegne hva gjelder treffsikkerhet for konkurs. Dette ble også understøttet av en rekke andre studier, se eksempelvis Shumway (2001) og Chava & Jarrow (2004). Sistnevnte studie fant i tillegg at regnskapsvariabler tilføyer modellene svært lite dersom markedsvariabler allerede er inkludert i modellen.

Agarwal & Taffler (2008) trakk frem flere forhold som gjør at bruk av kun regnskapsvariabler kan være problematisk. Innledningsvis presenterer regnskap historiske resultater, og er dermed ikke nødvendigvis informative hva gjelder fremtiden. Videre vil bruk av regnskapsprinsipper som historisk kost føre til at eiendeler ofte har annen verdi enn hva som er oppgitt i regnskapene. De pekte videre på at regnskap i hovedsak er utarbeidet på grunnlag av fortsatt drift og dermed ikke for å forutsi konkurs, samt at regnskap ofte er gjenstand for manipulasjon. Bruk av markedsvariabler vil på mange måter bøte med den fremlagte kritikken, gjennom at aksjekurs blir benyttet som proxy for selskapenes verdi (Shumway, 2001). Dette vil gjenspeile informasjon som er tilgjengelig i regnskapet, samtidig som at verdiene ikke blir påvirket. Det vil også reflektere i større grad fremtidige kontantstrømmer, og er ikke avhengig av tid eller utvalg (Agarwal & Taffler, 2008).

En implikasjon hva gjelder modeller som baserer seg på eller inkluderer markedsbaserte variabler er at selskapene må være registrert på børs (Berg, 2007). Hovedmajoriteten av norske selskaper faller inn under kategorien SMB, og disse selskapene er på grunn av deres størrelse typisk ikke børsregistrerte. Dette fører til at det ikke vil være mulig å benytte markedsbaserte variabler for slike selskaper. Ciampi (2015) trekker imidlertid i sin studie frem viktigheten av nettopp å inkludere andre variabler enn kun regnskapsvariabler i modeller for konkursprediksjon.

2.3 Ikke-finansielle variabler

I en litteraturstudie om forskning på konkursprediksjon for SMB viste Ciampi, Giannozzi, Marzi & Altman (2021) til at ikke-finansielle variabler for konkursprediksjon er et forskningsområde som med fordel kan utforskes nærmere. De trakk spesielt frem muligheten til å se nærmere på flere kvalitative variabler som omhandler ledelsesegenskaper og intellektuell kapital, og implementere disse i modeller for prediksjon av konkurs for SMB (Ciampi et al., 2021).

Altman et al. (2010) bygde videre på modellen fra Altman & Sabato (2007) for SMB bestående av opplysninger fra selskaper i Storbritannia for perioden 2000 - 2007. Datasettet inneholdt over 5,8 millioner poster med regnskapstall. Andelen insolvente selskaper i datasettet var 1,2 % og utgjorde 66 833 selskaper, og det ble benyttet logistisk regresjon for estimeringen. I denne studien inkluderte de ikke-finansielle variabler som størrelse og selskapets alder, og fant at prediksjonsevnen ble forbedret med opptil 13 %. Ciampi (2015) benyttet i likhet med Altman & Sabato (2010) logistisk regresjon, og i Ciampis studie ble det utført konkursprediksjon av 934 små selskaper i Italia. I modellen ble det inkludert flere ikke-finansielle variabler, og funnene tydet på at flere av disse var signifikante. Ciampi (2015) fant videre at det å inkludere ikke-finansielle variabler i tillegg til finansielle variabler førte til en signifikant forbedring av prediksjonsevnen, sammenlignet med prediksjon basert kun på finansielle variabler. Ciampi viste også til at de fleste tidligere publiserte studier av ikke-finansielle variabler for konkursprediksjon omhandlet mellomstore og store selskap, i likhet med Grunert, Norden & Weber (2005). I studien av Grunert et al. (2005) ble det benyttet data fra tyske selskaper som hadde en årlig omsetning på mellom EUR 25 og 250 millioner, og det ble sett på hvilke effekter det å inkludere ikke-finansiell informasjon ville tilføre kredittvurderinger. Funnene deres tydet på at det å kombinere finansiell og ikke-finansiell informasjon førte til en signifikant forbedring av prediksjonsevnen, sammenlignet med å kun benytte finansielle eller kun ikke-finansielle variabler. Denne konklusjonen ble forsterket av Brèdart et al. (2021) som fant at bruk av variabler som inneholdt informasjon om menneskelige ressurser forbedret prediksjonsevnen til modellene signifikant, i tillegg til

at andelen av type I-feil ble redusert. Liang (2020) grupperte de ikke-finansielle variablene i følgende tre kategorier; egenskaper ved lederskapet, egenskaper ved eierskapet og andre ikke-finansielle variabler.

2.3.1 Egenskaper ved lederskapet

Flere studier har pekt på at ikke-finansielle variabler som omhandler selskapets ledelse og styre har en betydning for risikoen for konkurs. Den kvantitative størrelsen på styret har i denne sammenheng blitt ansett som en viktig faktor (Manzaneque, Priego & Merino, 2016). Studien gjort av børsnoterte selskaper i Spania fant at det var en sammenheng mellom størrelsen på styret og sannsynligheten for økonomisk mislighold. Funnene pekte på at et styre bestående av flere deltakere vil inneha et større mangfold og økt tilgang på informasjon, noe som kan øke evnen til å kontrollere ledelsen i riktig retning. Dette momentet ble også støttet av tidligere forskning på området som pekte på at flere medlemmer i et styre vil ha positiv påvirkning på effektivitet og uavhengighet (Dalton, Daily, Johnson & Ellstrand, 1999; Pearce & Zahra, 1992). Guest (2009) viste til at informasjonsflyt og effektivitet i beslutningsprosesser innad i styret imidlertid kan svekkes dersom størrelsen på styret blir alt for stort.

I tillegg til størrelsen på styret, pekte Elloumi et al. (2001) på at selve sammensetningen av styret er av interesse for konkursprediksjon og kan forklare økonomisk mislighold. Antulov-Fantulin et al. (2021) utførte en studie på italienske kommuner, og fant at de sosialdemografiske egenskapene kjønn og gjennomsnittsalder på medlemmene i kommunestyret var blant de ti viktigste ikke-finansielle variablene for å predikere mislighold i kommunal sektor. Garcia & Herrero (2021) utførte i denne sammenheng en studie blant europeiske selskaper i perioden 2002 til 2019, og fant at sannsynligheten for konkurs reduseres dersom det er en høy andel kvinnelige styremedlemmer. Dette funnet ble også støttet av en studie gjort på kinesiske selskaper i perioden 2005 til 2016 som fremhevet at kjønns mangfold i ledelsen var en viktig faktor når det kom til konkurserisiko (Cho, Okafor, Ujah & Zhang, 2021). Når det gjaldt rollen som daglig leder, viste Platt & Platt (2012) til at funnene i

deres studie pekte mot at selskaper med yngre daglig leder oftere går konkurs enn selskaper med en eldre daglig leder. Hvorvidt daglig leders kjønn har betydning ble av Escribano-Nevas & Gemar (2021) funnet å være signifikant i en studie om kjønn og konkurs innen hotellbransjen i Spania. Det ble her funnet at selskaper med kvinnelige daglige ledere hadde mindre sjanse for å gå konkurs, sammenlignet med om daglig leder var en mann.

I enkelte selskaper er daglig leder og styrets leder samme person. Dette scenarioet omtales i litteraturen som CEO-dualitet, og har vært et omdiskutert tema og problemstilling innen næringsliv i lengre tid (Finkelstein & D'Aveni, 1994). Jensen (1993) pekte på at det er nødvendig med uavhengighet mellom ledelsen og styret for å kunne kontrollere driften på en forsvarlig måte og dermed unngå negative resultateffekter. Andre studier har imidlertid funnet at CEO-dualitet kan føre til et mer fleksibelt lederskap som bedrer effektiviteten i selskapet og har talt for at CEO-dualitet er en styrke for selskapet (Ciampi, 2015; Dowell, Shackell & Stuart, 2011; Lilienfeld-Toal & Ruenzi, 2014). Det har blitt utført flere studier på CEO-dualitet i forbindelse med konkursprediksjon, men de empiriske bevisene har ikke vært entydige (Mangena & Chamisa, 2008). Manzaneque et al. (2016) utførte en studie på spanske børsnoterte selskaper og forkastet sin nullhypotese om at CEO-dualitet førte til økt sannsynlighet for økonomisk mislighold. Regresjonskoeffisientene ga positive fortegn, men de viste til at variabelen ikke kunne regnes som signifikant for konkursprediksjon. Ciampi (2015) fant på sin side at CEO-dualitet var blant en av de ikke-finansielle variablene som var signifikant og som reduserte risikoen for mislighold gjennom sin studie på italienske små selskaper. Studien belyste videre at funnene tydet på at forholdet mellom selskapsstyring og konkurs var annerledes for små selskaper enn for mellomstore og store selskaper.

Duru, Lyengar og Zampelli (2016) analyserte amerikanske selskaper og fant at hvorvidt CEO-dualitet var positivt eller negativt for et selskap, hadde sammenheng med antall eksterne styremedlemmer. I tilfeller hvor andelen uavhengige styremedlemmer utgjorde en liten del av styret, fant de at CEO-dualitet har sterkere og større negative innvirkninger på selskapets driftsresultater. Dersom styret bestod av mange eksterne styremedlemmer fant de imidlertid at det kan være positivt at styrets leder og daglig leder var samme

person. Dette ble begrunnet blant annet med at lederskapet blir sterkt og entydig mot et felles mål (Lilienfeld-Toal & Ruenzi, 2014). Ciampi (2015) viste til at andelen eksterne styremedlemmer likevel ikke bør overstige 50 %. Dersom andelen holdt seg under dette nivået var dette en ikke-finansiell variabel som ble funnet signifikant og som reduserte sannsynligheten for mislighold blant små selskaper. Platt & Platt (2012) fant i sin studie at CEO-dualitet ikke var av betydning, men at andelen eksterne styremedlemmer og styrets størrelse hadde en positiv effekt på selskapenes økonomiske forhold.

I de tilfellene hvor styreleder og daglig leder ikke er samme person, fungerer styreleder som en sentral kommunikasjonskanal mellom styret og daglig leder, og innehar en kritisk lederposisjon (Krause, Semadeni & Withers, 2016). Styrelederen vil også i mange tilfeller ha en overvåkningsfunksjon gjennom å kontrollere og evaluere selskapets ledelse. Dette viser hvilken kritisk posisjon styrelederen i et selskap har, og er begrunnelsen for at styreleder direkte kan påvirke selskapets økonomiske resultater (Withers & Fitzg, 2017). Til tross for den viktige rollen en styreleder har, er hovedtyngden av forskningen på området imidlertid konsentrert rundt effektene av CEO-dualitet (Krause, Semadeni & Cannella, 2014). Hvorvidt styreleders personlige egenskaper som eksempelvis alder eller kjønn er av betydning for selskapenes økonomiske situasjon er mindre diskutert i litteraturen.

2.3.2 Egenskaper ved eierskapet

Liang et al. (2020) inkluderte 21 indikatorer for eierskap og ledelse sammen med Altman (1968) sine fem finansielle indikatorer fra Z-score modellen, og fant at det ikke førte til en signifikant forbedring av prediksjonsevnen sammenlignet med Z-score modellen. De testet deretter ulike metoder for utvalgelse av variabler, og fant at de oppnådde bedre prediksjonsevne ved å benytte stegvis diskriminantanalyse for utvalgelse av seks finansielle variabler og seks indikatorer for eierskap og ledelse blant henholdsvis 40 og 21 tilgjengelige indikatorer. Tang, Li, Tan & Shi (2020) fant at en variabel for andelen aksjer eid totalt av de ti største aksjonærene hadde stor betydning ved predikering av konkurs, trolig grunnet at en fordeling av selskapets aksjer på få aksjonærer ofte vil medføre at hver enkelt har mer

å tape ved en eventuell konkurs. Platt & Platt (2012) fant at eierandelen blant eksterne aksjonærer var høyere for selskaper som gikk konkurs, sammenlignet med selskaper som ikke gikk konkurs. Liang et al. (2016) fant at eierskapsstruktur, sammen med strukturen på styret, var de to kategoriene av ikke-finansielle variabler som hadde størst betydning ved predikering av konkurs. Kategorien for eierskapsstruktur inneholdt blant annet variabler som eierandel i styret, eierandel i ledelsen og eierandel for hovedaksjonær.

2.3.3 Andre ikke-finansielle variabler

Hvorvidt et selskap har plikt til å benytte ekstern revisor avhenger av selskapets størrelse (Aksjeloven, 1999). Det følger av tilhørende forskrift at generalforsamlingen har myndighet til å velge bort revisjon dersom selskapet har driftsinntekter under NOK 6 millioner, sum eiendeler er under NOK 23 millioner, og antall ansatte ikke utgjør mer enn ti årsverk i gjennomsnitt (Forskrift om terskelverdier, 2018). SMB vil på bakgrunn av dette ofte være i en situasjon hvor fravalg av revisjon kan være en aktuell problemstilling. Altman et al. (2010) fant at dersom et selskap ble revidert, økte sannsynligheten for konkurs. De pekte på at dette kan skyldes at en revisor trolig vil gripe inn dersom et selskap er teknisk insolvent, slik at insolvente selskaper forhindres fra å fortsette å handle. Verdien av å inkludere revisjonsvariabler ble også trukket frem av Cenciarelli et al. (2018) som i sin studie så på forholdet mellom sannsynligheten for konkurs og bistand fra ekstern revisor. Funnene deres pekte imidlertid på at sannsynligheten for konkurs ble redusert dersom selskapet benyttet en revisor med ekspertise innenfor bransjen selskapet opererte i, og forklarte dette ved at en revisor aktivt vil kunne fange opp tidlige signaler på økonomisk nød slik at ledelsen kan iverksette nødvendige tiltak og endringer i tide. Imidlertid fant de at dersom selskapet hadde store utgifter til revisjon, økte dette sannsynligheten for konkurs. Inkludering av revisorvariabler i konkursprediksjonsmodeller ble av studien funnet å øke prediksjonsevnen i slike modeller ved blant annet å redusere feilklassifiseringen av antall konkursselskaper som ikke-konkursselskaper.

Det er selskapets organisasjonsform og størrelse som avgjør om det foreligger regnskapsplikt (Regnskapsloven, 1999). Det følger av den aktuelle loven at alle aksjeselskaper uavhengig av størrelse faller inn under denne plikten jf. § 1-2. Med regnskapsplikt følger en plikt til å levere årsregnskap som minst inneholder resultat, balanse og noter for foregående regnskapsår. Regnskapsførerloven sier imidlertid ingenting om hvem som skal føre regnskapet, og det vil derfor være opp til det enkelte selskap om de ønsker å føre regnskapet selv eller å benytte en ekstern regnskapsfører for bistand og utarbeidelse av regnskapet. I litteraturgjennomgangen ble det ikke funnet studier innen konkursprediksjon som sier noe om sammenhengen mellom å benytte en autorisert regnskapsfører og sannsynligheten for konkurs.

3 Data

3.1 Beskrivelse av datasett

Studien tar utgangspunkt i et datasett bestående av alle ukonsoliderte årsregnskaper for norske aksje- og allmennaksjeselskaper i perioden 2006 - 2020, i tillegg til ukonsoliderte årsregnskaper som er innrapportert til norske myndigheter fra andre organisasjonsformer i samme tidsperiode (Wahlstrøm, 2022). Totalt består dette datasettet av i overkant av 4,3 millioner observasjoner. Hver observasjon er et avlevert årsregnskap, og inneholder regnskapsinformasjon og annen tilgjengelig informasjon om årsregnskapene fordelt på 296 forklaringsvariabler. Innholdet i dette datasettet er levert av Brønnøysundregistrene.

I denne studien skal verdien av å tilføye ikke-finansiell informasjon undersøkes. Det ble derfor nødvendig å innhente ikke-finansiell informasjon i tilknytning til årsregnskapene vi hadde tilgjengelig i det beskrevne datasettet. Datasettet som inneholder regnskapstall er derfor supplert med ikke-finansielle variabler som vi har lagd basert på informasjon om selskapenes styre, ledelse, aksjonærer og forretningsadresse levert av Enin AS. For denne studien vil de ikke-finansielle variablene bli gruppert i følgende tre kategorier; egenskaper ved lederskapet, egenskaper ved eierskapet og andre ikke-finansielle variabler. Denne inndelingen er basert på grunnlag av grupperinger som er benyttet i tidligere studier (Liang et al., 2020), og inspirasjonen for variablene er hentet fra litteraturgjennomgangen. Totalt inkluderer vi 25 ikke-finansielle variabler. En stor del av bidraget til denne masteroppgaven er konstrueringen av dette datasettet gjennom en møysommelig prosess. Resultatet av denne sammenkoblingen vi har gjennomført og bearbeidet gir et unikt datasett bestående av finansiell og ikke-finansiell informasjon om SMB i Norge.

3.2 Responsvariabel

Dersom et selskap ikke klarer å oppfylle sine økonomiske forpliktelser og betale tilbake sin utestående gjeld, har styret plikt til å vurdere om det foreligger konkurs (Altinn, 2020). Om et selskap er insolvent kan konkurs begjæres av selskapet selv, eller av fordringshaverne

(Konkursloven, 1986). Konkursloven § 61 gir følgende definisjon på insolvens: *Skyldneren er insolvent når denne ikke kan oppfylle sine forpliktelser etter hvert som de forfaller, medmindre betalingsudyktigheten må antas å være forbigående. Insolvens foreligger likevel ikke når skyldnerens eiendeler og inntekter til sammen antas å kunne gi full dekning for skyldnerens forpliktelser, selv om oppfyllelsen av forpliktelsene vil bli forsinket ved at dekning må søkes ved salg av eiendelene.*

Responsvariabelen som benyttes i denne oppgaven er *bankrupt_fs*, som angir om et selskap er konkurs eller ikke. Årsregnskapene i dette datasettet blir klassifisert som konkurs ved dummyvariabelen *konkurs_fs* som får verdien 1 dersom årsregnskapet er det siste årsregnskapet til selskapet, og selskapet er begjært konkurs i henhold til variabelen *konkursdato* (Wahlstrøm, 2022). Variabelen *konkursdato* angir datoen for konkursbegjæringen, enten begjært av selskapet selv eller av kreditorer. Manglende verdier for denne variabelen vil bety at det ikke er begjært konkurs.

3.3 Forprosessering

Det er blitt utført forprosessering av datasettet gjennom filtrering for å komme frem til et troverdig og pålitelig datasett. Tabell 1 viser hvordan antall observasjoner endrer seg gjennom de ulike kriteriene for filtrering. Tabellen viser også forholdet mellom konkurs og ikke-konkurs gjennom de ulike stegene fra opprinnelig datasett til endelig datasett.

Tabell 1: Filtrering av datasett

Tabellen viser filtreringsprosessen fra rådatasett til endelig datasett. For hver rad er det utført filtrering, og det oppgis antall observasjoner etter at filtreringen er gjennomført. Hver rad belyser også fordelingen av konkurs og ikke-konkurs.

Filtreringssteg	Antall observasjoner	Antall konkurs	Antall ikke-konkurs
Komplett datasett	4 248 493	51 127	4 197 366
Aksjeselskap SMB	3 699 633	49 214	3 650 419
Ekskludering av manglende data	3 699 626	49 214	3 650 412
Eiendeler <500 000	2 828 560	27 509	2 801 051
Ekskludering av bransjer	1 844 738	24 233	1 820 505

I denne studien vil det bli analysert data for SMB i Norge. SMB er et begrep som er gjenstand for flere ulike definisjoner. Den norske regjeringen har ved flere anledninger benyttet avgrensningen av selskaper med inntil 100 ansatte i sine beskrivelser av SMB (Depar-

telementene, 2019; Nærings - og handelsdepartementet, 2012). I EUs anbefaling 2003/361 som ble gitt av EU-kommisjonen defineres SMB som selskaper som har opptil 250 ansatte. Dersom informasjon om antall ansatte ikke er tilgjengelig sier definisjonen at SMB er selskaper som har en omsetning på maksimalt EUR 50 millioner eller med totale eiendeler som ikke overstiger EUR 43 millioner (European Commission, 2003). I datagrunnlaget som er tilgjengelig for denne studien er det mangelfull informasjon om antall ansatte, og grensene for omsetning og totale eiendeler vil derfor benyttes. Denne definisjonen er også anvendt i flere andre studier som omhandler lignende tema (Paraschiv et al., 2021; Schalck & Yankol-Schalck, 2021).

Alle selskapsformer utover aksjeselskaper er fjernet fra analysen. Dette er begrunnet med at tidligere forskning på området i hovedsak har vært rettet mot konkursprediksjon av større og børsnoterte selskaper (Paraschiv et al., 2021). I tillegg har nyere forskning argumentert for at det bør utvikles egne modeller for konkursprediksjon av SMB, gjerne inkludert ikke-finansielle variabler, siden markedsvariabler ofte er umulig å oppdrive for SMB (Altman et al., 2010; Ciampi, 2015; Ciampi et al., 2021).

Videre ekskluderer vi sju årsregnskaper fra analysen da disse observasjonene hadde utelukkende manglende verdier for alle forklaringsvariabler. I tillegg ekskluderes svært små selskaper fra datasettet ved å utelate selskaper med totale eiendeler under NOK 500 000. Dette er i tråd med tidligere studier på området, og er også benyttet i SEBRA-modellen (Bernhardsen & Larsen, 2007; Paraschiv et al., 2021).

Innenfor litteraturen for konkursprediksjon er det utbredt praksis å ekskludere bransjene finans og forsikringsvirksomhet, kraft og gassforsyning, vann, kloakk og avfallshåndtering, samt eiendomsvirksomhet (Mansi, Maxwell & Zhang, 2012). I denne studien vil derfor næringskodene D, E, K og L fra Standard for næringsgruppering (SN 2007) ekskluderes (Statistisk sentralbyrå, 2009b). Vi følger Sæther og Larsen (1999) og ekskluderer også holdingselskaper, næringskode 0. Videre vil selskaper som ikke har innrapportert bransjekode og offentlig sektor, næringskode O ekskluderes, slik som for eksempel i Paraschiv et al. (2021).

Etter fullstendig filtrering er gjennomført består datasettet av til sammen 1 844 738 årsregnskaper, fordelt på 257 590 unike selskaper. Antall konkurs og ikke-konkurs er henholdsvis 24 233 og 1 820 505, noe som gir en konkursfrekvens på 1,31 %. Til sammenligning bestod rådatasettet av 51 127 konkurs og 4 197 366 ikke-konkurs, noe som gir en andel av konkurser på 1,20 %. Dette vitner om at filtreringen har ekskludert forholdsvis lik andel av hver kategori i prosessen til endelig datasett. Tabell 2 viser fordelingen av antall observasjoner per regnskapsår, og hvor mange observasjoner som er konkurs og ikke-konkurs etter at filtreringen er gjennomført.

Tabell 2: Observasjoner per regnskapsår fordelt på konkurs og ikke-konkurs

Tabellen viser fordelingen av observasjoner basert på regnskapsår fordelt på konkurs og ikke-konkurs. Det presiseres at tabellen er laget etter filtreringsprosessen beskrevet i Tabell 1 er gjennomført. Kolonnene totalt, konkurs og ikke-konkurs angir antall observasjoner i de aktuelle regnskapsårene. Kolonnen frekvens angir andelen konkurser for hvert regnskapsår.

Regnskapsår	Totalt	Konkurs	Ikke-konkurs	Frekvens
2006	97 536	1 402	96 134	1,44 %
2007	103 985	2 205	101 780	2,12 %
2008	106 812	1 941	104 871	1,82 %
2009	107 371	1 783	105 588	1,66 %
2010	108 846	1 556	107 290	1,43 %
2011	111 607	1 619	109 988	1,45 %
2012	116 211	1 688	114 523	1,45 %
2013	120 621	1 653	118 968	1,37 %
2014	124 944	1 600	123 344	1,28 %
2015	129 323	1 507	127 816	1,17 %
2016	133 858	1 725	132 133	1,29 %
2017	139 068	1 754	137 314	1,26 %
2018	143 802	2 018	141 784	1,40 %
2019	147 920	1 209	146 711	0,82 %
2020	152 834	573	152 261	0,37 %
Totalt	1 844 738	24 233	1 820 505	1,31 %

For dette datasettet er gjennomsnittlig konkursfrekvens 1,31 %, men det er noen av regnskapsårene i datasettet som skiller seg ut. Dette gjelder særlig for regnskapsåret 2020, hvor konkursfrekvensen kun er 0,37 %. Dette skyldes i hovedsak at klassifiseringen av konkurs bygger på konkurserklæringer til og med mai 2021, slik at årsregnskaper til selskaper som egentlig skal klassifiseres som konkurs, ikke er klassifisert til dette fordi selskapene ikke var slått konkurs innen mai 2021. Videre kan den store nedgangen trolig også skyldes flere forhold tilknyttet koronapandemien. Dette gjelder blant annet kompensasjonsordningen for næringslivet og forsinket saksbehandlingstid i rettsinstansene, som begge er momenter som kan ha bidratt til å utsette konkurs for enkelte bedrifter som under normale omstendigheter

ville ha gått konkurs (Finansdepartementet, 2022; Statistisk sentralbyrå, 2020a). I tillegg har Skatteetaten vist tilbakeholdenhet når det kommer til å fremme konkursbegjæringer under pandemien, noe som også kan ha bidratt til den reduserte konkursfrekvensen (Skatteetaten, 2021; Statistisk sentralbyrå, 2020b).

En annen periode som skiller seg noe ut fra analyseperioden er årene 2007 til 2009 som har en noe høyere konkurstrate. Økningen av andel konkurser i denne perioden kan forklares med utspringet og ettervirkningene av finanskrisen som rammet store deler av verdensøkonomien i 2008 (Statistisk sentralbyrå, 2009a). Likevel er hele perioden 2006 - 2020 sett under ett relativt stabil når det gjelder andel konkurser. Dette vitner om en homogen analyseperiode uten de store svingningene.

3.4 Ikke-finansielle variabler

Fra ikke-finansiell informasjon lagde vi til sammen 25 variabler fordelt på 14 variabler for egenskaper ved lederskapet, 8 variabler for egenskaper ved eierskapet, samt 3 variabler innen kategorien andre ikke-finansielle variabler. Disse er presentert i Tabell 3. Korrelasjonsmatrisen for disse 25 ikke-finansielle variablene er gitt i Appendiks B. Denne matrisen viser at det er få variabler som korrelerer. Unntaket er *andel_st_medl_ikke_aksj* og *styre_eierandel*, som har en korrelasjon på -0,90, samt *sl_eierandel* og *styre_eierandel*, som har en korrelasjon på 0,87. I tillegg har variablene *dl_eierandel* og *sl_eierandel*, samt *log_aksj_antall* og *aksj_min* en korrelasjon på henholdsvis 0,86 og -0,86.

Høyt korrelerte variabler blir med i vår analyse. Dette er fordi vi lar metoden *Least Absolute Shrinkage Selector Operator* (LASSO) velge ut variabler som skal benyttes i logistiske regresjonsmodeller (se kapittel 4). LASSO velger variabler på en måte som unngår multikollinearitet. Ved å presentere LASSO for høyt korrelerte variabler, lar vi altså denne metoden avgjøre hvilke variabler som skal inkluderes og ikke.

Tabell 3: Beskrivelse av ikke-finansielle variabler

Tabellen gir en oversikt over de ikke-finansielle variablene denne studien benytter. For hver variabel er det gitt en beskrivelse av type variabel samt en kort beskrivelse av hva variabelen betyr.

Variabelnavn	Type	Beskrivelse
Egenskaper ved lederskapet		
alder_dl	Kontinuerlig	Alder på daglig leder
alder_sl	Kontinuerlig	Alder på styrets leder
styre_alder_snitt	Kontinuerlig	Gjennomsnittlig alder på styremedlemmer
styre_alder_std	Kontinuerlig	Standardavvik alder på styremedlemmer
styre_antall	Kontinuerlig	Antall styremedlemmer
styre_andel_kjonn_k	Kontinuerlig	Andel styremedlemmer som er kvinner
d_samme_dl_sl	Kategorisk	Dummy for CEO-dualitet (0:nei, 1:ja)
d_dl_i_styre	Kategorisk	Dummy for om daglig leder sitter i styret (0:nei, 1:ja)
d_dl_kjonn_k	Kategorisk	Dummy for kjønn på daglig leder (0:mann, 1:kvinne)
d_to_dl	Kategorisk	Dummy for om selskapet har to daglige ledere (0:nei, 1:ja)
d_dl_fylke	Kategorisk	Dummy om daglig leder bor i samme fylke som selskapet (0:nei, 1:ja)
d_sl_fylke	Kategorisk	Dummy om styrets leder bor i samme fylke som selskapet (0:nei, 1:ja)
d_sl_kjonn_k	Kategorisk	Dummy for kjønn på styrets leder (0:mann, 1:kvinne)
styre_andel_fylke	Kontinuerlig	Andel styremedlemmer som er bosatt i samme fylke som selskapet
Egenskaper ved eierskapet		
andel_st_medl_ikke_aksj	Kontinuerlig	Andel styremedlemmer som ikke er aksjonær
aksj_antall	Kontinuerlig	Antall aksjonærer ved balansedato
aksj_std	Kontinuerlig	Standardavvik eierandeler aksjonærer ved balansedato
aksj_maks	Kontinuerlig	Gjennomsnittlig eierandeler aksjonærer ved balansedato
aksj_min	Kontinuerlig	Eierandel hos minste aksjonær ved balansedato
dl_eierandel	Kontinuerlig	Eierandel hos daglig leder ved årets slutt
sl_eierandel	Kontinuerlig	Eierandel hos styrets leder ved årets slutt
styre_eierandel	Kontinuerlig	Eierandel samlet for alle styremedlemmer ved årets slutt
Andre ikke-finansielle variabler		
d_fravalg_revisjon	Kategorisk	Dummy om selskapet har valgt bort ekstern revisor (0: nei, 1: ja)
d_utarbeidet_regn_forer	Kategorisk	Dummy om regnskapsfører har utarbeidet regnskapet (0:nei, 1:ja)
d_bistand_regn_forer	Kategorisk	Dummy om regnskapsfører har bistått i regnskapet (0:nei, 1:ja)

Datasettet består av kategorivariabler som omhandler egenskaper hos daglig leder og styrets leder, samt bruken av regnskapsfører og revisor. Det må nevnes i denne sammenhengen at begrepet daglig leder blir i denne studien brukt som en fellesbetegnelse på rollen som øverste myndighet i selskapet, og inkluderer i tillegg forretningsfører og kontaktperson.

Tabell 4 viser at i 89 % av tilfellene er daglig leder bosatt i samme fylke som selskapets forretningsadresse. Tilsvarende for styrets leder er om lag 86 %. Det er 805 572 selskaper hvor daglig leder og styrets leder er samme person. Dette tilsvarer en andel på om lag 55 % av alle selskaper, og vitner om at CEO-dualitet er noe som forekommer hyppig blant norske selskaper. Videre har 85 % av selskapene mannlig daglig leder, og 88 % har mannlig styreleder. Selskaper som har to i rollen som daglig leder er svært sjelden, og forekommer i under 1 % av observasjonene.

Tabell 4: Deskriptiv statistikk av kategoriske ikke-finansielle variabler

Tabellen viser beskrivende statistikk for de kategoriske ikke-finansielle variablene på tvers av alle data i vårt endelige datasett. Kolonnen mest frekvent viser hvilken dummyverdi som forekommer oftest i datasettet. Kolonnene antall og andel angir i hvor mange observasjoner denne verdien opptrer med tilhørende andel. Det er også oppgitt hvor mange manglende verdier den aktuelle variabelen har gjennom perioden. Det må presiseres at for de fem variablene som omhandler revisjon, regnskapsfører samt fylke er den beskrivende statistikken kun for årene 2014 - 2020 på grunn av manglende verdier i foregående år.

Variabler	Mest frekvent	Antall mest frekvent	Andel mest frekvent	Manglende verdier
d_samme_dl_sl	dummy = 1	805 572	54,69 %	20,16 %
d_dl_i_styre	dummy = 1	1 109 611	74,59 %	19,36 %
d_dl_kjonn_k	dummy = 0	1 335 402	84,94 %	14,78 %
d_sl_kjonn_k	dummy = 0	1 450 096	88,36 %	11,04 %
d_to_dl	dummy = 0	1 560 475	99,26 %	14,78 %
d_fravalg_revisjon	dummy = 0	552 365	60,55 %	6,12 %
d_utarbeidet_regn_forer	dummy = 1	591 958	64,52 %	5,58 %
d_bistand_regn_forer	dummy = 1	575 455	62,77 %	5,65 %
d_dl_fylke	dummy = 1	737 216	88,88 %	14,65 %
d_sl_fylke	dummy = 1	723 478	85,94 %	13,37 %

Tabell 5 gir en oversikt over beskrivende statistikk av kontinuerlige ikke-finansielle variabler. Alder på daglig leder i selskapene varierer fra 9 til 100 år, med et gjennomsnitt på 49 år. Når det gjelder styrets leder er gjennomsnittlig alder 51 år, med et spenn fra 7 til 100 år. Videre ligger 1. og 99. persentilen på henholdsvis 27 og 73 år for daglig leder og 28 og 75 år for styrets leder. Dette tyder på at datasettet inneholder utliggere med ekstremverdier, noe vi håndterer ved logaritmisk transformasjon og *winsorizing* (se kapittel 4).

Tabell 5: Deskriptiv statistikk av kontinuerlige ikke-finansielle variabler

Tabellen inneholder beskrivende statistikk av kontinuerlige ikke-finansielle variabler. For hver variabel er det oppgitt gjennomsnittsverdi på tvers av alle observasjoner, standardavvik, minimumsverdi og maksimumsverdi. Kolonnen manglende verdier angir hvor stor del av datasettet som mangler verdier for aktuell variabel for perioden. Persentilene angir observasjoner mindre eller lik den gitte prosentisatsen. Det må presiseres at for de variablene som gjelder eierandel, aksjonærer og fylke er den beskrivende statistikken kun for årene 2014 - 2020 på grunn av manglende verdier i foregående år.

Variabler	Snitt	Std	Min	1 %	25 %	50 %	75 %	99 %	Maks	Manglende verdier
alder_dl	49,55	10,61	9,00	27,00	42,00	49,00	57,00	74,00	100,00	21,47 %
alder_sl	51,55	11,00	7,00	27,00	44,00	51,00	59,00	76,00	100,00	20,16 %
styre_alder_snitt	50,26	9,71	7,00	28,00	44,00	50,00	57,00	73,00	100,00	13,41 %
styre_alder_std	3,13	4,61	0,00	0,00	0,00	0,00	5,10	16,50	32,50	13,41 %
styre_antall	2,13	1,35	1,00	1,00	1,00	2,00	3,00	6,00	19,00	9,22 %
styre_andel_kjonn_k	0,16	0,29	0,00	0,00	0,00	0,00	0,25	1,00	1,00	9,22 %
andel_st_medl_ikke_aksj	0,48	0,46	0,00	0,00	0,00	0,50	1,00	1,00	1,00	3,04 %
dl_eierandel	0,43	0,44	0,00	0,00	0,00	0,33	1,00	1,00	1,00	7,84 %
sl_eierandel	0,41	0,44	0,00	0,00	0,00	0,25	1,00	1,00	1,00	4,14 %
styre_eierandel	0,51	0,47	0,00	0,00	0,00	0,50	1,00	1,00	1,00	3,04 %
aksj_antall	2,92	27,08	1,00	1,00	1,00	1,00	2,00	23,00	9470,00	1,10 %
aksj_std	0,05	0,10	0,00	0,00	0,00	0,00	0,01	0,44	0,50	1,10 %
aksj_maks	0,81	0,26	0,00	0,20	0,51	1,00	1,00	1,00	1,00	1,10 %
aksj_min	0,70	0,38	0,00	0,00	0,33	1,00	1,00	1,00	1,00	1,10 %
styre_andel_fylke	0,87	0,29	0,00	0,00	1,00	1,00	1,00	1,00	1,00	20,02 %

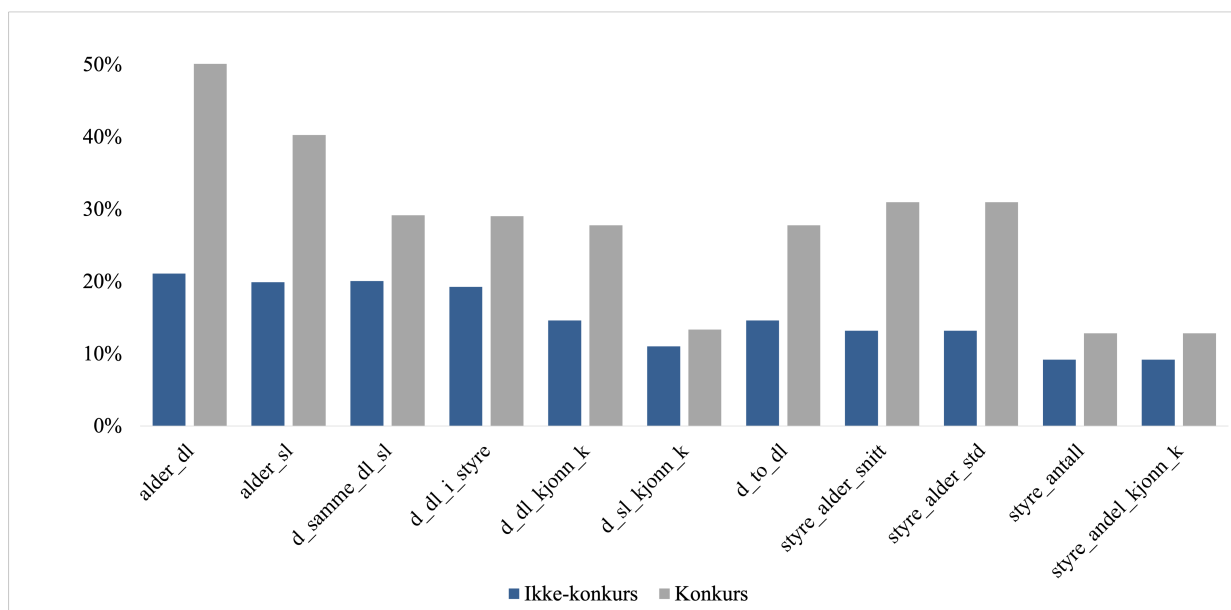
Alder for styrets medlemmer ligger i samme område som for styrets leder, og vitner om at det er stor homogenitet mellom styrets leder og styremedlemmer når det gjelder alder. Det kan nevnes at i de tilfellene hvor selskapet kun har oppført styreleder og ikke styremedlemmer, består styret kun av styrets leder. For øvrig i kategorien styre inngår det i tillegg til styremedlemmer også nestleder i de tilfellene det eksisterer. Eventuelle varamedlemmer eller varaledere ble ikke inkludert til å være en del av styret. Størrelsen på styret hos selskapene varierer fra 1 til 19 styremedlemmer. Et gjennomsnitt på 2,13 medlemmer og et standardavvik på 1,35 viser at hovedtyngden av observasjonene ligger i nedre del hva gjelder antall styremedlemmer. Dette samsvarer godt med datagrunnlaget vårt som er SMB, som følgelig har færre medlemmer i styret enn store børsnoterte selskaper. Kjønnssammensetningen i styret viser at det i gjennomsnitt er 16 % kvinner i styret, og at andelen menn følgelig i sterk overvekt med en andel på 84 %. Videre har i gjennomsnitt 51 % av styremedlemmene eierandeler i selskapet gjennom aksjer. Når det gjelder daglig leder og styrets leder har disse rollene i gjennomsnitt eierandel på henholdsvis 43 % og 41 % selskapet. Styret som helhet sitter samlet sett på 51 % av selskapenes aksjer i gjennomsnitt. Det er oppgitt at maksimalt antall aksjonærer på tvers av alle selskaper er 9470. 99. persentilen gir imidlertid en verdi på 23 aksjonærer, noe som vitner om at denne variabelen inneholder ekstremverdier.

3.5 Manglende verdier

De finansielle variablene fra datasettet som inneholder regnskapstall fra Registerenheten i Brønnøysund er komplett, og har ingen manglende verdier for regnskapstallene. Dette datasettet inneholder imidlertid tilleggsvariabler hvor enkelte variabler har manglende verdier. Tre av disse variablene, *fravalg_revisjon*, *bistand_regnskapsforer* og *utarbeidet_regnskapsforer* inkluderes i våre modeller som ikke-finansielle variabler. Disse er kun tilgjengelige fra og med regnskapsår 2012, og vil derfor ikke bli vurdert før denne perioden. Den ikke-finansielle informasjonen som er innhentet fra Enin AS har manglende verdier i varierende grad. Data om selskapenes aksjonærer er tilgjengelig fra og med 2014. Videre

viser oversikten over antall manglende verdier som er presentert i Appendiks C at variablene som inneholder informasjon om fylkesnummer har svært mange manglende verdier i regnskapsårene før 2014. Variablene som inneholder fylkesnummer vil derfor, sammen med variablene for regnskapsfører, revisor og aksjonærvariablene, inkluderes i egne modeller som kun benytter de nevnte årene. Dette vil bli beskrevet nærmere i kapittel 4.

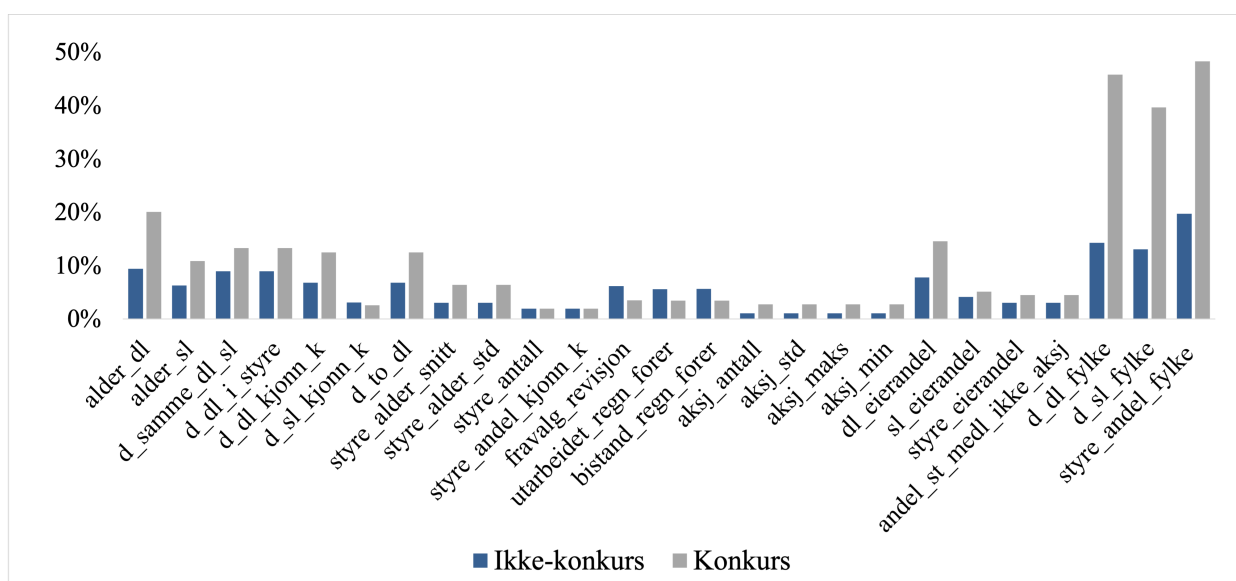
Figur 1 viser hvordan andelen av manglende verdier fordeler seg på henholdsvis konkurs og ikke-konkurs for de ikke-finansielle variablene som vil inkluderes i analysen for perioden 2006 - 2020. Andelen av manglende observasjoner er vesentlig høyere blant observasjonene som er konkurs enn ikke-konkurs. Dette er uheldig, da fordelingen av informasjon i data-settet blir skjev, og modellene vil ha mer tilgjengelig informasjon om selskaper som ikke har gått konkurs enn selskaper som har gått konkurs.



Figur 1: Manglende verdier for ikke-finansielle variabler 2006 - 2020

Figuren viser manglende verdier for ikke-finansielle variabler som vil bli benyttet i hele analyseperioden 2006 - 2020. De blå stolpene angir andelen manglende verdier for ikke-konkurs, og de grå stolpene viser andelen manglende verdier for konkurs. Andelen er oppgitt i prosent.

Figur 2 viser fordeling av andel manglende verdier fordelt på konkurs og ikke-konkurs for variablene i perioden 2014 - 2020. For denne perioden er det jevnt over mindre manglende verdier for alle variablene, noe som knyttes til økt dekning av data hos Enin AS de siste regnskapsårene. Variablene som er knyttet til fylkesnummer er de variablene som har høyest andel manglende verdier i perioden. Det er benyttet k-nærmeste naboer (kNN) for utfylling av manglende verdier, dette beskrives nærmere i kapittel 4.



Figur 2: Manglende verdier for ikke-finansielle variabler 2014 - 2020

Figuren viser manglende verdier for ikke-finansielle variabler som vil bli benyttet til analyser i perioden 2014 - 2020. De blå stolpene angir andelen manglende verdier for ikke-konkurs, og de grå stolpene viser andelen manglende verdier for konkurs. Andelen er oppgitt i prosent.

3.6 Finansielle variabler

Tabell 6 viser de finansielle variablene denne studien vil benytte for å vurdere betydningen av ikke-finansielle variabler. De finansielle variablene baserer seg på den utvidede SEBRA-modellen (Bernhardsen & Larsen, 2007) og Altman (1968) sin Z-score modell. For variablene som inngår i SEBRA-modellen er imidlertid aldersvariabelen, som angir antall år etter etablering, samt størrelsesvariabelen som angir selskapets totale eiendeler, transformert til logaritmisk verdi. Dette er i samsvar med hva andre lignende studier har gjort

på området (Paraschiv et al., 2021). For variablene som angår Altman er det i denne studien benyttet variabelen *ek_gjeld*, bokført egenkapital som andel av sum gjeld, fra Altman (2000) sin reviderte Z'-score modell.

Tabell 6: Beskrivelse av finansielle variabler

Tabellen gir en oversikt over de finansielle variablene denne studien benytter, med tilhørende forklaring. Variablene baserer seg på den utvidede SEBRA-modellen (Bernhardsen & Larsen, 2007) samt Altman (1968) sin Z-score modell.

Variabelnavn	Kategori	Beskrivelse
SEBRA		
res_gjeld	Inntjening	Ordinært resultat før av- og nedskrivninger i prosent av total gjeld
ek_tk	Soliditet	Egenkapital i prosent av total kapital
d_bek_iek	Soliditet	Dummy om bokført egenkapital er mindre enn innskutt egenkapital (0:nei, 1:ja)
likv_kg_omset	Likviditet	Likvide midler minus kortsiktig gjeld i prosent av omsetning
log_alder	Alder	Logaritmen av år siden etablering
log_tk	Størrelse	Logaritmen av total kapital i faste kroner
levgjeld_tk	Likviditet	Leverandørgjeld i prosent av total kapital
avg_tk	Likviditet	Skyldige offentlige avgifter i prosent av total kapital
Altman		
arbkap_tk	Likviditet	Arbeidskapital som andel av sum eiendeler
opptj_ek_tk	Lønnsomhet	Opptjent egenkapital som andel av sum eiendeler
EBIT_tk	Gjeld	EBIT som andel av sum eiendeler
ek_gjeld	Soliditet	Bokført egenkapital som andel av sum gjeld
salginntekt_tk	Lønnsomhet	Salgsinntekter som andel av sum eiendeler

4 Metode

I likhet med at variablene som er benyttet for konkursprediksjon har endret seg gjennom historien, har også metodene som er blitt benyttet vært gjennom en utvikling. Altman (1968) sin anerkjente Z-score modell benytter seg av multippel diskriminantanalyse som metode for konkursprediksjon. Dette er en statistisk metode som benyttes til klassifikasjonsproblemer hvor forskjeller mellom to eller flere kategorier undersøkes (James, Witten, Hastie & Tibshirani, 2013, s. 138). Metoden ble introdusert av Ronald A. Fisher (1936), men på grunn av dens restriktive antagelser har metoden vist seg å være problematisk å benytte innen økonomi og finans (Eisenbeis, 1977). For å bøte på disse problemene utviklet Ohlson (1980) O-score modellen som benyttet en betinget logistisk regresjonsmodell. Ohlson (1980) begrunner dette valget av metode med at man da unngår en del av problemene som er knyttet til multivariat diskriminantanalyse, slik som dårlig tolkbarhet og at det stilles enkelte krav til fordelingsegenskapene til forklaringsvariablene. Altman & Sabato (2007) gjennomførte en sammenligningsstudie på disse metodene, og funnene deres tydet på at multivariat diskriminantanalyse trolig har dårligere prediksjonsevne enn logistisk regresjon når det benyttes de samme variablene. På bakgrunn av metodens gode egenskaper har logistisk regresjon blitt en dominerende klassifikasjonsmetode i moderne kredittvurderinger og finansiell forskning generelt (Audrino, Kostrov & Ortega, 2019; Jones, Johnstone & Wilson, 2015).

Schumway (2001) argumenterer for at *discrete hazard models* er en mer nøyaktig metode å benytte for konkursprediksjon. Dette er en variant av logistisk regresjon hvor det benyttes flere årsregnskaper fra hvert selskap til prediksjonen, og ikke bare det foregående året. Dette er en svært anerkjent metode i dag, og benyttes både av banker og andre innen finanslitteraturen. Jones, Johnstone & Wilson (2017) understreker at statistiske metoder som logistisk regresjon og diskriminantanalyse er gode metoder for konkursprediksjon, men at også mer avanserte teknikker innen maskinlæring kan bedre modellene. Dette understrekes av Ravi Kumar & Ravi (2007) som i sin studie belyste at enkle statistiske metoder gjerne kan suppleres med mer avanserte metoder innen maskinlæring. Av norske studier på

området må studien av Næss, Wahlstrøm, Helland & Kjærland (2017) trekkes frem, hvor de finner at maskinlæringsteknikker imidlertid estimerer konkurs marginalt dårligere enn statistiske metoder, men at denne forskjellen ikke er signifikant.

4.1 Balansering og estimering

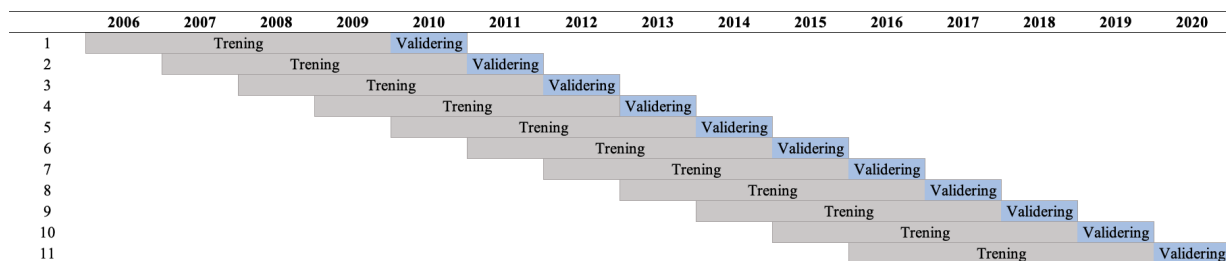
Balansering av datasett er et tema som har vært mye omdiskutert i litteraturen innenfor konkursprediksjon. Li & Sun (2012) hevder at flere studier om konkursprediksjon ikke i tilstrekkelig grad har tatt med i vurderingen at den aktuelle populasjonen i virkeligheten består av ubalanserte data, og finner i sin studie at balansering av datasettet fører til en forbedret prediksjonsevne. I realiteten er det en liten andel av alle selskaper som går konkurs, og dersom hele datasettet benyttes vil det medføre at andelen ikke-konkurs vil være langt større enn andelen selskaper som har gått konkurs. Lopez et al. (2013) peker på at en slik ubalanse mellom klassene kan føre til det som kan omtales som et *klasseubalanse problem*. Videre viser de til at det som regel er slik at det er den minste klassen, minoritetsklassen, som er av størst interesse, og at feilklassifikasjon av minoritetsklassen derfor ofte kan medføre store kostnader.

Lopez et al. (2013) viser til at det i ubalanserte datasett kan oppstå problemer ved at klassifiseringsalgoritmene tenderer mot å foretrekke majoritetsklassen, slik at minoritetsklassen ender opp med en høy feilklassifikasjonsrate. Li & Sun (2012) påpeker også at for et ubalansert datasett hvor majoriteten av observasjonene er selskaper som ikke er konkurs, vil selv en modell som predikerer alle observasjoner som ikke-konkurs oppnå en svært høy treffsikkerhet, men at en slik modell i realiteten ikke vil ha noen verdi.

Zmijewski (1984) retter imidlertid kritikk mot bruk av balanserte datasett, og viser til at bruk av ikke-tilfeldige utvalg kan føre til estimeringsbias i form av valgbasert utvalgsbias, og i så måte føre til forvrengte modeller. Ohlson (1980) benyttet et ubalansert datasett, og bruken av ubalanserte datasett som representerer den virkelige populasjonen er vanlig praksis i litteraturen i nyere tid (se for eksempel Bernhardsen & Larsen, 2007; Paraschiv et al., 2021). På bakgrunn av dette vil vi derfor benytte et ubalansert datasett.

I denne studien vil det bli predikert konkurs ett år frem i tid, en tidshorisont som er mye anvendt i tidligere forskning (Tian, Yu & Guo, 2015). Vi følger Paraschiv et al. (2021) og benytter *rolling window* og *forward validation procedure* som valideringsteknikker. Dette innebærer at det benyttes foregående regnskapsår til å trene modellene, hvor tilpasningen videre evalueres på treningssettet. Deretter testes prestasjonen på det påfølgende regnskapsåret i valideringssettet. På denne måten vil man ikke benytte fremtidige år i modellene for å predikere konkurs. Vi benytter fire regnskapsår til å trene modellene og ett regnskapsår til validering. Dette gir følgelig en 80/20-fordeling mellom treningssett og valideringssett, noe som sikrer at modellene har tilstrekkelig mengde data for å trene modellene, men samtidig nok data i valideringssettet for å sikre gode evalueringer.

En fordel med å benytte *rolling window* som valideringsteknikk, fremfor alternativet *expanding window*, er at alle treningssett vil inneholde like mange regnskapsår, noe som er gunstig for sammenlignbarheten. Dette vil også utelukke at modellene benytter for gammel og udatert informasjon til trening av modeller. Ved å benytte flere regnskapsår for å predikere neste års konkurs vil konkurranserisikoen over tid bli vurdert.



Figur 3: Inndeling i treningssett og valideringssett

Figuren viser hvordan perioden som benyttes som treningssett og valideringssett forskyves ett år for hver fold, en metode som kalles *rolling window*. Ved å benytte denne metoden vil det for alle folder være like mange år i hvert treningssett.

4.2 Håndtering av manglende observasjoner

De manglende observasjonene i datasettet er utfylt med k-nærmeste naboer (kNN). For å unngå problemer med *data leakage* er datasettet delt for hvert regnskapsår, slik at kNN kun benytter verdier fra det aktuelle året til å fylle ut manglende verdier. For kategori-variablene er det benyttet $k=1$, og for de kontinuerlige variablene er det benyttet $k=3$.

Dette innebærer at kNN benytter den nærmeste naboen for utfylling av de manglende kategorivariablene, og de tre nærmeste naboene for å fylle ut de manglende observasjonene for de kontinuerlige variablene. Videre benytter kNN-algoritmen selskapenes alder, totale eiendeler, inntekter og lønnskostnader som sammenligningsgrunnlag, da disse antas å tilsammen gi en god beskrivelse av de ulike selskapenes karakteristikk. Dummyvariablene *fravalg_revisjon*, *bistand_regnskapsforer* og *utarbeidet_regnskapsforer* er kun tilgjengelige fra og med regnskapsår 2012, og er derfor ikke utfylt i årene før 2012. For aksjonærvariablene er det kun fylt ut manglende verdier for regnskapsårene 2014 til 2020. Dette fordi data om aksjonærer ikke er tilgjengelig fra årene før 2014.

4.3 Håndtering av ekstremverdier

Ekstremverdier, eller utliggere, er observasjoner som avviker og som klart skiller seg fra øvrige observasjoner, og kan i noen tilfeller skyldes feil i datasettet (James et al., 2013, s. 96). Slik presentert i forrige kapittel inneholder datagrunnlaget i denne studien ekstremverdier, og for å dempe effekten av slike utliggere i datagrunnlaget er det utført logaritmisk transformasjon av de ikke-finansielle variablene *alder_dl*, *alder_sl*, *styre_alder_snitt*, *styre_antall* og *aksj_antall*. Videre følger vi Paraschiv et al. (2021) og benytter den logaritmiske transformasjonen av total kapital inspirert av SEBRA-modellen (Eklund et al., 2001), og logaritmen av antall år siden etablering. For de øvrige variablene, med unntak av dummyvariablene, blir deres verdier begrenset mellom 1. og 99. persentil. Dette kalles ofte *winsorizing* og er en fremgangsmåte som er mye benyttet i annen litteratur (se for eksempel Paraschiv et al., 2021; Pelja & Wahlstrøm, 2021; Shumway, 2001; Tian et al., 2015). Dette utføres for hvert regnskapsår, og begrenser i så måte effekten av ekstremverdier i datasettet.

4.4 Oppsett av modellene

For å vurdere betydningen av å inkludere ikke-finansielle variabler for konkursprediksjon vil det i denne studien bli konstruert åtte ulike modeller. Modellene består av finansielle

og ikke-finansielle variabler og er kombinasjoner av variabelsettene presentert i Tabell 7.

Tabell 7: Oversikt over variabelsett

Tabellen gir en oversikt over de ulike variabelsettene som vil bli benyttet til analysen. For hver modell vil det være oppgitt hvilke av disse variabelsettene som inngår. Variabelsett I og II er utelukkende ikke-finansielle variabler. Variabelsett Altman og variabelsett SEBRA inneholder kun finansielle variabler.

Variabelsett I	Variabelsett II	Variabelsett Altman	Variabelsett SEBRA
log_alder_dl	andel_st_medlikke_aksj	arbkap_tk	res_gjeld
log_alder_sl	log_aksj_antall	opptj_ek_tk	ek_tk
log_styre_alder_snitt	aksj_std	EBIT_tk	d_bek_iek
styre_alder_std	aksj_maks	ek_gjeld	likv_kg_omset
log_styre_antall	aksj_min	salginntekt_tk	log_alder
styre_andel_kjonn_k	dl_eierandel		log_tk
d_dl_kjonn_k	sl_eierandel		levgjeld_tk
d_sl_kjonn_k	styre_eierandel		avg_tk
d_samme_dl_sl	d_sl_fylke		
d_dl_i_styre	d_dl_fylke		
d_to_dl	styre_andel_fylke		
	d_fralag_revisjon		
	d_utarbeidet_regn_forer		
	d_bistand_regn_forer		

Tabell 7 viser at denne studien vil benytte fire ulike variabelsett. De ikke-finansielle variablene er fordelt over to ulike variabelsett; variabelsett I og variabelsett II. Årsaken til denne fordelingen er manglende observasjoner for en del av de ikke-finansielle variablene i første del av analyseperioden. Omfanget av manglende verdier og hvilke variabler det gjelder er illustrert i Appendiks C. Variablene som inngår i variabelsett I er variabler som har lite manglende observasjoner gjennom perioden, og som derfor vil bli inkludert i hele analyseperioden. Når det gjelder variabelsett II har disse variablene, enten delvis eller fullstendig, manglende verdier frem til og med 2013. Disse variablene vil derfor kun inkluderes i analysene fra og med regnskapsår 2014, og det vil av den grunn kun foreligge modellevaluering for regnskapsår 2018, 2019 og 2020. Dette kan illustreres med Figur 3 som ble presentert i kapittel 4.1. De nevnte modellene vil kun benytte fold 9, 10 og 11, hvor de trenes i årene 2014 til 2017 og testes i regnskapsår 2018, og videre trenes i 2015 til 2018 og testes i regnskapsår 2019, og deretter trenes i årene 2016 til 2019 og testes i regnskapsår 2020. For resterende modeller vil modellene bli evaluert for årene 2010 til 2020. Videre er variablene i variabelsett Altman inspirert av de finansielle variablene som inngår i Altman (1968) sin Z-score modell. Avslutningsvis er variablene i variabelsett SEBRA inspirert av variablene i SEBRA-modellen.

Tabell 8 gir en oversikt over studiens modeller, hvilke variabelsett som inngår i hver modell og hvilke metoder som blir benyttet. Modell 1A og 1B består utelukkende av ikke-finansielle variabler, hvor modell 1A inneholder de ikke-finansielle variablene fra variabelsett I, og modell 1B består både av variabelsett I og II. Dersom alle variablene for 1A og 1B benyttes ved trening av modellene ved logistisk regresjon, forventer vi at det vil oppstå problemer som multikollinearitet. Vi anvender derfor LASSO for å velge variabler før disse benyttes for trening.

Modell 2A, 2B og 2C består av ulike kombinasjoner av finansielle variabler fra modellen til Altman (1968) og våre ikke-finansielle variabler. Videre består modell 3A, 3B og 3C av ulike kombinasjoner av finansielle variabler fra SEBRA-modellen og våre ikke-finansielle variabler. Modellene 2A og 3A er inspirert av henholdsvis Altman (1968) og SEBRA-modellen, og inneholder utelukkende finansielle variabler. Variablene til disse modellene er sammensatt på en slik måte at de skal dekke hovedaspektene ved selskapenes økonomi uten at det oppstår problemer som multikollinearitet. Modellene 2B og 3B inneholder både de finansielle variablene i henholdsvis 3A og 3B, men i tillegg også de ikke-finansielle variablene fra variabelsett I. Videre inneholder modellene 2C og 3C de samme variablene som i henholdsvis 2B og 3B, i tillegg til de ikke-finansielle variablene fra variabelsett II. Av samme grunn som nevnt ovenfor, benytter vi også LASSO for valg av variabler for 2B, 2C, 3B og 3C. Modellene 1B, 2C og 3C benyttes kun for regnskapsårene 2014 – 2020 siden variabelsett II kun er utviklet for disse årene.

Tabell 8: Oversikt over modeller

Tabellen viser en oversikt over hvilke variabelsett som inngår i modellene denne studien presenterer. Modell 1A og 1B består utelukkende av ikke-finansielle variabler. Modell 2A, 2B og 2C er ulike kombinasjoner av ikke-finansielle variabler og variabler inspirert av Altman sin Z-score modell (1968). Modell 3A, 3B og 3C består av ulike kombinasjoner av ikke-finansielle variabler og variabler inspirert av SEBRA-modellen (2007).

Modell	Metode	Variabler
Modell 1A	LASSO + Logistisk regresjon	Variabelsett I
Modell 1B	LASSO + Logistisk regresjon	Variabelsett I + Variabelsett II
Modell 2A	Logistisk regresjon	Variabelsett Altman
Modell 2B	LASSO + Logistisk regresjon	Variabelsett I + Variabelsett Altman
Modell 2C	LASSO + Logistisk regresjon	Variabelsett I + Variabelsett II + Variabelsett Altman
Modell 3A	Logistisk regresjon	Variabelsett SEBRA
Modell 3B	LASSO + Logistisk regresjon	Variabelsett I + Variabelsett SEBRA
Modell 3C	LASSO + Logistisk regresjon	Variabelsett I + Variabelsett II + Variabelsett SEBRA

4.5 LASSO

I denne studien vil vi benytte LASSO for seleksjon av variabler for å lage modeller med et redusert antall parametere (Nokeri, 2021, s. 50). Variablene som blir valgt ut vil videre bli benyttet i logistiske regresjoner for å evaluere prestasjon og for å undersøke koeffisientverdier og deres signifikans. En av fordelene ved å bruke LASSO er at denne metoden vil håndtere eventuelle problemer med multikollinearitet, i tillegg til at den minimerer den numeriske ustabiliteten som kan oppstå som en følge av at modellen er overtilpasset (Schalck & Yankol-Schalck, 2021). Et annet moment er tolkbarhet. Ved at LASSO velger ut de forklaringsvariablene med størst effekt og setter andre koeffisienter til 0 vil man oppnå modeller med økt tolkbarhet (Tibshirani, 1996). I tillegg gir dette informasjon om hvilke variabler som er viktigst, siden LASSO vil velge de variablene som er av størst betydning først.

LASSO-koeffisientene estimeres ved å minimere:

$$-\ell(\beta, \beta_0) + \lambda \|\beta\|_1 \quad (2)$$

hvor $\ell(\beta, \beta_0)$ er gitt ved:

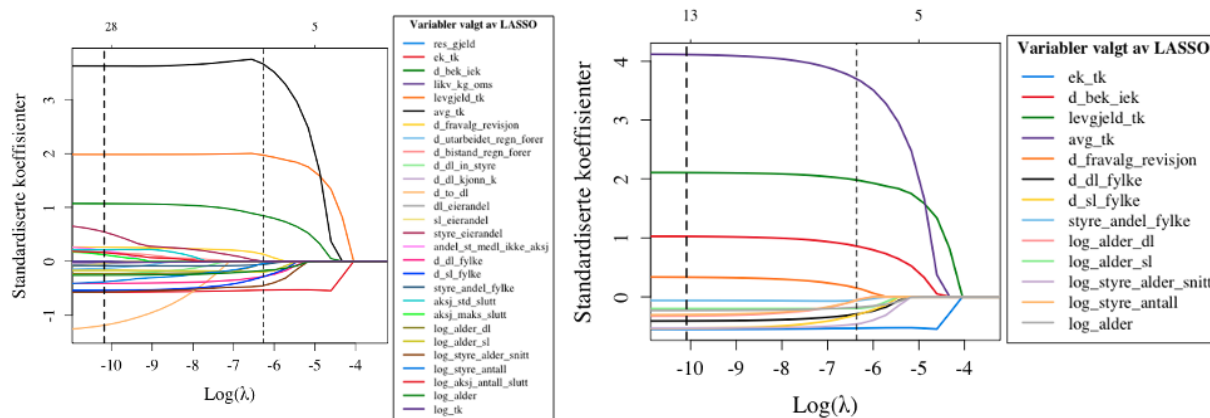
$$\ell(\beta, \beta_0) = \sum_{n=1}^N [y \odot (X\beta + \iota\beta_0) - \log(\iota + \exp(X\beta + \iota\beta_0))] \quad (3)$$

Her er y angitt som en vektor av faktiske klassifiseringer av konkurs og ikke-konkurs, hvor verdien 1 angir konkurs og verdien 0 angir ikke-konkurs. X er videre en matrise med én rad per observasjon og én kolonne per forklaringsvariabel. β angir modellens koeffisienter som må estimeres, \odot angir videre Hadamard-produktet, mens ι er en vektor med kun 1-tall og som har en lengde tilsvarende $N * 1$. N er her angitt som antall observasjoner, som i denne analysen vil tilsvare antall årsregnskaper. Videre er λ definert som en positiv tuningparameter og $\|\beta\|_1$ er ℓ_1 -normen av β (Paraschiv et al., 2021). ℓ_1 -normen vil medføre

at effekten av enkelte koeffisienter vil krympes helt til null dersom tuningparameteren λ er stor nok. På denne måten vil LASSO føre til variabelseleksjon hvor uavhengige variabler som ikke har betydning elimineres bort (James et al., 2013, s. 219). Valg av λ er derfor av stor betydning. Dersom $\lambda = 0$ vil estimering bli lik minste kvadraters metode, og ingen av variablene vil bli ekskludert fra modellen. Om λ settes alt for høyt vil dette resultere i at alle koeffisientene settes til null, noe som vil tilsvare ingen effekt av tilhørende uavhengige variabler (James et al., 2013, s. 220).

4.5.1 Parametervalg

I denne studien vil det blant annet benyttes AUC som prestasjonsmål, noe som forklares nærmere i kapittel 4.9. For valg av tuningparameteren λ vil det benyttes 3-fold kryssvalidering, hvor verdien av λ som gir den beste gjennomsnittlige verdien for *in-sample* AUC blant alle verdier av λ som er inkludert i grid-søket identifiseres. Verdien av λ som så velges er den verdien som gir en *in-sample* AUC som ligger ett standardavvik fra den beste. Grunnen til dette er at verdien av λ som gir beste AUC ofte inkluderer svært mange variabler, noe som kan føre til kompliserte modeller og overtilpasning.



Figur 4: Valg av tuningparameter λ

Figuren viser hvordan valg av λ påvirker antall variabler som velges ut av LASSO for modell 3C i valideringsår 2019. Den uthevede vertikale stiplede linjen i plottene verdien for λ som gir best AUC og den smalere vertikale stiplede linjen i plottene viser verdien for λ som gir AUC ett standardavvik fra beste. Plottet til venstre viser variabler som velges av LASSO dersom verdien av λ som gir best AUC benyttes. Plottet til høyre viser variabler som velges ved bruk av λ som gir AUC ett standardavvik fra beste.

Figur 4 viser et eksempel fra modell 3C med valideringsår 2019. Illustrasjonen viser at bruk av verdien for λ som gir en AUC ett standardavvik fra beste AUC vil føre til at LASSO velger færre variabler, og i dette tilfellet reduseres antall valgte variabler fra 28 til 13, mens *in-sample* AUC kun reduseres fra 0,867 til 0,864. Begge har en *out-of-sample* AUC på 0,856.

4.6 Logistisk regresjon

I denne studien vil vi benytte logistisk regresjon som metode for å anslå sannsynligheten for konkurs. Logistisk regresjon er en statistisk klassifikasjonsmetode som benyttes når responsvariabelen Y er en kategorivariabel med et gitt antall mulige utfall. Metoden beregner hvor mye den naturlige logaritmen til sannsynligheten for $Y=1$ endres når X endres med én enhet, og angir sannsynligheten for at Y tilhører en gitt kategori (James et al., 2013, s. 130). I denne sammenhengen vil responsvariabelen stå ovenfor to mulige klasser; konkurs og ikke-konkurs. I logistisk regresjon benyttes den logaritmiske funksjonen gjengitt i Ligning 4.

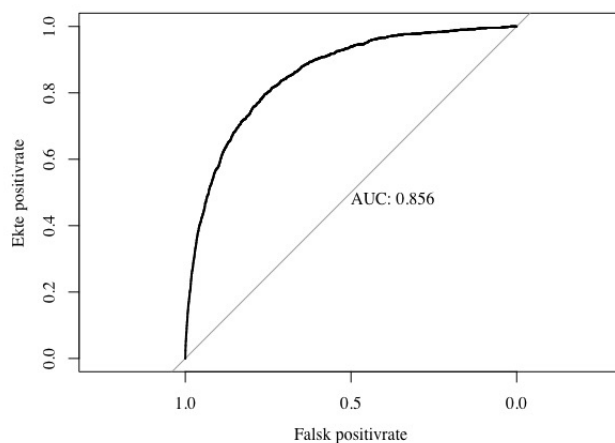
$$p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \quad (4)$$

Ved bruk av denne ligningen unngår man ulogiske output som ikke naturlig faller inn under de gitte kategoriene. Ved estimering av koeffisientene minimeres den negative av $\ell(\beta, \beta_0)$ som definert i ligning (3) (James et al., 2013, s. 132). Vi lar logistisk regresjon vurdere sannsynligheten for konkurs over flere år ved å trene modellene med årsregnskaper fra samme selskaper over flere år, som vist i Figur 3. Dette er omtalt i litteraturen som *descrete hazard models* (Shumway, 2001), og er benyttet i utstrakt grad i nyere studier (Paraschiv et al., 2021). I motsetning til statiske modeller, som bruker kun én enkel observasjon per selskap, tar *descrete hazard models* hensyn til at selskaper endrer seg over tid.

4.7 Evaluering av modellene

4.7.1 AUC

Et mye anvendt prestasjonsmål for konkursprediksjonsmodeller er AUC, som måler modellens evne til å klassifisere konkurs og ikke-konkurs korrekt (Bernhardsen & Larsen, 2007; Tian et al., 2015). AUC er en verdi i intervallet $[0.5,1]$, hvor høy verdi indikerer god forklaringskraft for modellen. Ofte blir verdier mellom 0,7 og 0,8 definert som akseptable, verdier mellom 0,8 og 0,9 anses som utmerket, mens verdier over 0,9 klassifiseres som fremragende (Hosmer, Lemeshow & Sturdivant, 2013, s. 177). Verdien angir arealet under modellens *receiver operating characteristic curve* (ROC-kurve), som er en kurve definert som ekte positivrate mot falsk positivrate for terskelverdier mellom 0 og 1 (Hosmer et al., 2013, s. 176). Figur 5 er en illustrasjon av en ROC-kurve med tilhørende AUC-verdi konstruert for en av denne studiens modeller. En optimal ROC-kurve vil plassere seg i øvre venstre hjørne, slik at arealet under grafen blir størst mulig (James et al., 2013, s. 147).



Figur 5: Illustrasjon av ROC-kurve

Figuren viser eksempel på en ROC-kurve med tilhørende AUC-verdi for modell 3B og regnskapsår 2010. Kurven viser bytteforholdet mellom falsk positivrate og ekte positivrate, og angir en AUC-verdi på 0,856.

4.7.2 Brier score

Brier score måler gjennomsnittet av de kvadrerte forskjellene mellom de predikerte verdiene og de faktiske utfallene, og ønskes derfor så lav som mulig. Desto nærmere 0 *Brier score*

er, jo nærmere er modellen de faktiske utfallene, og en score på 0 vil dermed indikere en perfekt prediksjon (Tian et al., 2015). *Brier score* er et mye anvendt prestasjonsmål, og er benyttet for flere studier innen konkurrerprediksjon (Paraschiv et al., 2021; Tian et al., 2015).

4.7.3 AIC

Akaike information criterion (AIC) angir den relative mengden av informasjon en modell ikke fanger opp og blir derfor mye anvendt som mål på hvor godt en modell presterer, (se for eksempel Paraschiv et al., 2021; Tian et al., 2015). Det er ønskelig å miste minst mulig informasjon, og derfor vil et lavt tall indikere at kvaliteten på modellen er bedre og prediksjonen følgelig mer pålitelig. AIC benytter antall variabler som inngår i modellen som straff, og gir en verdi basert på det optimale balansepunktet mellom kompleksitet og nøyaktighet (Tian et al., 2015).

4.7.4 Nøyaktighetsratio

Nøyaktighetsratio blir ofte benyttet som prestasjonsmål for konkurrerprediksjonsmodeller og defineres som AUC fratrukket 0,5 og deretter multiplisert med 2 (Pelja & Wahlstrøm, 2021; Tian et al., 2015). Dette gir en verdi i intervallet $[0,1]$ hvor 1 er en perfekt modell og 0 indikerer en modell basert på tilfeldigheter.

4.8 Repliserbarhet

Denne studien har benyttet Python for uthenting av data fra Enin AS. Samme programmeringsspråk ble også benyttet til sammenkobling av uthentede ikke-finansielle variabler sammen med datasettet som inneholdt finansielle variabler. Python ble valgt blant annet på bakgrunn av datasettets størrelse og omfang. Videre ble R benyttet til studiens analyser. Datasettet som er benyttet vil ikke publiseres grunnet begrensninger fra dataleverandørene.

5 Resultater

Dette kapitlet inneholder en vurdering av hvorvidt inkludering av ikke-finansiell informasjon i konkursprediksjonsmodeller vil øke deres prediksjonsevne, og eventuelt hvilke ikke-finansielle variabler som virker å ha størst betydning. Dette vil bli gjennomført gjennom å benytte variabelsettene som er inspirert av SEBRA-modellen (Bernhardsen & Larsen, 2007) og Z-score modellen til Altman (1968) som sammenligningsgrunnlag. For å måle prestasjonen av ikke-finansielle variabler alene, vil to av modellene utelukkende bestå av ikke-finansielle variabler, som presentert i Tabell 8. LASSO-plott til samtlige modeller vil bli presentert i Appendiks D.

5.1 Modell 1

5.1.1 Modell 1A

Tabell 9 gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsett I for å inngå i Modell 1A. Tabellen viser også koeffisientverdier og z-verdier når de valgte variablene blir benyttet i en logistisk regresjon, samt modellens tilpasning *in-sample* og hvor godt den predikerer *out-of-sample*. Dette presenteres for hver fold som vist i Figur 3.

Variabelen *d_dl_i_styre*, en dummy for om daglig leder sitter i styret, enten som styremedlem eller som styreleder, har en z-verdi med absoluttverdi større enn 3,29 og er dermed signifikant på 0,1 %-nivå gjennom hele perioden. I tillegg er dette variabelen som er valgt ut først av LASSO flest ganger for modell 1A, som vist i Appendiks E. Variabelen har negative fortegn, noe som virker fornuftig da det kan tenkes at det er en styrke for selskapet at daglig leder også sitter i styret. Videre er variabelen *d_samme_dl_sl*, som angir tilfellene hvor samme person innehar rollen både som daglig leder og styrets leder, valgt ut i alle år med unntak av 2019 og 2020. Variabelen har en z-verdi med absoluttverdi større enn 3,29 i alle periodene, og er dermed signifikant på 0,1 %-nivå. Med negative fortegn indikeres det at risikoen for konkurs reduseres dersom begge de nevnte rollene besittes av samme person. Dette er i tråd med funnene til Ciampi (2015), som fant at risikoen for økonomisk

mislighold ble redusert i de tilfeller hvor daglig leder og styrets leder er samme person.

Logaritmen av alderen på styrets leder og daglig leder, oppgitt som henholdsvis *log_alder_sl* og *log_alder_dl*, samt logaritmen av styrets gjennomsnittsalder, *log_styre_alder_snitt*, er med z-verdier med absoluttverdi større enn 3,29 alle signifikante på 0,1 %-nivå i alle regnskapsårene. Med negative fortegn impliserer dette at økt alder på personene som besitter disse rollene gir lavere sannsynlighet for konkurs, noe som virker fornuftig, da en høyere alder ofte innebærer økt erfaring og kompetanse. Dette støttes også av funnene til Platt & Platt (2012) som fant at risikoen for konkurs økte ved redusert alder på daglig leder, og Elloumi et al. (2001) som fant at gjennomsnittsalderen til styret var en av de mest betydningsfulle ikke-finansielle variablene for å forklare økonomisk mislighold. Koeffisientenes størrelse antyder at gjennomsnittsalderen på styret og alderen på styrets leder har større effekt enn alderen til daglig leder. Dette gjenspeiles også i rekkefølgen variablene blir valgt ut av LASSO på, da *log_alder_sl* og *log_styre_alder_snitt* blir valgt før *log_alder_dl* i alle årene.

Tabell 9: Resultater modell 1A

Tabellen gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsett I, og som er benyttet i en logistisk regresjon. Tabellen viser koeffisientverdiene fra den logistiske regresjonen med tilhørende z-verdier i parentes, samt modellens tilpasning *in-sample* ved Brier score, AUC og AIC, samt hvor godt den predikerer *out-of-sample* ved Brier score, AUC og nøyaktighetsratio. Dette presenteres for hver fold som vist i Figur 3.

	2010	2011	2012	2013	2014
log_alder_dl	-0,52(-7,33)	-0,56(-7,96)	-0,62(-8,45)	-0,56(-7,58)	-0,55(-7,25)
log_alder_sl	-0,78(-8,41)	-0,75(-8,25)	-0,69(-7,17)	-0,60(-6,02)	-0,75(-7,15)
log_styre_alder_snitt	-0,69(-7,53)	-0,79(-8,70)	-0,90(-9,35)	-1,00(-9,85)	-0,96(-8,75)
d_dl_i_styre	-0,39(-11,71)	-0,56(-17,17)	-0,70(-20,26)	-0,80(-22,50)	-0,77(-21,24)
d_samme_dl_sl	-0,30(-8,48)	-0,30(-8,40)	-0,30(-8,09)	-0,43(-10,61)	-0,41(-10,19)
styre_alder_std	-0,04(-10,61)	-0,04(-10,83)	-0,05(-13,05)	-0,03(-8,29)	-0,03(-7,44)
log_styre_antall				-0,25(-9,17)	-0,31(-10,80)
d_to_dl	-0,80(-2,96)				
d_dl_kjonn_k	0,15(4,65)				
(intercept)	4,00(14,38)	4,57(16,96)	5,01(18,36)	5,01(18,49)	5,31(20,46)
In-sample Brier score	0,035	0,034	0,031	0,029	0,028
In-sample AUC	0,620	0,645	0,665	0,681	0,679
AIC	72 216	73 380	68 475	66 193	65 510
Out-of-sample Brier score	0,030	0,030	0,028	0,027	0,025
Out-of-sample AUC	0,666	0,689	0,693	0,661	0,635
Nøyaktighetsratio	0,331	0,377	0,385	0,322	0,269

	2015	2016	2017	2018	2019	2020
log_alder_dl	-0,61(-8,09)	-0,74(-9,48)	-0,83(-10,36)	-0,89(-11,24)	-0,99(-12,77)	-0,94(-11,86)
log_alder_sl	-0,76(-6,97)	-0,83(-7,34)	-0,93(-8,20)	-0,72(-6,48)	-0,56(-5,14)	-0,45(-4,03)
log_styre_alder_snitt	-0,89(-7,72)	-0,80(-6,56)	-0,70(-5,52)	-0,84(-6,60)	-0,98(-7,94)	-1,09(-8,49)
d_dl_i_styre	-0,64(-17,87)	-0,50(-13,59)	-0,30(-8,15)	-0,18(-4,7)	-0,28(-9,15)	-0,20(-6,35)
d_samme_dl_sl	-0,42(-10,65)	-0,38(-9,65)	-0,33(-8,45)	-0,27(-7,23)		
styre_alder_std	-0,03(-6,96)	-0,02(-5,09)	-0,02(-3,69)			
log_styre_antall	-0,35(-11,93)	-0,39(-12,33)	-0,36(-11,02)	-0,42(-15,44)	-0,29(-12,93)	-0,26(-11,34)
d_to_dl		-2,05(-5,02)	-2,12(-5,19)	-2,04(-5,39)	-1,37(-5,47)	-0,95(-4,84)
d_dl_kjonn_k		0,11(3,36)				0,11(3,27)
(intercept)	5,33(21,68)	5,59(23,70)	5,80(25,81)	5,64(26,03)	5,79(28,07)	5,43(25,91)
In-sample Brier score	0,027	0,026	0,025	0,025	0,025	0,023
In-sample AUC	0,673	0,671	0,664	0,660	0,660	0,658
AIC	66 522	66 174	67 151	68 568	72 539	70 578
Out-of-sample Brier score	0,024	0,025	0,024	0,026	0,020	0,015
Out-of-sample AUC	0,659	0,654	0,652	0,655	0,647	0,651
Nøyaktighetsratio	0,319	0,307	0,304	0,310	0,294	0,302

Størrelsen på styret og spredningen på styremedlemmenes alder, presentert gjennom henholdsvis variablene *log_styre_antall* og *styre_alder_std* er signifikant på 0,1 %-nivå da z-verdiene har absoluttverdi større enn 3,29, men velges ikke ut i alle regnskapsår. Variablen *log_styre_antall* har negative fortegn og indikerer at dersom antall styremedlemmer øker, vil dette redusere risikoen for konkurs. Dette virker fornuftig, og kan trolig knyttes til at et økt antall medlemmer i styret vil bety økt mangfold og økt samlet kompetanse, noe Manzanque et al. (2016) fant at reduserte sannsynligheten for økonomisk mislighold. Videre har variabelen *styre_alder_std* også negative fortegn, noe som indikerer at dersom det blir økt spredning i alderen til styrets medlemmer vil dette redusere risikoen for konkurs. Dette vir-

ker fornuftig, da det innebærer at styrets medlemmer er representert i ulike aldersgrupper, og ofte kan ha ulike perspektiver og erfaringer. Det skal dog nevnes at koeffisientverdiene for denne variabelen er svært lav.

Enkelte selskaper har to personer i rollen som daglig leder, angitt her som *d_to_dl*. Denne variabelen blir valgt ut i seks regnskapsår og antyder at selskapet har mindre risiko for å gå konkurs om det er to personer som besitter denne rollen samtidig. Forekomsten av dette tilfellet utgjør imidlertid i denne studien kun under 1 % av alle observasjoner, og det kan derfor diskuteres hvor pålitelig dette funnet er. Det at fenomenet opptrer svært sjeldent kan være en mulig årsak til at dette er et område innen litteraturen som er lite diskutert. At variabelen har negative fortegn kan imidlertid virke fornuftig da to daglige ledere kan samlet sett inneha mer kompetanse, noe som kan tenkes å redusere sannsynligheten for konkurs.

Kjønnsvariablene *styre_andel_kjonn_k* og *d_sl_kjonn_k* ble av LASSO ikke funnet å være betydningsfulle, og er derfor ikke inkludert i modell 1A. Dette er i strid med Garcia & Herrero (2021) og Cho et al. (2021) som fremhever at kjønns sammensetningen i styret er av stor betydning når det gjelder sannsynligheten for konkurs. Variabelen *d_dl_kjonn_k* blir derimot valgt ut og er positiv og signifikant på 0,1 %-nivå i regnskapsår 2010, 2016 og 2020 med z-verdier med absoluttverdi større enn 3,29. Dette antyder at risikoen for konkurs øker dersom daglig leder er kvinne. Sammenlignet med resterende signifikante dummyvariabler er imidlertid forklaringskraften til denne variabelen svært beskjeden. Disse funnene fraviker fra studien av Escribano-Navas & Gemar (2021) som fant at selskaper med kvinnelige ledere har en redusert sjanse for å gå konkurs.

5.1.2 Modell 1B

Tabell 10 gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene I og II for å inngå i Modell 1B. Tabellen viser også koeffisientverdier og z-verdier når de valgte variablene blir benyttet i en logistisk regresjon, samt modellens tilpasning *in-sample* og hvor godt den predikerer *out-of-sample*. Siden variablene som inngår i variabelsett II

kun er tilgjengelig fra 2014 vil denne modellen kun estimeres for de tre årene 2018, 2019 og 2020, som tilsvarer fold 9, 10 og 11 som vist i Figur 3.

Tabell 10: Resultater modell 1B

Tabellen gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene I og II, og som er benyttet i en logistisk regresjon. Tabellen viser koeffisientverdiene fra den logistiske regresjonen med tilhørende z-verdier i parentes, samt modellens tilpasning *in-sample* ved Brier score, AUC og AIC, samt hvor godt den predikerer *out-of-sample* ved Brier score, AUC og nøyaktighetsratio. Dette presenteres for valideringsårene 2018, 2019 og 2020, som utgjør fold 9, 10 og 11 som vist i Figur 3.

	2018	2019	2020
log_alder_dl	-0,80(-10,02)	-0,88(-11,48)	-0,86(-10,94)
log_alder_sl	-0,53(-4,66)	-0,44(-4,01)	-0,35(-3,07)
log_styre_alder_snitt	-0,81(-6,25)	-0,93(-7,42)	-1,02(-7,91)
log_styre_antall	-0,56(-13,68)	-0,35(-13,04)	-0,33(-11,89)
d_dl_in_styre	-0,30(-7,08)	-0,28(-8,37)	-0,19(-5,85)
d_to_dl	-2,15(-5,68)	-1,35(-5,37)	-0,93(-4,73)
d_dl_fylke	-0,51(-12,97)	-0,55(-13,67)	-0,52(-12,29)
aksj_std	0,73(3,24)	0,37(2,12)	0,43(2,42)
aksj_min	-0,03(-0,28)	-0,23(-4,40)	-0,23(-4,31)
d_sl_fylke	-0,59(-12,10)	-0,59(-12,09)	-0,59(-11,12)
styre_andel_fylke	-0,18(-3,04)	-0,13(-2,08)	-0,10(-1,44)
d_fravalg_revisjon	0,44(15,97)	0,44(16,80)	0,42(15,46)
d_bistand_regn_forer	0,22(7,55)	0,27(9,52)	0,28(9,80)
styre_alder_std	-0,01(-2,86)		
dl_eierandel	-0,27(-4,84)		
sl_eierandel	-0,28(-3,56)		
styre_eierandel	1,10(12,28)		
andel_st_medl_ikke_aksj	0,62(7,73)		
d_dl_kjonn_k	0,05(1,58)		
d_samme_dl_sl	-0,14(-3,46)		
log_aksj_antall	0,13(2,80)		
(Intercept)	4,54(18,60)	5,51(25,51)	5,25(23,96)
In-sample Brier score	0,024	0,025	0,023
In-sample AUC	0,711	0,706	0,698
AIC	66 679	70 713	69 045
Out-of-sample Brier score	0,026	0,020	0,015
Out-of-sample AUC	0,668	0,659	0,668
Nøyaktighetsratio	0,336	0,318	0,335

I Tabell 10 fremkommer det at variablene *log_alder_dl*, *log_alder_sl*, *log_styre_alder_snitt* og *d_dl_i_styre*, som ble valgt ut for alle år i modell 1A, også velges i denne modellen. Variablene har z-verdier med absoluttverdi større enn 3,29, noe som innebærer et signifikansnivå på 0,1 %. Fortegnene er negative, noe som sammenfaller med resultatet fra modell 1A. Variabelen *log_styre_alder_snitt* blir valgt ut først av LASSO i to av årene, og *log_alder_sl* velges først i ett av årene, som vist i Appendiks E. Videre er variablene *log_styre_antall* og *d_to_dl* også valgt ut i alle regnskapsårene, har negative fortegn, og er med sine z-verdier med absoluttverdi større enn 3,29 signifikante på 0,1 %-nivå slik som for modell 1A.

Variablene *d_dl_fylke* og *d_sl_fylke* angir henholdsvis om daglig leder og styrets leder har bostedsadresse i samme fylke som selskapets forretningsadresse. Videre angir *styre_andel_fylke* hvor stor andel av styremedlemmene som bor i samme fylke som selskapet. Disse tre variablene har alle negative fortegn, noe som indikerer at risikoen for konkurs reduseres dersom personene som innehar disse rollene er bosatt i samme fylke som selskapet. Dette virker fornuftig, da det kan tenkes at fysisk nærhet til selskapet bidrar til økt oversikt og kontroll. Koeffisientene for daglig leder og styrets leder er med z-verdier med absoluttverdi større enn 3,29 signifikante på 0,1 %-nivå, men for styret er denne kun signifikant på 1 %-nivå i 2018 og 5 %-nivå i 2019, har z-verdier med absoluttverdi større enn henholdsvis 2,58 og 1,96. Den er imidlertid ikke signifikant i 2020. Forøvrig er variabelen *d_dl_fylke* valgt ut nest først av LASSO i to av årene, og som tredje variabel i ett av årene, noe som understreker viktigheten av denne variabelen, som vist i Appendiks E.

Videre har variabelen *aksj_std*, standardavvik eierandeler aksjonærer ved balansedato, z-verdi med absoluttverdi større enn 2,58 i 2018 og større enn 1,96 i 2019 og 2020, som tilsvarer signifikansnivå på henholdsvis 1 % og 5 %. Fortegnet er positivt, noe som indikerer at dersom det er stor spredning i størrelsen på eierandelene vil dette kunne øke risikoen for konkurs. Dette virker fornuftig, da det kan tenkes at dersom aksjene er spredt på mange personer vil hver enkelt ta et mindre ansvar og i så måte vil risikoen for konkurs øke. Variabelen *aksj_min*, eierandel hos minste aksjonær ved balansedato, har z-verdi med absoluttverdi større enn 3,29, og er med det signifikant på 0,1 %-nivå i 2019 og 2020. Den er imidlertid ikke signifikant i 2018. Variabelen har negative fortegn, noe som betyr at dersom eierandelen til den minste aksjonæren øker, vil modellen estimere mindre risiko for konkurs. Dette virker fornuftig, siden stor eierandel hos minste aksjonær sannsynligvis betyr at det er færre aksjonærer, noe som kan føre til at hver enkelt føler et større ansvar. Dette har likhetstrekk med funnene til Tang et al. (2020), som fant at andelen aksjer eid totalt av de ti største aksjonærene hadde stor betydning for konkursprediksjon. Dersom det er få aksjonærer, og aksjonæren eier en stor andel av aksjene i et selskap, vil vedkommende kunne påvirkes sterkere av selskapets økonomiske situasjon, og vil ofte ha en større innvirkning

på den overordnede planen for selskapet og dermed ta større ansvar.

Variablene *d_fravalg_revisjon*, dummy for om selskapet har valgt bort ekstern revisor, og *d_bistand_regn_forer*, dummy for om regnskapsfører har bistått i regnskapet, er begge signifikante på 0,1 %-nivå med sine z-verdier med absoluttverdi større enn 3,29. Med positivt fortegn vil dette indikere at dersom selskapet har valgt bort ekstern revisor vil dette kunne øke risikoen for konkurs. Dette er i strid med funnene til Altman et al. (2010) som fant at sannsynligheten for konkurs øker dersom selskapet blir revidert, noe de forklarer med at en revisor trolig vil gripe inn og stanse driften i insolvente selskaper. Det at våre funn peker mot det motsatte kan likevel forklares med at selskaper som har mulighet til å velge bort revisjon er begrenset til driftsinntekter opp til NOK 6 millioner, balansesum opp til NOK 23 millioner og gjennomsnittlig antall ansatte opp til 10 (Forskrift om terskelverdier, 2018), og er dermed mindre enn selskaper som ikke har mulighet til å velge vekk revisjon. Siden risikoen for konkurs synes å avta ved økende selskapsstørrelse, (se utledning av størrelsesvariabelen *log.tk* fra variabelsett SEBRA i kapittel 2.1.1), kan det derfor tenkes at det er en størrelseseffekt som blir fanget opp av variabelen *fravalg_revisjon*. I tillegg finner Cenciarelli et al. (2018) at risikoen for konkurs avtar ved bruk av revisorer med ekspertise innenfor bransjen selskapet driver i. Dette forklarer de ved at en revisor vil kunne sammenligne selskapets prestasjoner med lignende selskaper og på den måten fange opp faresignalene tidlig nok til at ledelsen kan iverksette de tiltak og endringer som kan være nødvendige for å unngå konkurs.

Mer overraskende er det at variabelen *d_bistand_regn_forer* har positivt fortegn, noe som indikerer at risikoen for konkurs synes å øke dersom regnskapsfører har bistått i regnskapet. Her kan det imidlertid bemerkes at variabelen *d_utarbeidet_regn_forer*, som angir om regnskapsfører har *utarbeidet* regnskapet ikke er valgt ut av LASSO, og at det derfor gjelder kun tilfeller med sporadisk bistand. Likevel kan det tenkes at selskaper som fører regnskapet selv oppnår en bedre økonomistyring ved at de som fører regnskapet er ansatt i bedriften og dermed har intern innsikt i selskapet.

Videre er det åtte variabler som kun blir valgt ut i 2018. En mulig forklaring på at det

er så mange flere variabler som velges ut i 2018 sammenlignet med to påfølgende årene kan være at det er en større andel manglende verdier i tidligere regnskapsår. Variabelen *styre_alder_std*, standardavvik alder på styremedlemmer, har z-verdi med absoluttverdi større enn 2,58 og dermed signifikant på 1 %-nivå. Fortegnet er negativt, noe som indikerer at dersom spredningen i alderen til styrets medlemmer øker, vil dette redusere risikoen for konkurs. Dette sammenfaller med funnene i modell 1A.

Variabelen *styre_eierandel* er med z-verdi med absoluttverdi større enn 3,29 signifikant på 0,1 %-nivå. Fortegnet er positivt, noe som indikerer at dersom styremedlemmenes samlede eierandel øker, vil risikoen for konkurs øke. Dette peker i motsatt retning av studien til Platt og Platt (2012), hvor de fant at en økt andel eksterne aksjonærer økte risikoen for konkurs. Imidlertid blir variabelen kun utvalgt for det ene regnskapsåret. En mulig forklaring kan være at det for små selskaper ofte vil være få aksjonærer, og trolig få aksjonærer som ikke sitter i styret, sammenlignet med større selskaper. Det kan derfor tenkes at det også er en størrelseseffekt som fanges opp av denne variabelen, da risikoen for konkurs er forventet å avta med økende selskapsstørrelse (se utledning av størrelsesvariabelen *log_tk* fra variabelsett SEBRA i kapittel 2.1.1).

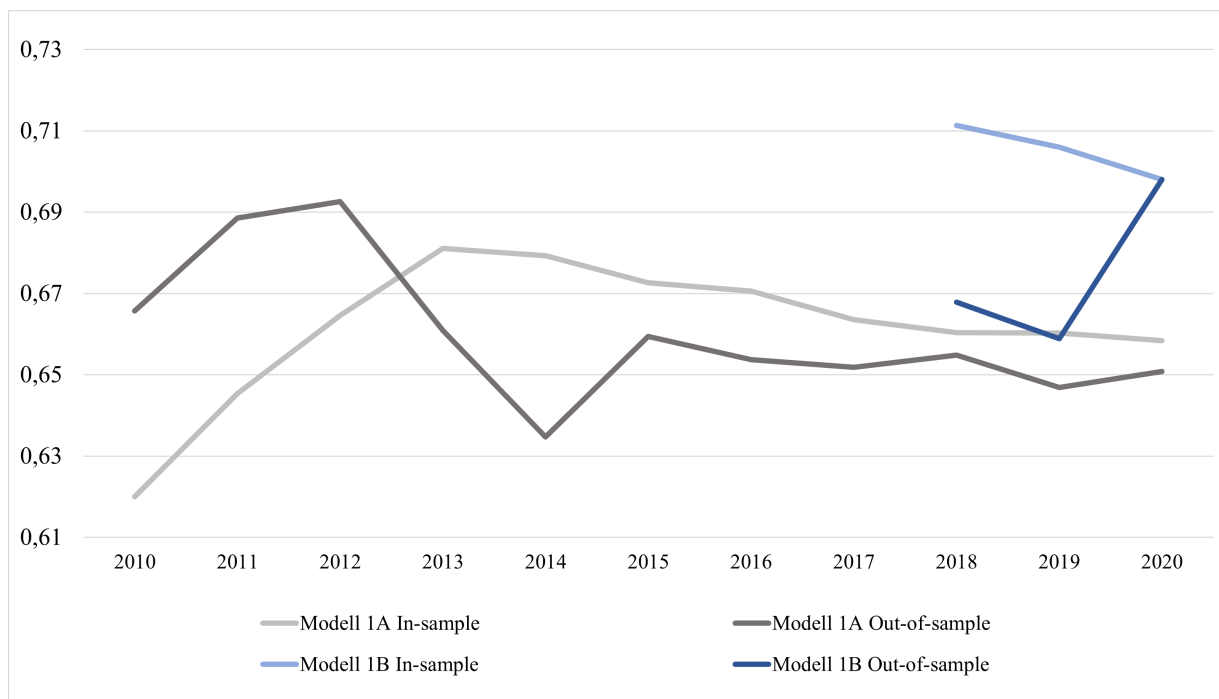
Variablene *dl_eierandel*, daglig leders eierandel, og *sl_eierandel*, styreleders eierandel, er med sine z-verdier med absoluttverdi større enn 3,29 signifikant på 0,1 %-nivå. Fortegnene er negative, noe som indikerer at dersom eierandelen til daglig leder eller styrets leder øker, vil risikoen for konkurs reduseres. Dette virker fornuftig, da disse nøkkelpersonene ofte vil ha mye å tape dersom selskapet går konkurs, og dermed vil jobbe aktivt for å hindre at dette skjer. Tilfellet hvor daglig leder og styrets leder er samme person, CEO-dualitet, blir presentert gjennom variabelen *d_samme_dl_sl*. Med negative fortegn og z-verdi med absoluttverdi større enn 3,29, tyder dette på at sannsynligheten for konkurs reduseres dersom dette er et scenario, og at funnet er signifikant på 0,1 %-nivå. Dette virker fornuftig, og flere studier peker på at CEO-dualitet kan gi et fleksibelt og effektivt lederskap (Ciampi, 2015; Dowell et al., 2011; Lilienfeld-Toal & Ruenzi, 2014). Det sammenfaller også med funnene i modell 1A.

Variabelen *log_aksj_antall* angir den logaritmiske transformasjonen av antall aksjonærer ved årets slutt er med sin z-verdi på 2,80 signifikant på 1 %-nivå. Fortegnet er positivt, noe som indikerer at dersom antall aksjonærer øker, vil risikoen for konkurs øke. Dette kan knyttes til studien til Tang et al. (2020), hvor de fant at variabelen for andel aksjer eid totalt av de ti største aksjonærene var av stor betydning for prediksjon av konkurs. Dersom den samlede eierandelen blant de ti største aksjonærene er høy vil dette også kunne indikere at det er få aksjonærer i selskapet. Om aksjene er spredt på få personer vil hver enkelt kunne ha mer å tape, sammenlignet med selskaper hvor risikoen er spredt over flere aksjonærer, og det kan derfor tenkes at hver enkelt vil ta et større ansvar.

Variabelen *andel_st_medl_ikke_aksj*, andelen styremedlemmer som ikke er aksjonærer, har z-verdi med absoluttverdi større enn 3,29 og dermed signifikant på 0,1 %-nivå. Fortegnet er positivt, noe som indikerer at risikoen for konkurs øker dersom andelen styremedlemmer som ikke er aksjonærer øker. En mulig forklaring på dette kan være at styremedlemmer som også er aksjonærer har større incentiv for å unngå konkurs, sammenlignet med styremedlemmer som ikke er aksjonærer. Det kan avslutningsvis nevnes at variabelen som omhandler om hvorvidt daglig leder er kvinne, *d_dl_kjonn_k* blir valgt ut av LASSO, men kan ikke anses som et signifikant funn da variabelen har en z-verdi på 1,58.

5.1.3 Modellevaluering

Figur 6 viser en sammenligning av AUC-verdiene til modell 1A og 1B. Det fremkommer at modell 1B har en bedre prestasjon enn modell 1A basert på AUC, men det må presiseres at det kun er sammenligningsgrunnlag for valideringsårene 2018, 2019 og 2020. Hverken modell 1A eller 1B kan betegnes som akseptable modeller, noe som tyder på at de ikke-finansielle variablene alene ikke fanger opp nok informasjon til å kunne predikere konkurs på en god måte.



Figur 6: Sammenligning av AUC-verdier for modell 1A og 1B

Figuren viser en sammenligning av AUC-verdiene til modell 1A og 1B. Den lysegrå og mørkegrå linjen viser henholdsvis *in-sample* og *out-of-sample* AUC til modell 1A for de ulike årene i analyseperioden 2010 til 2020. Den lyseblå og mørkeblå linjen viser henholdsvis *in-sample* og *out-of-sample* AUC til modell 1B for analyseårene 2018, 2019 og 2020.

Når det gjelder nøyaktighetsratio ligger denne verdien mellom 0,269 og 0,385 for modell 1A, og mellom 0,318 og 0,336 for modell 1B. Dette er lave verdier, noe som er forventet siden disse verdiene henger sammen med AUC. Videre har modell 1A en *in-sample* Brier score mellom 0,023 og 0,035, mot 0,023 og 0,025 for modell 1B. *Out-of-sample* Brier score er mellom 0,015 og 0,030 for modell 1A, mot 0,015 og 0,026 for modell 1B. Disse verdiene ønskes så nær 0 som mulig, og modell 1B kommer derfor noe bedre ut enn modell 1A basert på dette kriteriet. Modell 1A har AIC-verdier i området 66 679 til 70 713 for perioden 2018 til 2020, hvor til sammenligning modell 1B har verdier mellom 68 568 og 72 539 for tilsvarende år. Dette er et mål som straffer antall variabler som er inkludert i modellen, og modell 1B kommer likevel ut med et lavere tall enn modell 1A for alle de tre årene. Dette tyder på at de ekstra variablene som er lagt til i modell 1B gir ekstra forklaringskraft ved prediksjon av konkurs.

5.2 Modell 2

5.2.1 Modell 2A

Modell 2A består av variabelsett Altman. For denne modellen er det ikke benyttet LASSO for utvelgelse av variablene, og variabelsettet er benyttet i sin helhet i en logistisk regresjon. Tabellen viser koeffisientverdier og z-verdier fra den logistiske regresjonen, samt modellens tilpasning *in-sample* og hvor godt den predikerer *out-of-sample*. Dette presenteres for hver fold som vist i Figur 3.

Tabell 11: Resultater modell 2A

Tabellen gir en oversikt over variablene i variabelsett Altman benyttet i en logistisk regresjon. Tabellen viser koeffisientverdiene fra den logistiske regresjonen med tilhørende z-verdier i parentes, samt modellens tilpasning *in-sample* ved Brier score, AUC og AIC, samt hvor godt den predikerte *out-of-sample* ved Brier score, AUC og nøyaktighetsratio. Dette presenteres for hver fold som vist i Figur 3.

	2010	2011	2012	2013	2014
arbkap.tk	-0,71(-21,94)	-0,75(-23,38)	-0,80(-24,05)	-0,84(-24,94)	-0,93(-27,21)
opptj_ek.tk	-0,37(-14,74)	-0,39(-16,22)	-0,38(-15,79)	-0,40(-16,72)	-0,36(-14,75)
EBIT.tk	-1,99(-42,20)	-1,86(-40,49)	-1,77(-37,65)	-1,69(-34,87)	-1,63(-33,86)
ek_gjeld	-0,00(-2,84)	-0,00(-3,06)	-0,00(-2,34)	-0,00(-2,42)	-0,00(-3,42)
salginntekt.tk (intercept)	0,13(24,44) -4,26(-234,93)	0,13(24,44) -4,26(-234,93)	0,14(25,97) -4,43(-237,15)	0,16(28,19) -4,50(-237,21)	0,16(29,06) -4,54(-238,86)
In-sample Brier score	0,033	0,033	0,030	0,028	0,027
In-sample AUC	0,808	0,809	0,810	0,810	0,809
AIC	65 140	66 609	62 453	60 714	60 021
Out-of-sample Brier score	0,031	0,030	0,028	0,027	0,025
Out-of-sample AUC	0,803	0,810	0,813	0,815	0,797
Nøyaktighetsratio	0,606	0,620	0,627	0,631	0,593

	2015	2016	2017	2018	2019	2020
arbkap.tk	-0,97(-28,10)	-1,01(-28,93)	-1,03(-30,03)	-0,98(-29,12)	-0,99(-30,56)	-0,97(-29,76)
opptj_ek.tk	-0,31(-12,74)	-0,28(-11,38)	-0,26(-10,65)	-2,83(-11,84)	-0,30(-12,99)	-0,30(-13,11)
EBIT.tk	-1,62(-33,75)	-1,56(-32,52)	-1,50(-31,76)	-1,48(-31,53)	-1,41(-31,18)	-1,39(-30,29)
ek_gjeld	-0,00(-2,73)	-0,00(-2,98)	-0,00(-2,49)	-0,00(-1,63)	-0,00(-0,95)	-0,00(-1,51)
salginntekt.tk (intercept)	0,17(31,54) -4,45(-241,95)	0,18(32,39) -4,63(-243,45)	0,18(32,89) -4,65(-246,63)	0,18(34,46) -4,69(-250,05)	0,19(36,53) -4,50(-237,21)	0,19(35,71) -4,75(-256,79)
In-sample Brier score	0,026	0,025	0,024	0,024	0,024	0,023
In-sample AUC	0,807	0,805	0,803	0,800	0,803	0,804
AIC	60 923	60 575	61 368	62 541	65 962	64 086
Out-of-sample Brier score	0,024	0,025	0,024	0,025	0,020	0,014
Out-of-sample AUC	0,802	0,802	0,796	0,805	0,810	0,831
Nøyaktighetsratio	0,605	0,604	0,593	0,611	0,621	0,663

Tabell 11 viser at alle variablene med unntak av *ek_gjeld* har z-verdier med absoluttverdi større enn 3,29 og dermed har et signifikansnivå på 0,1 % for alle år. Med unntak av *salgsinntekt.tk* har alle variablene negative fortegn, noe som indikerer at dersom de øker så vil risikoen for konkurs reduseres. Dette er som forventet, og på bakgrunn av økonomisk teori

gjelder denne forventningen om negative fortegn for alle variablene i dette variabelsettet, slik drøftet i kapittel 2.1. For variabelen *salgsinntekt.tk*, salgsinntekter som andel av sum eiendeler, er fortegnet positivt. Dette er i strid med forventningen om negativt fortegn, men koeffisientverdiene er imidlertid ikke veldig høye. En mulig forklaring kan likevel være at dersom selskapet har lav total kapital i forhold til salgsinntekter, kan selskapet tenkes å være mindre robust for nedgangstider og perioder med lavere salgsinntekter, sammenlignet med selskaper som har en høyere andel total kapital i forhold til salgsinntekt. Videre viser Altman (2000) til at denne variabelen måler kapitalens omløpshastighet, og at den derfor er industri-sensitiv med store variasjoner fra bransje til bransje.

Variabelen *ek.gjeld*, sum egenkapital som andel av sum gjeld, har z-verdi med absoluttverdi større enn 2,58 og er signifikant på 1 %-nivå i årene 2010, 2011, og 2014 - 2016, og på 5 %-nivå i årene 2012, 2013 og 2017, hvor z-verdiene har absoluttverdi større enn 1,96. Verdien til koeffisientene er svært lave i alle år.

5.2.2 Modell 2B

Tabell 12 gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene Altman og I for å inngå i Modell 2B. Tabellen viser også koeffisientverdier og z-verdier når de valgte variablene blir benyttet i en logistisk regresjon, samt modellens tilpasning *in-sample* og hvor godt den predikerer *out-of-sample*. Dette presenteres for hver fold som vist i Figur 3.

Tabell 12: Resultater modell 2B

Tabellen gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene Altman og I, og som er benyttet i en logistisk regresjon. Tabellen viser koeffisientverdiene fra den logistiske regresjonen med tilhørende z-verdier i parentes, samt modellens tilpasning *in-sample* ved Brier score, AUC og AIC, samt hvor godt den predikerer *out-of-sample* ved Brier score, AUC og nøyaktighetsratio. Dette presenteres for hver fold som vist i Figur 3.

	2010	2011	2012	2013	2014
log_styre_alder_snitt		-0,89(-9,71)	-1,52(-24,04)	-1,53(-23,73)	-0,94(-9,13)
log_alder_sl		-0,71(-8,00)			-0,84(-8,62)
log_styre_antall					
d_dl_i_styre		-0,57(-23,38)	-0,67(-26,54)	-0,81(-31,62)	-0,78(-29,76)
log_alder_dl					
opptj_ek_tk	-0,65(-31,41)	-0,39(-16,11)	-0,38(-15,70)	-0,40(-16,56)	-0,37(-15,02)
EBIT_tk	-2,12 (46,02)	-1,78(-38,64)	-1,68(-35,48)	-1,54(-32,12)	-1,50(31,07)
salginntekt_tk		0,11(20,96)	0,13(22,43)	0,14(24,43)	0,14(24,92)
arbkap_tk		-0,74(-22,92)	-0,79(-23,63)	-0,84(-24,65)	-0,92(-26,53)
(intercept)	-4,08(-312,65)	-2,19(8,62)	1,84(7,52)	1,87(7,52)	2,84(11,16)
In-sample Brier score	0,034	0,033	0,030	0,028	0,027
In-sample AUC	0,802	0,817	0,820	0,824	0,825
AIC	66 290	65 602	61 284	59 281	58 595
Out-of-sample Brier score	0,031	0,029	0,027	0,026	0,024
Out-of-sample AUC	0,794	0,829	0,832	0,821	0,798
Nøyaktighetsratio	0,588	0,658	0,664	0,642	0,596

	2015	2016	2017	2018	2019	2020
log_styre_alder_snitt	-0,98(-8,80)	-0,70(-5,72)	-0,55(-4,28)	-0,71(-5,40)	-0,84(-6,54)	-1,03(-7,79)
log_alder_sl	-0,70(-6,53)	-0,70(-5,93)	-0,73(-6,39)	-0,56(-4,86)	-0,48(-4,32)	-0,35(-3,07)
log_styre_antall	-0,41(-17,51)	-0,43(-18,12)	-0,37(-15,34)	-0,40(-16,39)	-0,38(-16,11)	-0,35(-14,65)
d_dl_i_styre	-0,72(-27,27)	-0,56(-20,21)				
log_alder_dl		-0,50(-6,27)	-0,64(-7,67)	-0,67(-8,12)	-0,67(-8,48)	-0,59(-7,38)
opptj_ek_tk	-0,36(-14,60)	-0,34(-13,90)	-0,36(-14,58)	-0,39(-16,16)	-0,41(-17,58)	-0,41(-17,51)
EBIT_tk	-1,56(-32,35)	-1,52(-31,53)	-1,52(-31,84)	-1,48(-31,47)	-1,40(-30,88)	-1,38(-30,00)
salginntekt_tk	0,15(26,35)	0,15(26,33)	0,15(26,57)	0,15(27,73)	0,15(29,12)	0,15(28,50)
arbkap_tk	-0,88(-25,51)	-0,90(-25,72)	-0,89(-26,01)	-0,84(-24,94)	-0,85(-26,26)	-0,84(-25,72)
(intercept)	2,67(11,00)	3,28(13,05)	3,03(12,87)	3,04(13,27)	3,27(14,97)	3,11(14,07)
In-sample Brier score	0,026	0,025	0,024	0,024	0,024	0,023
In-sample AUC	0,825	0,825	0,821	0,819	0,821	0,822
AIC	59 367	59 120	60 085	61 159	64 420	62 660
Out-of-sample Brier score	0,024	0,024	0,024	0,025	0,020	0,014
Out-of-sample AUC	0,818	0,816	0,818	0,822	0,824	0,843
Nøyaktighetsratio	0,636	0,632	0,635	0,643	0,649	0,687

Tabell 12 viser at variabelen *log_alder_sl*, logaritmen av styreleders alder, velges i åtte av elleve regnskapsår. Denne er signifikant på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29, og har negative fortegn, noe som kan tolkes som at dersom alderen til styrets leder øker, vil risikoen for konkurs reduseres. Dette er fornuftig, da økt alder kan bety mer erfaring. Videre er variabelen *log_styre_alder_snitt*, den logaritmiske transformasjonen av gjennomsnittsalderen til styret, utvalgt i alle regnskapsår unntatt 2010, og er med det den av de ikke-finansielle variablene som velges ut i flest regnskapsår. I tillegg blir den valgt ut som første eller andre ikke-finansielle variabel i alle årene den er inkludert, som vist i Appendix E. Den har z-verdier med absoluttverdi større enn 3,29 og dermed signifikant på

0,1 %-nivå. Fortegnene er negative, noe som indikerer at dersom gjennomsnittsalderen på styremedlemmene øker, forventes risikoen for konkurs å reduseres. Dette virker fornuftig, da det kan knyttes til at eldre styremedlemmer trolig har mer erfaring, og at økt erfaring kan bidra til bedre økonomisk styring. Dette speiler funnene til Elloumi et al. (2001), som fant at gjennomsnittsalderen til styret var blant topp ti av de viktigste ikke-finansielle variablene for å forklare økonomisk mislighold.

Variabelen *log_styre_antall*, den logaritmiske transformasjonen av antall styremedlemmer, er valgt ut i regnskapsårene 2015 - 2020, og er med sine z-verdier med absoluttverdi større enn 3,29 signifikant på 0,1 %-nivå. Denne variabelen har også negative fortegn, og indikerer dermed at risikoen for konkurs er forventet å reduseres dersom antall styremedlemmer øker, og samsvarer i så måte med funnene til Platt & Platt (2012) som fant at økt størrelse på styret hadde en positiv effekt på selskapers økonomiske forhold. Dette støttes også av funnene til Manzanque et al. (2016), som fant at en økt mengde styremedlemmer reduserte risikoen for økonomisk mislighold.

Dummyvariabelen *d_dl_i_styre*, for om daglig leder sitter i styret, enten som styreleder eller som medlem, blir valgt ut i seks av elleve regnskapsår, og er den av de ikke-finansielle variablene som blir valgt ut først av LASSO flest ganger, som vist i Appendiks E. Den har z-verdier med absoluttverdi større enn 3,29 og er dermed signifikant på 0,1 %-nivå. Den har negativt fortegn, noe som antyder at dersom daglig leder er styremedlem eller styreleder vil risikoen for konkurs reduseres. Dette virker fornuftig, da det kan tenkes at daglig leders tilstedeværelse i styret bidrar til økt oversikt og kontroll, og peker i retning av funnene til Ciampi (2015), som fant at dersom daglig leder også har rollen som styreleder reduserte dette sannsynligheten for økonomisk mislighold.

Variabelen *log_alder_dl*, logaritmen av daglig leders alder, er valgt i regnskapsårene 2016 til 2020. Den har z-verdier med absoluttverdi større enn 3,29 og er med det signifikant på 0,1 %-nivå. Fortegnet er negativt, noe som indikerer at risikoen for konkurs reduseres dersom alderen til daglig leder øker. Dette er i tråd med funnene til Platt & Platt (2012), som fant at risikoen for konkurs øker for selskaper som hadde en yngre daglig leder, sammen-

lignet med selskaper som hadde en eldre daglig leder. Dette virker fornuftig, da det trolig kan skyldes økt erfaring ved økt alder. Alle fortegnene til de ikke-finansielle variablene som blir valgt ut i modellen samsvarer dermed med resultatene fra modell 1A og 1B.

Når det gjelder variablene fra variabelsett Altman er de to variablene *opptj_ek.tk*, opptjent egenkapital som andel av sum eiendeler, og *EBIT.tk*, inntjening før renter og skatt som andel av sum eiendeler, valgt i alle regnskapsårene. De er signifikante på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29 og har negativt fortegn som forventet og drøftet i modell 2A. Videre er variablene *salgsinntekt.tk*, salgsinntekter som andel av sum eiendeler, og *arbkap.tk*, arbeidskapital som andel av sum eiendeler, valgt ut i alle regnskapsår med unntak av 2010, og er signifikant på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29. Det bemerkes at resultatene fra 2010 skiller seg ut sammenlignet med øvrige år, ved at det kun er to variabler som velges ut. En mulig forklaring kan være at 2010 er preget av manglende data, da de tidligere årene i datasettet inneholdt en større mengde manglende verdier. Variabelen *arbkap.tk* har negative fortegn, mens *salgsinntekt.tk* har positive fortegn for alle regnskapsår. Dette er i strid med forventningen om negativt fortegn, men sammenfaller med resultatet som diskutert i modell 2A. Variabelen *ek.gjeld*, sum egenkapital som andel av sum gjeld, som synes å ha minst betydning i modell 2A er ikke valgt i noen av regnskapsårene.

5.2.3 Modell 2C

Tabell 13 gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene Altman, I og II for å inngå i Modell 2C. Tabellen viser også koeffisientverdier og z-verdier når de valgte variablene blir benyttet i en logistisk regresjon, samt modellens tilpasning *in-sample* og hvor godt den predikerer *out-of-sample*. Siden variablene som inngår i variabelsett II kun er tilgjengelig fra 2014 vil denne modellen kun estimeres for de tre årene 2018, 2019 og 2020, som tilsvarende fold 9, 10 og 11 som vist i Figur 3.

Tabell 13: Resultater modell 2C

Tabellen gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene Altman, I og II, og som er benyttet i en logistisk regresjon. Tabellen viser koeffisientverdiene fra den logistiske regresjonen med tilhørende z-verdier i parentes, samt modellens tilpasning *in-sample* ved Brier score, AUC og AIC, samt hvor godt den predikerer *out-of-sample* ved Brier score, AUC og nøyaktighetsratio. Dette presenteres for valideringsårene 2018, 2019 og 2020, som utgjør henholdsvis fold 9, 10 og 11 som vist i Figur 3.

	2018	2019	2020
d_sl_fylke	-0,53(-10,70)	-0,44(-9,26)	-0.55(-10.28)
styre_andel_fylke	-0,17(-2,70)	-0,17(-2,86)	-0.04(-0.67)
d_fravalg_revisjon	0,53(18,73)	0,67(25,76)	0.51(18.03)
d_dl_fylke	-0,40(-9,66)	-0,44(-11,12)	-0.43(-10.02)
log_alder_dl	-0,59(-7,12)	-0,50(-6,42)	-0.53(-6.61)
log_alder_sl	-0,40(-3,44)	-0,55(-5,30)	-0.23(-2.03)
log_styre_alder_snitt	-0,61(-4,61)	-0,70(-5,70)	-0.91(-6.74)
styre_eierandel	0,28(8,86)	0,28(9,01)	
log_styre_antall	-0,31(-12,09)		-0.28(-10.88)
arbkap_tk	-0,80(-23,94)	-0,86(-27,03)	-0.81(-24.94)
opptj_ek_tk	-0,39(-16,10)	-0,34(-14,85)	-0.41(-17.31)
EBIT_tk	-1,45(-30,94)	-1,33(-29,52)	-1.36(-29.70)
salginntekt_tk	0,17(29,31)	0,18(33,84)	0.17(30.42)
(Intercept)	2,19(9,27)	2,64(11,52)	2.33(10.13)
In-sample Brier score	0,023	0,024	0,022
In-sample AUC	0,834	0,834	0,835
AIC	59 749	63 249	61 438
Out-of-sample Brier score	0,025	0,020	0,014
Out-of-sample AUC	0,824	0,826	0,845
Nøyaktighetsratio	0,648	0,651	0,691

Tabell 13 viser at variablene *d_dl_fylke* og *d_sl_fylke*, dummyer for om henholdsvis daglig leder og styreleder er bosatt i samme fylke som selskapet, og *styre_andel_fylke*, andelen styremedlemmer som er bosatt i samme fylke som selskapet, er valgt ut i alle regnskapsårene. Både *d_dl_fylke* og *d_sl_fylke* er signifikante på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29. Fortegnene er negative, noe som indikerer at risikoen for konkurs reduseres dersom daglig leder eller styreleder bor i samme fylke som selskapet har sin forretningsadresse. Dette virker fornuftig, da fysisk nærhet til selskapet kan tenkes å være et moment som bidrar til økt oversikt og kontroll. Variabelen *styre_andel_fylke* er signifikant på 1 %-nivå med z-verdier med absoluttverdi større enn 2,58, men kun i 2018 og 2019. Den peker i retning av at det er gunstig dersom en størst mulig andel av styret er bosatt i samme fylke som selskapet, men det må påpekes at koeffisientene har lave verdier og at variabelen kun er signifikant i to av årene. Disse funnene samsvarer med resultatene for disse fylkesvariablene i modell 1B .

Variabelen *d_fravalg_revisjon*, en dummy for om selskapet har valgt bort ekstern revisor, er

signifikant på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29, og inkluderes i alle tre årene. Den har positivt fortegn, noe som kan tyde på at dersom selskapet velger bort ekstern revisor vil risikoen for konkurs øke. Dette samsvarer med funnene og vurderingene som er gjort av disse i modell 1B.

Variablene *log_alder_dl*, *log_alder_sl* og *log_styre_alder_snitt*, de logaritmiske transformasjonene av alder til daglig leder, styreleder og gjennomsnittsalderen til styret, er signifikante på henholdsvis 0,1 % og 5 %-nivå med z-verdier med absoluttverdi større enn 3,29 og 1,96. De har negative fortegn, noe som peker mot at økt alder hos disse nøkkelpersonene er negativt assosiert med risikoen for konkurs, i likhet med funnene i modell 2B. Variabelen *log_styre_alder_snitt* er den av de ikke-finansielle variablene som blir valgt ut først av LASSO i alle tre årene, som vist i Appendiks E.

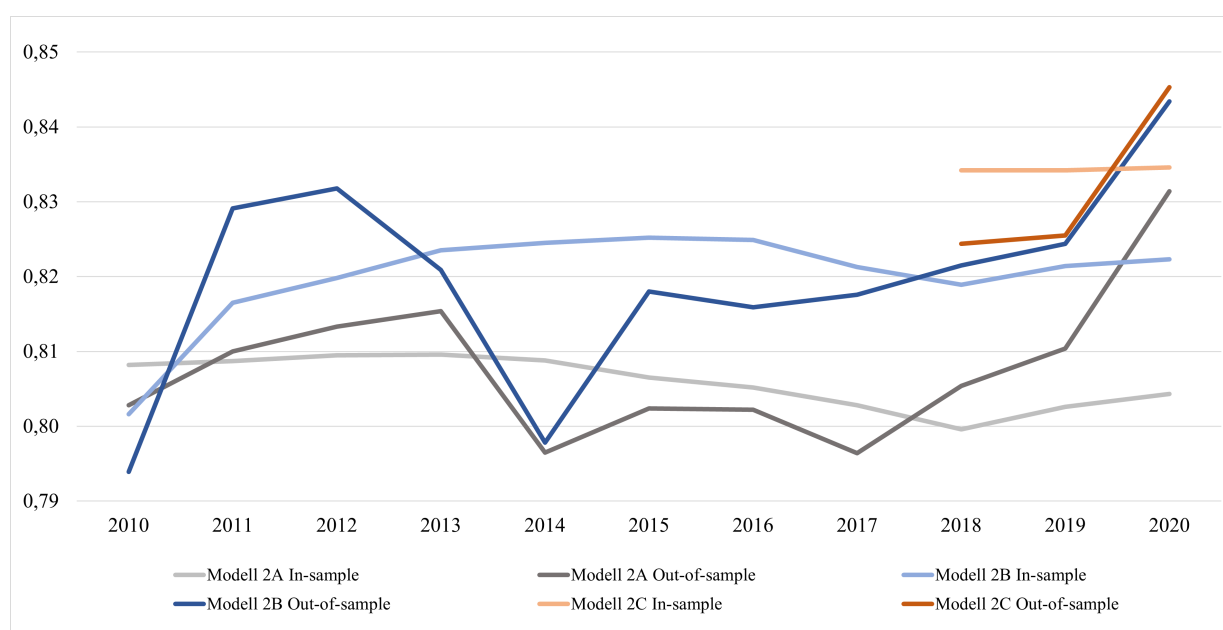
Videre blir variabelen *styre_eierandel*, samlet eierandel for alle styremedlemmer ved årets slutt, valgt ut for regnskapsårene 2018 og 2019. Den er signifikant på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29. Fortegnet er positivt, noe som indikerer at risikoen for konkurs øker ved økt eierandel blant styrets medlemmer. Dette samsvarer med funnene og argumentasjonen for dette i modell 1B. Variabelen *log_styre_antall*, logaritmen til antall styremedlemmer, blir valgt ut i 2018 og 2020. Den er signifikant på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29. Fortegnet er negativt, og indikerer at risikoen for konkurs avtar ved økende styrestørrelse. Dette er tilsvarende funn som i modell 2B, og det er gjort lignende funn i tidligere studier (Manzaneque et al., 2016; Platt & Platt, 2012).

I likhet med i modell 2B er det variablene *opptj_ek_tk*, opptjent egenkapital som andel av sum eiendeler, *EBIT_tk*, inntjening før renter og skatt som andel av sum eiendeler, *salgsinntekt_tk*, salgsinntekter som andel av sum eiendeler, og *arbkap_tk*, arbeidskapital som andel av sum eiendeler, som blir valgt ut fra variabelsett Altman. Variabelen *ek_gjeld*, sum egenkapital som andel av sum gjeld, blir ikke valgt i noen av årene. Alle de utvalgte variablene fra variabelsett Altman er signifikante på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29, og har, med unntak av variabelen *salgsinntekt_tk*, negativt fortegn som forventet,

og tilsvarer funnene i modell 2A og 2B.

5.2.4 Modellevaluering

Figur 7 viser en sammenligning av AUC-verdier for modell 2A, 2B og 2C, og det fremkommer av grafene at modell 2B jevnt over oppnår en noe bedre AUC enn modell 2A i de fleste regnskapsår. Best ut kommer modell 2C, men resultatene kan imidlertid ikke sammenlignes direkte da modell 2C kun er testet for tre av regnskapsårene.



Figur 7: Sammenligning av AUC-verdier for modell 2A, 2B og 2C

Figuren viser en sammenligning av AUC-verdiene til modell 2A, 2B og 2C. Den lysegrå og mørkegrå linjen viser henholdsvis *in-sample* og *out-of-sample* AUC til modell 2A for de ulike årene i analyseperioden 2010 til 2020. Den lyseblå og mørkeblå linjen viser henholdsvis *in-sample* og *out-of-sample* AUC til modell 2B for tilsvarende periode. Den lyseoransje og mørkeoransje viser henholdsvis *in-sample* og *ut-of-sample* til modell 2C for analyseårene 2018, 2019 og 2020.

Når det gjelder *Brier score* presterer modellene tilnærmet likt. Modell 2B har en noe bedre prestasjon hva gjelder *out-of-sample* *Brier score* enn modell 2A for flere av regnskapsårene, og alle tre modellene scorer likt i 2018, 2019 og 2020. Videre kommer modell 2B bedre ut enn modell 2A på prestasjonsmålet AIC for alle regnskapsår bortsett fra 2010, og modell 2C kommer best ut av de tre modellene i de tre årene den er estimert, 2018, 2019 og 2020. Dette er interessant fordi målet AIC gir en verdi basert på balansen mellom kompleksitet

og nøyaktighet, og det at modell 2C kommer best ut her taler for at prestasjonen øker ved å inkludere flere ikke-finansielle variabler. Når det kommer til nøyaktighetsratio kommer modell 2B bedre ut enn modell 2A i alle år bortsett fra 2010, og modell 2C kommer best ut av de tre modellene i årene 2018, 2019 og 2020, med verdier på henholdsvis 0,648, 0,651 og 0,691.

I følge evalueringmålene har modell 2B jevnt over en bedre prestasjon enn modell 2A. Videre har modell 2C en noe bedre prestasjon enn modell 2A og 2B, men det må tas i betraktning at modellen kun er testet for tre regnskapsår. Likevel tyder resultatene på at det å inkludere ikke-finansielle variabler gir økt forklaringskraft.

5.3 Modell 3

5.3.1 Modell 3A

Modell 3A består av variabelsett SEBRA. For denne modellen er det ikke benyttet LASSO for utvelgelse av variablene, og variabelsettet er benyttet i sin helhet i en logistisk regresjon. Tabellen viser koeffisientverdier og z-verdier fra den logistiske regresjonen, samt modellens tilpasning *in-sample* og hvor godt den predikerer *out-of-sample*. Dette presenteres for hver fold som vist i Figur 3.

Tabell 14: Resultater modell 3A

Tabellen gir en oversikt over variablene i variabelsett SEBRA benyttet i en logistisk regresjon. Tabellen viser koeffisientverdiene fra den logistiske regresjonen med tilhørende z-verdier i parentes, samt modellens tilpasning *in-sample* ved Brier score, AUC og AIC, samt hvor godt den predikerte *out-of-sample* ved Brier score, AUC og nøyaktighetsratio. Dette presenteres for hver fold som vist i Figur 3.

	2010	2011	2012	2013	2014
res_gjeld	-0,00(-0,56)	-0,00(-0,84)	-0,00(-0,13)	-0,00(-0,03)	-0,00(-1,02)
ek_tk	-0,63(-21,839)	-0,60(-21,55)	-0,57(-20,15)	-0,59(-20,59)	-0,60(-20,06)
d_bek_iek	1,42(45,06)	1,39(44,55)	1,38(41,98)	1,32(39,54)	1,27(37,82)
likv_kg_omset	-0,00(-13,36)	-0,00(-13,40)	-0,00(-11,81)	-0,00(-11,54)	-0,00(-11,55)
log_alder	-0,34(-30,44)	-0,34(-30,74)	-0,33(-28,29)	-0,34(-28,01)	-0,34(-29,46)
log_tk	-0,04(-3,44)	-0,03(-2,68)	-0,04(-3,38)	-0,04(-3,10)	-0,04(-2,97)
levgjeld_tk	1,86(37,06)	1,89(37,82)	1,93(37,27)	1,93(37,45)	2,03(38,12)
avg_tk	2,83(24,83)	2,83(25,39)	2,95(25,98)	3,06(26,87)	3,26(28,87)
(intercept)	4,05(-22,66)	-4,15(-23,06)	-4,15(-23,07)	-4,21(-22,80)	-4,27(-22,98)
In-sample Brier score	0,032	0,032	0,029	0,028	0,026
In-sample AUC	0,846	0,848	0,849	0,852	0,854
AIC	60 280	61 631	57 865	56 174	55 449
Out-of-sample Brier score	0,029	0,029	0,028	0,026	0,025
Out-of-sample AUC	0,849	0,853	0,854	0,857	0,849
Nøyaktighetsratio	0,697	0,707	0,707	0,714	0,698

	2015	2016	2017	2018	2019	2020
res_gjeld	-0,00(0,13)	-0,00(-0,38)	-0,00(0,60)	-0,00(1,57)	-0,00(2,70)	-0,00(1,82)
ek_tk	-0,53(-17,70)	-0,53(-17,40)	-0,51(-16,58)	-0,51(-17,09)	-0,54(-18,75)	-0,56(-19,18)
d_bek_iek	1,22(36,87)	1,15(34,55)	1,11(33,29)	1,07(32,33)	1,07(33,21)	1,07(32,51)
likv_kg_omset	-0,00(-11,92)	-0,00(-12,11)	-0,00(-11,91)	-0,00(-12,13)	-0,00(-13,39)	-0,00(-13,09)
log_alder	-0,36(-32,68)	-0,37(-34,06)	-0,37(-36,18)	-0,37(-31,64)	-0,36(-31,29)	-0,36(-30,64)
log_tk	-0,05(-4,15)	-0,07(-6,10)	-0,09(-7,23)	-0,11(-8,93)	-0,10(-8,84)	-0,09(-7,92)
levgjeld_tk	2,10(39,38)	2,05(37,98)	2,07(38,47)	2,06(38,49)	2,00(38,69)	1,97(37,33)
avg_tk	3,61(32,68)	3,71(34,06)	3,68(31,69)	3,87(36,40)	3,79(36,81)	3,73(35,71)
(intercept)	-4,10(-21,87)	-3,73(-19,78)	-3,51(-18,67)	-3,20(-17,10)	-3,24(-18,16)	-3,45(-19,23)
In-sample Brier score	0,026	0,024	0,024	0,023	0,024	0,022
In-sample AUC	0,854	0,854	0,855	0,855	0,854	0,854
AIC	56 066	55 869	56 450	57 572	60 933	59 368
Out-of-sample Brier score	0,023	0,024	0,023	0,024	0,019	0,014
Out-of-sample AUC	0,854	0,859	0,853	0,849	0,854	0,872
Nøyaktighetsratio	0,709	0,719	0,706	0,697	0,708	0,744

Tabell 14 viser at seks av åtte variabler er signifikant på 1 %-nivå med z-verdier med absoluttverdi større enn 2,58 i alle regnskapsårene i analyseperioden. Variabelen *log_tk* har z-verdier med absoluttverdi større enn 1,96 og er dermed signifikant på 5 %-nivå gjennom hele perioden. Variabelen *res_gjeld* skiller seg ut ved at den kun er signifikant i regnskapsår 2019 med en z-verdi på 2,70. Effekten av denne variabelen dette året er imidlertid begrenset på grunn av koeffisientens størrelse.

Variablene *avg_tk* og *levgjeld_tk* har begge positive fortegn. Dette indikerer at dersom skyldige offentlige avgifter i prosent av totalkapital eller leverandørgjeld i prosent av totalkapital øker, medfører dette økt risiko for konkurs. Dummyvariabelen *d_bek_iek* har også positive

fortegn, noe som indikerer at dersom bokført egenkapital er mindre enn innskutt egenkapital vil dette øke risikoen for konkurs. Fortegnene for disse variablene er som forventet basert på økonomisk teori, og slik utledet i kapittel 2.1.1.

Videre har variablene *ek_tk*, *res_gjeld*, *likv_kg_oms*, *log_alder* og *log_tk* alle negative fortegn. Variabelen *ek_tk*, egenkapital i prosent av totalkapital, indikerer at en høyere egenkapitalandel vil medføre en lavere risiko for konkurs. Videre for inntjeningsvariabelen *res_gjeld*, ordinært resultat før av- og nedskrivninger i prosent av total gjeld, så indikerer de negative fortegnene at risikoen for konkurs reduseres når forholdet mellom inntjening og betalingsforpliktelser øker. For *likv_kg_oms*, likvide midler minus kortsiktig gjeld i prosent av omsetning, indikerer de negative fortegnene at dersom andelen betalingsmidler øker, vil risikoen for konkurs reduseres. For variablene *log_alder* og *log_tk*, logaritmen av henholdsvis antall år siden etablering av selskapet og totalkapital, peker de negative fortegnene mot at dersom selskapets alder eller størrelse øker, vil dette medføre en redusert risiko for konkurs. De negative fortegnene for alle disse variablene er også som forventet ut i fra økonomisk teori og begrunnet i kapittel 2.1.1.

5.3.2 Modell 3B

Tabell 15 gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene SEBRA og I for å inngå i Modell 3B. Tabellen viser også koeffisientverdier og z-verdier når de valgte variablene blir benyttet i en logistisk regresjon, samt modellens tilpasning *in-sample* og hvor godt den predikerer *out-of-sample*. Dette presenteres for hver fold som vist i Figur 3.

Tabell 15: Resultater modell 3B

Tabellen gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene SEBRA og I, og som er benyttet i en logistisk regresjon. Tabellen viser koeffisientverdiene fra den logistiske regresjonen med tilhørende z-verdier i parentes, samt modellens tilpasning *in-sample* ved Brier score, AUC og AIC, samt hvor godt den predikerer *out-of-sample* ved Brier score, AUC og nøyaktighetsratio. Dette presenteres for hver fold som vist i Figur 3.

	2010	2011	2012	2013	2014
log_styre_alder_snitt	-0,47(-6.87)	-0,51(-7.68)	-0,73(-10,65)	-0,68(-9,85)	-0,66(-9,75)
d_dl_i_styre	-0,48(-19,22)	-0,59(-23,88)	-0,69(-27,23)	-0,85(-32,74)	-0,86(-32,31)
log_styre_antall					-0,12(-4,45)
log_alder_sl					
styre_alder_std		-0,03(-7,81)			-0,03(-6,90)
log_alder_dl					
d_samme_dl_sl					
d_to_dl					
ek_tk	-0,62(-21,61)	-0,60(-21,22)	-0,57(-20,16)	-0,60(-20,64)	-0,57(-19,47)
d_bek_iek	1,39(44,12)	1,36(43,39)	1,34(41,06)	1,27(38,28)	1,19(35,41)
log_alder	-0,32(-27,71)	-0,31(-26,58)	-0,28(-23,75)	-0,29(-24,36)	-0,29(-23,23)
levgjeld_tk	1,81(35,78)	1,81(36,18)	1,96(38,29)	2,00(38,15)	2,03(38,52)
avg_tk	2,69(23,50)	2,68(23,91)	3,34(31,40)	3,51(32,83)	3,41(30,75)
likv_kg_omset	-0,00(-12,57)	-0,00(-12,18)			
log_tk	-0,07(-6,53)	-0,06(-5,61)			-0,07(-5,83)
(intercept)	-1,42(-4,55)	-1,30(-4,26)	-1,70(-6,46)	-1,81(-6,81)	-0,65(-2,06)
In-sample Brier score	0,032	0,032	0,029	0,027	0,026
In-sample AUC	0,850	0,854	0,855	0,860	0,864
AIC	59 882	60 956	57 231	55 227	54 403
Out-of-sample Brier score	0,028	0,028	0,027	0,025	0,024
Out-of-sample AUC	0,856	0,865	0,864	0,859	0,849
Nøyaktighetsratio	0,712	0,730	0,728	0,718	0,698

	2015	2016	2017	2018	2019	2020
log_styre_alder_snitt	-0,54(-4,57)	-0,47(-3,85)	-0,31(-2,38)	-0,43(-3,24)	-0,49(-3,76)	-0,66(-4,88)
d_dl_i_styre	-0,75(-27,95)	-0,51(-13,24)	-0,42(-14,12)	-0,27(-8,78)		
log_styre_antall	-0,16(-5,46)	-0,25(-7,55)	-0,28(-11,25)	-0,31(-12,15)	-0,29(-11,92)	-0,29(-11,72)
log_alder_sl	-0,07(-0,60)	-0,26(-2,25)	-0,47(-4,06)	-0,29(-2,50)	-0,23(-2,11)	-0,16(-1,36)
styre_alder_std	-0,03(-6,44)	-0,03(-4,95)				
log_alder_dl			-0,11(-1,37)	-0,20(-2,47)	-0,32(-3,92)	-0,32(-3,95)
d_samme_dl_sl		-0,18(-4,32)				
d_to_dl		-2,00(-4,87)				
ek_tk	-0,52(-17,29)	-0,52(-17,04)	-0,51(-16,52)	-0,52(-17,07)	-0,58(-19,17)	-0,59(-20,49)
d_bek_iek	1,20(35,78)	1,14(33,99)	1,12(33,39)	1,09(32,81)	1,10(34,07)	1,06(32,32)
log_alder	-0,31(-25,24)	-0,30(-24,44)	-0,31(-24,92)	-0,31(-24,56)	-0,29(-22,94)	-0,28(-22,22)
levgjeld_tk	2,02(37,65)	1,98(36,52)	2,02(37,43)	2,01(37,69)	1,96(37,87)	2,04(39,36)
avg_tk	3,48(31,23)	3,57(32,41)	3,73(34,41)	3,70(34,57)	3,59(34,73)	4,01(40,92)
likv_kg_omset	-0,00(-10,87)	-0,00(-11,24)	-0,00(-11,12)	-0,00(-11,30)	-0,00(-12,54)	
log_tk	-0,07(-5,64)	-0,08(-6,40)	-0,07(-5,77)	-0,08(-6,37)	-0,06(-4,77)	
(intercept)	-0,80(-2,52)	-0,17(-0,57)	0,09(0,29)	0,25(0,79)	0,14(0,495)	-0,55(-2,34)
In-sample Brier score	0,025	0,024	0,024	0,023	0,024	0,022
In-sample AUC	0,863	0,862	0,860	0,858	0,857	0,856
AIC	55 106	55 061	55 963	57 148	60 490	59 153
Out-of-sample Brier score	0,023	0,023	0,023	0,024	0,019	0,014
Out-of-sample AUC	0,858	0,863	0,856	0,853	0,856	0,872
Nøyaktighetsratio	0,715	0,726	0,712	0,705	0,713	0,744

Tabell 15 viser at det kun er *log_styre_alder_snitt* av de ikke-finansielle variablene som blir valgt ut i alle regnskapsårene. Den er også den av de ikke-finansielle variablene som

blir valgt ut nest først av LASSO flest ganger, som vist i Appendiks E. Variabelen har z-verdier med absoluttverdi større enn 2,58, og dermed signifikant på 1 %-nivå gjennom hele perioden. Koeffisientene er negative, noe som indikerer at en høyere gjennomsnittsalder på personene som sitter i styret reduserer sjansen for å gå konkurs. Dette er som forventet, da det kan antas at et slikt styre vil inneha mer kompetanse og erfaring enn et tilsvarende ungt styre, noe som kan virke preventivt mot konkurs. En annen variabel som hyppig blir valgt ut som en variabel av betydning er *d_dl_i_styre*. Variabelen er med sine z-verdier med absoluttverdi større enn 3,29 signifikant på 0,1 %-nivå i alle periodene hvor denne variabelen inngår i modellen. I likhet med i modell 2B har variabelen også i denne modellen fått negative fortegn, som vil si at dersom daglig leder også sitter i styret reduserer dette sannsynligheten for at selskapet går konkurs. Denne variabelen er også den blant de ikke-finansielle variablene som blir valgt ut først av LASSO flest ganger, og den blir i flere av årene også valgt ut før SEBRA-variablene *likv_kg_oms* og *log_tk*, som vist i Appendiks E. Dette vitner om at denne ikke-finansielle variabelen er av stor betydning hva gjelder risikoen for konkurs.

Logaritmen av antall styremedlemmer, *log_styre_antall*, blir tatt med i årene 2014 til 2020, og med z-verdier med absoluttverdi større enn 3,29 er den signifikant på 0,1 %-nivå. Fortegnet er negativt og indikerer at dersom antall styremedlemmer øker, forventes risikoen for konkurs å reduseres. Variabelen *log_alder_sl*, logaritmen av styreleders alder, inkluderes i regresjonsmodellene for årene 2015 til 2020 med negativt fortegn og indikerer at høyere alder på styrets leder reduserer sannsynligheten for konkurs. Variabelen er imidlertid med z-verdier i området -0,60 til -4,06 kun signifikant i enkelte år, og betydningen er derfor varierende gjennom analyseperioden. Variablene *styre_alder_std* og *log_alder_dl* forekommer i fire av regnskapsårene med negative fortegn. Variablene *d_samme_dl_sl* og *d_to_dl* blir kun valgt ut som relevante i regnskapsåret 2010, med z-verdi med absoluttverdi større enn 3,29 som tilsvarer signifikansnivå på 0,1 %. Sistnevnte variabel har i 2010 en koeffisientverdi som er vesentlig høyere enn de resterende dummyvariablene dette året. Fortegnene til disse variablene er også negative, og indikerer at risikoen for konkurs reduseres dersom daglig

leder og styrets leder er samme person, eller dersom det er to i rollen som daglig leder i samme selskap. Fortegnene til samtlige av de ikke-finansielle variablene som blir valgt ut i modell 3B er dermed som forventet basert på resultatene i de øvrige modellene.

Av de utvalgte finansielle variablene har alle bortsett fra *likv_kg_omset* og *log_tk* z-verdier med absoluttverdi større enn 3,29 og er dermed signifikant på 0,1 %-nivå, og blir valgt ut for alle regnskapsår. Samtlige av de finansielle variablene har også i denne modellen fortegn som forventet, slik drøftet i modell 3A. Sammenlignet med modell 3A, som kun inkluderer de finansielle variablene basert på SEBRA-modellen, registreres det at variabelen *res_gjeld* ikke blir valgt ut av LASSO i modell 3B. Dette er ikke overraskende på bakgrunn av det varierende signifikansnivået og de lave regresjonskoeffisientene variabelen hadde i modell 3A. Dette vitner om at de ikke-finansielle variablene har erstattet forklaringskraften denne variabelen hadde i modell 3A.

5.3.3 Modell 3C

Tabell 16 gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene SEBRA, I og II for å inngå i Modell 3C. Tabellen viser også koeffisientverdier og z-verdier når de valgte variablene blir benyttet i en logistisk regresjon, samt modellens tilpasning *in-sample* og hvor godt den predikerer *out-of-sample*. Siden variablene som inngår i variabelsett II kun er tilgjengelig fra 2014 vil denne modellen kun estimeres for de tre årene 2018, 2019 og 2020, som tilsvarer fold 9, 10 og 11 som vist i Figur 3.

Tabell 16: Resultater modell 3C

Tabellen gir en oversikt over hvilke variabler som blir valgt av LASSO fra variabelsettene SEBRA, I og II, og som er benyttet i en logistisk regresjon. Tabellen viser koeffisientverdiene fra den logistiske regresjonen med tilhørende z-verdier i parentes, samt modellens tilpasning *In-sample* ved Brier score, AUC og AIC, samt hvor godt den predikerer *out-of-sample* ved Brier score, AUC og nøyaktighetsratio. Dette presenteres for valideringsårene 2018, 2019 og 2020, som utgjør fold 9, 10 og 11 som vist i Figur 3.

	2018	2019	2020
d_sl_fylke	-0.51(-10,31)	-0.54(-10.83)	-0.57(-10.57)
styre_andel_fylke	-0.16(-2,64)	-0.06(-0.98)	-0.03(-0.33)
d_fravalg_revisjon	0,23(7,34)	0.34(12.46)	0,24(8.67)
d_dl_fylke	-0,37(-8,97)	-0.41(-9.99)	-0,49(-9.45)
log_styre_antall	-0,27(-10,50)	-0.31(-12.56)	-0,26(-9.79)
log_alder_dl	-0,18(-2,12)	-0.33(-4.08)	-0,27(-3.24)
log_alder_sl	-0,20(-1,74)	-0.20(-1.75)	-0,08(-0.72)
log_styre_alder_snitt	-0,44(-3,26)	-0.53(-4.02)	-0,62(-4.55)
styre_eierandel	0,33(9,46)		0,26(3.94)
andel_st_medl_ikke_aksj			-0,00(-0.02)
d_bistand_regn_forer	0,07(2,44)		
d_dl_i_styre	-0,36(-10,68)		
d_to_dl	-1,97(-5,19)		
ek_tk	-0,54(-17,34)	-0.55(-19.06)	-0,59(-20.22)
d_bek_iek	1,08(32,15)	1.03(31.92)	1.09(32.82)
levgjeld_tk	2,03(37,53)	2.11(41.12)	1.95(36.86)
avg_tk	3,70(34,11)	4.12(42.79)	3.58(34.88)
log_alder	-0,28(-21,52)	-0.23(-18.48)	-0,26(-20.08)
likv_kg_omset	-0,00(-10,49)		-0,00(-12.01)
log_tk	-0,28(-21,52)		
(Intercept)	-0,37(-1,02)	-0.20(-0.85)	-0,56(-2.26)
In-sample Brier score	0,023	0,024	0,022
In-sample AUC	0,867	0,864	0,864
AIC	56 095	59 708	58 067
Out-of-sample Brier score	0,024	0,019	0,014
Out-of-sample AUC	0,854	0,856	0,871
Nøyaktighetsratio	0,708	0,713	0,742

Tabell 16 viser at variabelene *d_sl_fylke* og *d_dl_fylke* har z-verdier med absoluttverdi større enn 3,29, noe som gjør at de er signifikante på 0,1 %-nivå i samtlige år. Variabelen *styre_andel_fylke* blir også valgt ut i alle tre år, men er kun signifikant på 1 %-nivå i 2018 med absoluttverdi større enn 2,58, og ikke signifikant i øvrige år. De har alle negative fortegn, noe som indikerer at dersom henholdsvis daglig leder og styrets leder er bosatt i samme fylke som selskapet, eller dersom andelen styremedlemmer som bor i samme fylke som selskapet øker, vil dette redusere sannsynligheten for konkurs. Dette er et område som er lite undersøkt i litteraturen, men det må virke fornuftig at høy tilstedeværelse av selskapets øverste ledelse er positivt, og dette samsvarer også med funnene i modell 1B og 2C.

Videre har også variablene *log_alder_dl*, *log_alder_sl* og *log_styre_alder_snitt* negative for-

tegn, noe som indikerer at dersom henholdsvis alder på daglig leder, styrets leder eller gjennomsnittsalderen på styrets medlemmer øker, så reduseres sjansene for konkurs. Dette samsvarer med tidligere funn i de øvrige modellene og understreker at økende alder med antageligvis økende erfaring og kompetanse, vil virke preventivt på konkurs. Variablene *log_alder_dl* og *log_alder_sl* har imidlertid z-verdier med absoluttverdier som varierer mellom 0,72 og 4,08, og er dermed ikke signifikante i alle år. Variabelen *log_styre_alder_snitt* er med sine z-verdier med absoluttverdi større enn 3,29 signifikant på 0,1 %-nivå i alle tre årene.

Variabelen *d_fravalg_revisjon* er med sine høye z-verdier signifikant på 0,1 %-nivå i alle tre regnskapsårene. Den har positivt fortegn, noe som tyder på at dersom et selskap velger bort ekstern revisor vil dette øke sjansene for at selskapet går konkurs. Dette samsvarer med funnene og diskusjonen rundt denne variabelen gjort i modell 2C. Det påpekes at størrelsesvariabelen *log_tk* fra variabelsett SEBRA ikke blir valgt ut i denne modellen, noe som kan indikere at *d_fravalg_revisjon* fanger opp et størrelsesmoment i form av at selskaper som har mulighet til å velge vekk revisjon er mindre enn selskaper som ikke har mulighet til det.

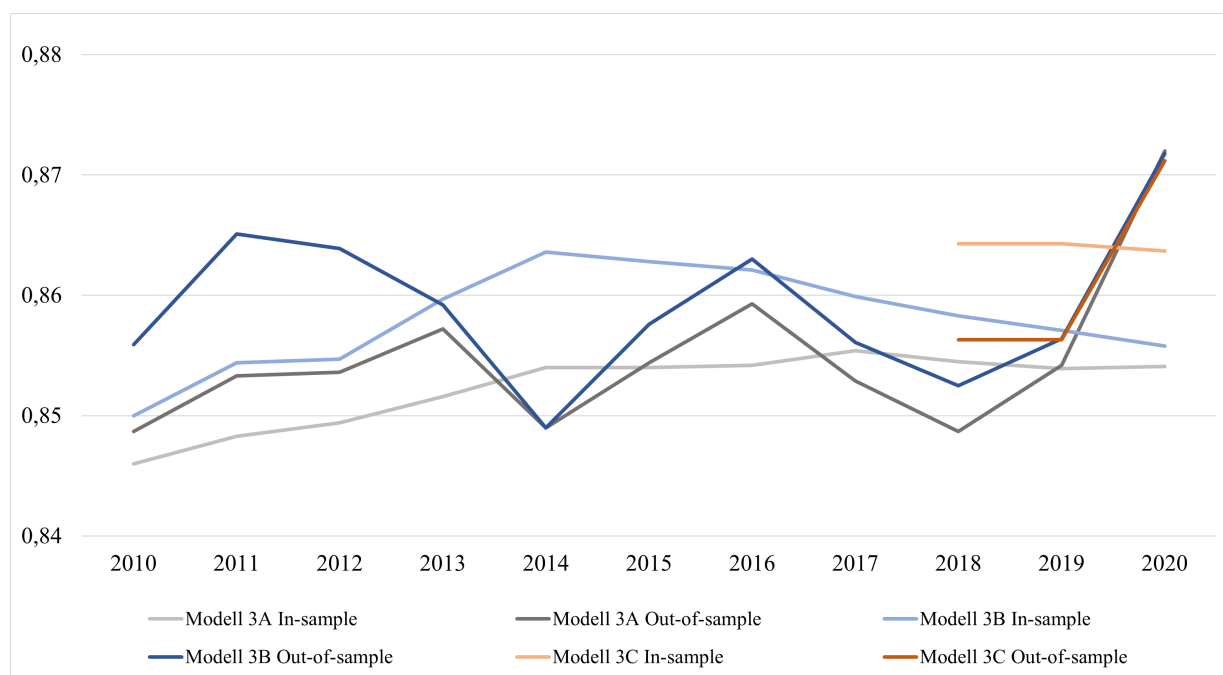
Variabelen *styre_eierandel* inkluderes i modellen i 2018 og 2020, og er signifikant på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29. Med positivt fortegn indikerer dette at sjansene for konkurs øker med en høy eierandel innad i styret. Variablene *d_dl_i_styre*, dummyvariabel for om daglig leder sitter i styret, og *d_to_dl*, dummyvariabel for om det er to i rollen som daglig leder i et selskap, blir valgt ut i 2018. De er signifikante på 0,1 %-nivå med z-verdier med absoluttverdi større enn 3,29. De har negativt fortegn, noe som indikerer at risikoen for konkurs reduseres om daglig leder sitter i styret, eller dersom et selskap har to daglige ledere. Dette sammenfaller med funnene i tidligere modeller hvor disse variablene har blitt inkludert. Videre er også variabelen *d_bistand_regnskapsfører*, dummyvariabel for om selskapet har fått bistand av regnskapsfører til å utarbeide regnskapet, valgt ut i 2018. Den har z-verdi med absoluttverdi større enn 1,96 og er dermed signifikant på 5 %-nivå. Fortegnet er positivt, noe som antyder at risikoen for konkurs øker dersom selskapet har

fått bistand av regnskapsfører. Dette tilsvarer funnet slik drøftet i modell 1B. Variabelen *andel_st_medl_ikke_aksj*, andelen styremedlemmer som ikke er aksjonærer, blir inkludert i modellen i 2020. Med en regresjonskoeffisient tilnærmet lik null, og en z-verdi på -0,02, er variabelen imidlertid ikke signifikant og effekten uansett begrenset.

Når det gjelder de finansielle variablene blir ikke *res_gjeld* valgt ut som en kritisk variabel i noen av årene, i likhet med i modell 3B. Resterende SEBRA-variabler er med sine z-verdier med absoluttverdi større enn 3,29 alle signifikante på 0,1 %-nivå i de årene de er inkludert, og har fortegn som forventet og drøftet i modell 3A.

5.3.4 Modellevaluering

Figur 8 viser *in-sample* og *out-of-sample* AUC for de tre ulike modellene 3A, 3B og 3C.



Figur 8: Sammenligning av AUC-verdier for modell 3A, 3B og 3C

Figuren viser en sammenligning av AUC-verdiene til modell 3A, 3B og 3C. Den lysegrå og mørkegrå linjen viser henholdsvis *in-sample* og *out-of-sample* AUC til modell 3A for de ulike årene i analyseperioden 2010 til 2020. Den lyseblå og mørkeblå linjen viser henholdsvis *in-sample* og *out-of-sample* AUC til modell 3B for tilsvarende periode. Den lyseoransje og mørkeoransje viser henholdsvis *in-sample* og *out-of-sample* til modell 3C for analyseårene 2018, 2019 og 2020.

Figuren viser at modell 3A, modellen med kun finansielle variabler, har laveste AUC i samt-

lige år i analyseperioden både når det gjelder *in-sample* og *out-of-sample*. Det er imidlertid mye mer interessant å observere at modell 3B, som inneholder både ikke-finansielle variabler og variabler inspirert av SEBRA-modellen øker *out-of-sample* AUC i store deler av perioden, og spesielt for årene 2011 og 2012. Dette vitner om at de ikke-finansielle variablene har fanget opp momenter som kan være relevant for konkursprediksjonsmodeller. En sammenligning av modellene 3C og 3B viser at det er en sterk forbedring i *in-sample* AUC, men at *out-of-sample* AUC er tilnærmet lik.

Når det kommer til nøyaktighetsratio presterer modell 3B like bra eller bedre enn modell 3A og 3C i alle regnskapsår. Videre for *Brier score* kommer modell 3B og 3C likt ut eller bedre enn modell 3A for alle regnskapsår, dette gjelder både *in-sample* og *out-of-sample*. Når det gjelder prestasjonsmålet AIC kommer modell 3B bedre ut enn modell 3A i alle årene, og videre kommer modell 3C best ut av alle tre modellene i de tre årene den er estimert. Resultatene viser at det å inkludere flere variabler gir økt forklaringskraft. Dette betyr at å inkludere ikke-finansielle variabler har en merverdi.

Avslutningsvis må det påpekes at det for flere av årene i de ulike modellene observeres at *out-of-sample* AUC får en høyere verdi enn *in-sample* AUC. Dette kan skyldes at datasettet stammer fra en stabil periode, uten de store svingningene. Lignende funn er også observert i andre studier som benytter samme datasett for regnskapsdata (Paraschiv et al., 2021).

6 Konklusjon

I likhet med Altman et al. (2010) og Ciampi (2015) finner også denne studien at inkludering av ikke-finansiell informasjon kan forbedre konkursprediksjon for SMB. For å besvare første del av problemstillingen:

I hvilken grad kan ikke-finansielle variabler forbedre konkursprediksjon for SMB?

så viser resultatene fra denne studien at modellene som inkluderer ikke-finansielle variabler, basert på målekriteriene AUC, AIC, nøyaktighetsratio og Brier score, oppnår en bedre prestasjon enn modellene som kun inneholder finansielle variabler, og at inkluderingen av ikke-finansielle variabler gir modellene økt forklaringskraft. Videre for del to av problemstillingen:

Hvilke ikke-finansielle variabler og kategorier har størst betydning?

så er det hvorvidt daglig leder sitter i styret (*dl_i_styre*) som blir valgt ut som den mest betydningsfulle ikke-finansielle variabelen flest ganger, etterfulgt av logaritmen av gjennomsnittsalderen til styremedlemmene (*log_styre_alder_snitt*) og logaritmen til alderen til styreleder (*log_alder_sl*). Alle disse variablene tilhører kategorien *egenskaper ved lederskapet*. Dette innebærer at disse variablene blir vurderert av LASSO til å ha størst betydning blant de ikke-finansielle variablene. Det må påpekes at variabelen *dl_i_styre* også fanger opp de tilfeller hvor daglig leder er styreleder, og fanger i så måte også opp CEO-dualitet. CEO-dualitet er en mye omtalt variabel i tidligere forskning, og funnene i denne studien peker mot at dersom daglig leder er styreleder eller medlem av styret vil risikoen for konkurs reduseres. Dette er nært beslektet med funnene til Ciampi (2015), som fant at CEO-dualitet reduserte sannsynligheten for økonomisk mislighold. Det at variablene *log_styre_alder_snitt* og *log_alder_sl* velges ut som to av de viktigste ikke-finansielle variablene er noe som samsvarer med funnene til Antulov-Fantulin et al. (2021), som fant at gjennomsnittsalderen til

styret var blant de ikke-finansielle variablene av størst betydning. Det bemerkes imidlertid at i motsetning til Antulov-Fantulin et al. (2021) som også fant at kjønns sammensetningen i styret var blant de viktigste variablene, finner ikke denne studien at kjønn er blant variablene av betydning. Det skal dog bemerkes at andelen kvinner i styret i gjennomsnitt kun er 16 % i dette datasettet, som vist i Tabell 5.

Dersom man ser på modellene som også inkluderer variabelsett II isolert sett, er det logaritmen av gjennomsnittsalderen til styremedlemmene (*log_styre_alder_snitt*) som blir valgt ut som den mest betydningsfulle ikke-finansielle variabelen flest ganger, etterfulgt av hvorvidt daglig leder bor i samme fylke som selskapet (*d_dl_fylke*) og logaritmen til alderen til styreleder (*log_alder_sl*). Alle disse variablene er også i kategorien *egenskaper ved lederskapet*. Imidlertid er disse modellene kun estimert for årene 2018, 2019 og 2020, men det kan likevel gi en indikasjon på hvilke variabler som er av størst betydning.

Dersom det kan utvikles en modell for konkursprediksjon bestående utelukkende av ikke-finansielle variabler som oppnår er god prediksjonsevne, vil dette kunne være svært nyttig eksempelvis i tilfeller hvor det skal vurderes lånesøknader for helt nystartede bedrifter uten tilgang på regnskapsdata. En sammenligning av prestasjonsmålene for modellene viser at modellene med kun ikke-finansielle variabler presterer dårligst. Dette indikerer at modeller som *kun* inkluderer ikke-finansielle variabler ikke fanger opp like mye informasjon som modeller som også inneholder variabler basert på regnskapstall. For regnskapsår 2011 og 2012 viser imidlertid prestasjonsmålet *out-of-sample* AUC for modell 1A en score på 0,69. Videre har modell 1B en *out-of-sample* AUC på 0,67 i 2020. Dette er tett opp under grensen hva antas som en akseptable verdier for en modell (Hosmer et al., 2013, s. 177), og taler for at dette er et område som bør forskes videre på.

Videre finner studien at inkluderingen av ikke-finansielle variabler fører til at variablene som gir uttrykk for ordinært resultat før av- og nedskrivninger som andel av gjeld (*res_gjeld*), samt logaritmen av totalkapitalen målt i faste kroner (*log_tk*) fra SEBRA-modellen, samt variabelen som beskriver egenkapital som andel av totalkapital (*ek_gjeld*) fra Z-score modellen ikke blir valgt ut av LASSO. Det er også observert at *out-of-sample* AUC får en

høyere verdi enn *in-sample* AUC for flere av modellene i flere perioder, men at dette kan forklares ved at datasettet stammer fra en stabil periode. Andre studier som benytter samme datasett for regnskapsdata har rapportert liknende funn (Paraschiv et al., 2021).

Kritikk av oppgaven

Variabelen *dl_i_styre* inkluderer både de tilfeller hvor daglig leder er styremedlem og styreleder, og fanger derfor opp noe av det samme som variabelen *samme_dl_sl*. Dette kan være en medvirkende årsak til at *samme_dl_sl* ikke blir valgt, og det kunne derfor ha vært interessant å skille disse slik at *dl_i_styre* kun fanger opp de tilfeller hvor daglig leder er styremedlem. I tillegg er det varierende kvalitet på data levert av Enin AS. En del av dette skyldes manglende data, i tillegg er det observert en større andel manglende verdier blant selskapene som har gått konkurs. Dette fører til at datasettet inneholder mindre informasjon om de selskapene som har gått konkurs, noe som ikke er heldig og som kan ha påvirket resultatene.

Implikasjoner og videre forskning

Det er flere momenter ved denne studien det ville vært interessant å undersøke nærmere. Modellene som inkluderer variabelsett II, altså modell 2C og 3C, har lovende tendenser og presterer godt. På bakgrunn av dette ville det vært svært interessant å inkludere disse variablene for flere regnskapsår når dette blir tilgjengelig. Dette følger for å se om den gode prestasjonen vedvarer over tid, og om det er en homogenitet i hvilke ikke-finansielle variabler som virker å være av betydning. Videre har modell 1B, modellen som kun benytter ikke-finansielle variabler, AUC-verdier som begynner å nærme seg akseptable i 2020. Et spørsmål man kan stille seg i den forbindelse er om det er mulig å utvikle egne modeller bestående utelukkende av ikke-finansiell informasjon som har akseptable prestasjoner. Dette ville i så måte vært av stor betydning eksempelvis i tilfeller hvor lånesøknader må vurderes med manglende eller begrenset tilgang på regnskapsinformasjon. Det ville også være interessant å benytte andre metoder innen maskinlæring for å se om funnene er stabile.

Referanser

- Agarwal, V. & Taffler, R. (2008). Comparing the performance of market-based and accounting-based bankruptcy prediction models. *Journal of Banking & Finance*, 32(8), 1541–1551. doi: 10.1016/j.jbankfin.2007.07.014.
- Aksjeloven. (1999). *Lov om aksjeselskaper (LOV-1997-06-13-44)*. Lovdata. Hentet 2022-04-25 fra https://lovdata.no/dokument/NL/lov/1997-06-13-44/**
- Altinn. (2020). *Konkurs i aksjeselskap*. Hentet 2022-03-24 fra <https://www.altinn.no/starte-og-drive/avvikling-sletting-og-konkurs/konkurs/konkurs-i-aksjeselskap/>
- Altman, E.I. (1968). Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, 23(4), 589–609. doi: 10.2307/2978933.
- Altman, E.I. (2000). Predicting financial distress of companies: revisiting the Z-Score and ZETA® models. *Handbook of Research Methods and Applications in Empirical Finance*, 428–456. doi: 10.4337/9780857936097.00027.
- Altman, E.I., Haldeman, R.G. & Narayanan, P. (1977). ZETA ANALYSIS. A new model to identify bankruptcy risk of corporations. *Journal of Banking & Finance*, 1(1), 29–54. doi: 10.1016/0378-4266(77)90017-6.
- Altman, E.I., Iwanicz-Drozowska, M., Laitinen, E.K. & Suvas, A. (2014). Distressed Firm and Bankruptcy Prediction in an International Context: A Review and Empirical Analysis of Altman’s Z-Score Model. *SSRN Electronic Journal*, doi: 10.2139/ssrn.2536340.
- Altman, E.I. & Sabato, G. (2007). Modelling Credit Risk for SMEs: Evidence from the U.S. Market. *Abacus*, 43(3), 332–357. doi: 10.1111/j.1467-6281.2007.00234.x.
- Altman, E.I., Sabato, G. & Wilson, N. (2010). The value of non-financial information in small and medium-sized enterprise risk management. *The Journal of Credit Risk*, 6(2), 33. doi: 10.21314/JCR.2010.110,.
- Antulov-Fantulin, N., Lagravinese, R. & Resce, G. (2021). Predicting bankruptcy of local government: A machine learning approach. *Journal of Economic Behavior &*

-
- Organization*, 183, 681–699. doi: 10.1016/j.jebo.2021.01.014.
- Audrino, F., Kostrov, A. & Ortega, J.-P. (2019). Predicting U.S. Bank Failures with MIDAS Logit Models. *Journal of Financial and Quantitative Analysis*, 54(6), 2575–2603. doi: 10.1017/S0022109018001308.
- Beaver, W.H. (1966). Financial Ratios As Predictors of Failure. *Journal of Accounting Research*, 4, 71–111. doi: 10.2307/2490171.
- Beaver, W.H., McNichols, M.F. & Rhie, J.W. (2005). Have Financial Statements Become Less Informative? Evidence from the Ability of Financial Ratios to Predict Bankruptcy. *Review of Accounting Studies*, 10(1), 93–122. doi: 10.1007/s11142-004-6341-9.
- Berg, D. (2007). Bankruptcy prediction by generalized additive models. *Applied Stochastic Models in Business and Industry*, 23(2), 129–143. doi: 10.1002/asmb.658.
- Bernhardsen, E. & Larsen, K. (2007). Modelling av kreditrisiko i foretakssektoren – videreutvikling av SEBRA-modellen. *Penger og Kreditt*, 2(7), 60–66.
- Brédart, X., Séverin, E. & Veganzones, D. (2021). Human resources and corporate failure prediction modeling: Evidence from Belgium. *Journal of Forecasting*, 40(7), 1325–1341. doi: 10.1002/for.2770.
- Cenciarelli, V.G., Greco, G. & Allegrini, M. (2018). External audit and bankruptcy prediction. *Journal of Management and Governance*, 22(4), 863–890. doi: 10.1007/s10997-018-9406-z.
- Chava, S. & Jarrow, R.A. (2004). Bankruptcy Prediction With Industry Effects. *Review of Finance*, 8(4), 537–569. doi: 10.1093/rof/8.4.537.
- Cho, E., Okafor, C., Ujah, N. & Zhang, L. (2021). Executives’ gender-diversity, education, and firm’s bankruptcy risk: Evidence from China. *Journal of Behavioral and Experimental Finance*, 30, 100500. doi: 10.1016/j.jbef.2021.100500,.
- Ciampi, F. (2015). Corporate governance characteristics and default prediction modeling for small enterprises. An empirical analysis of Italian firms. *Journal of Business Research*, 68(5), 1012–1025. doi: 10.1016/j.jbusres.2014.10.003.
- Ciampi, F., Giannozzi, A., Marzi, G. & Altman, E.I. (2021). Rethinking SME default
-

-
- prediction: a systematic literature review and future perspectives. *Scientometrics*, 126(3), 2141–2188. doi: 10.1007/s11192-020-03856-0.
- Dalton, D.R., Daily, C.M., Johnson, J.L. & Ellstrand, A.E. (1999). Number of Directors and Financial Performance: A Meta-Analysis. *The Academy of Management Journal*, 42(6), 674–686. doi: 10.2307/256988.
- Departementene. (2019). *Småbedriftslivet - Strategi for små- og mellomstore bedrifter*. Nærings- og fiskeridepartementet. Hentet fra <https://www.regjeringen.no/globalassets/departementene/nfd/dokumenter/vedlegg/smabedriftslivet-uu.pdf>
- Dowell, G.W.S., Shackell, M.B. & Stuart, N.V. (2011). Boards, CEOs, and surviving a financial crisis: Evidence from the internet shakeout. *Strategic Management Journal*, 32(10), 1025–1045. doi: 10.1002/smj.923.
- Duru, A., Iyengar, R.J. & Zampelli, E.M. (2016). The dynamic relationship between CEO duality and firm performance: The moderating role of board independence. *Journal of Business Research*, 69(10), 4269–4277. doi: 10.1016/j.jbusres.2016.04.001.
- Edmister, R.O. (1972). An Empirical Test of Financial Ratio Analysis for Small Business Failure Prediction. *The Journal of Financial and Quantitative Analysis*, 7(2), 1477–1493. doi: 10.2307/2329929.
- Eisenbeis, R.A. (1977). Pitfalls in the Application of Discriminant Analysis in Business, Finance, and Economics. *The Journal of Finance*, 32(3), 875–900. doi: 10.2307/2326320.
- Eklund, T., Larsen, K. & Bernhardsen, E. (2001). Modell for analyse av kredittrisiko i foretakssektoren. *Penger og Kreditt*, 2(1).
- Elloumi, F. & Gueyié, J. (2001). Financial distress and corporate governance: an empirical analysis. *Corporate Governance: The international journal of business in society*, 1(1), 15–23. doi: 10.1108/14720700110389548.
- Escribano-Navas, M. & Gemar, G. (2021). Gender and Bankruptcy: A Hotel Survival Econometric Analysis. *Sustainability*, 13(12), 6782. doi: 10.3390/su13126782.
- European Commission. (2003). *Internal Market, Industry, Entrepreneurship and SMEs*.
-

-
- Hentet 2022-04-07 fra https://ec.europa.eu/growth/smes/sme-definition_en
- Evans, D.S. (1987). The Relationship Between Firm Growth, Size, and Age: Estimates for 100 Manufacturing Industries. *The Journal of Industrial Economics*, 35(4), 567–581. doi: 10.2307/2098588.
- Finansdepartementet. (2022). *Kompensasjonsordning for næringslivet*. Hentet 2022-04-24 fra <https://www.regjeringen.no/no/tema/okonomi-og-budsjett/finansmarkedene/kompensasjonsordningen/id2698845/>
- Finanstilsynet. (2021). Finansielt utsyn juni 2021. Hentet fra <https://www.finanstilsynet.no/contentassets/de019705b5094a37ace5105e8b74b76d/finansielt-utsyn-juni-2021.pdf>
- Finkelstein, S. & D’Aveni, R.A. (1994). CEO Duality as a Double-Edged Sword: How Boards of Directors Balance Entrenchment Avoidance and Unity of Command. *The Academy of Management Journal*, 37(5), 1079–1108. doi: 10.2307/256667.
- Fisher, R.A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7(2), 179–188. doi: 10.1111/j.1469-1809.1936.tb02137.x.
- Forskrift om terskelverdier. (2018). *Forskrift om terskelverdier for beslutning om å unnlate revisjon etter aksjeloven § 7-6*. Lovdata. Hentet 2022-04-21 fra <https://lovdata.no/dokument/SF/forskrift/2018-01-03-7>
- García, C.J. & Herrero, B. (2021, november). Female directors, capital structure, and financial distress. *Journal of Business Research*, 136, 592–601. doi: 10.1016/j.jbusres.2021.07.061.
- Gissel, J.L., Giacomino, D. & Akers, M.D. (2007). A Review of Bankruptcy Prediction Studies: 1930-Present. *Journal of Financial Education*, 33, 1–42.
- Grunert, J., Norden, L. & Weber, M. (2005). The role of non-financial factors in internal credit ratings. *Journal of Banking & Finance*, 29(2), 509–531. doi: 10.1016/j.jbankfin.2004.05.017.
- Guest, P.M. (2009). The impact of board size on firm performance: evidence from the UK. *The European Journal of Finance*, 15(4), 385–404. doi: 10.1080/13518470802466121.
- Hillegeist, S.A., Keating, E.K., Cram, D.P. & Lundstedt, K.G. (2004). Assessing
-

-
- the Probability of Bankruptcy. *Review of accounting studies*, 9(1), 5–34. doi: 10.1023/B:RAST.0000013627.90884.b7.
- Hjelseth, I.N., Arbatli, E., Solheim, H. & Vatne, B.H. (2020). Anslag på bankenes tap på utlån til ikke-finansielle foretak. *Norges Bank, Finansiell stabilitet*, 10(20).
- Hoff, K.G. & Helbæk, M. (2021). *Økonomistyring 2 - driftsregnskap og budsjettering* (7. utg.). Universitetsforlaget.
- Hosmer, D.W., Lemeshow, S. & Sturdivant, R.X. (2013). *Applied Logistic Regression*. John Wiley & Sons.
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An Introduction to Statistical Learning* (vol. 103). Springer New York. doi: 10.1007/978-1-4614-7138-7.
- Jensen, M.C. (1993). The Modern Industrial Revolution, Exit, and the Failure of Internal Control Systems. *The Journal of Finance*, 48(3), 831–880. doi: 10.1111/j.1540-6261.1993.tb04022.x.
- Jones, S., Johnstone, D. & Wilson, R. (2015). An empirical evaluation of the performance of binary classifiers in the prediction of credit ratings changes. *Journal of Banking & Finance*, 56, 72–85. doi: 10.1016/j.jbankfin.2015.02.006.
- Jones, S., Johnstone, D. & Wilson, R. (2017). Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks. *Journal of Business Finance & Accounting*, 44(1-2), 3–34. doi: 10.1111/jbfa.12218.
- Konkursloven. (1986). *Lov om gjeldsforhandling og konkurs. LOV-1984-06-08-58*. Hentet 2022-05-07 fra <https://lovdata.no/dokument/NL/lov/1984-06-08-58>
- Krause, R., Semadeni, M. & Cannella, A.A. (2014). CEO Duality: A Review and Research Agenda. *Journal of Management*, 40(1), 256–286. doi: 10.1177/0149206313503013.
- Krause, R., Semadeni, M. & Withers, M.C. (2016). That special someone: When the board views its chair as a resource. *Strategic Management Journal*, 37(9), 1990–2002. doi: 10.1002/smj.2444.
- Li, H. & Sun, J. (2012). Forecasting business failure: The use of nearest-neighbour support vectors and correcting imbalanced samples – Evidence from the Chinese hotel industry. *Tourism Management*, 33(3), 622–634. doi: 10.1016/j.tourman.2011.07.004.

-
- Liang, D., Lu, C.-C., Tsai, C.-F. & Shih, G.-A. (2016). Financial ratios and corporate governance indicators in bankruptcy prediction: A comprehensive study. *European Journal of Operational Research*, 252(2), 561–572. doi: 10.1016/j.ejor.2016.01.012.
- Liang, D., Tsai, C.-F., Lu, H.-Y.R. & Chang, L.-S. (2020). Combining corporate governance indicators with stacking ensembles for financial distress prediction. *Journal of Business Research*, 120, 137–146. doi: 10.1016/j.jbusres.2020.07.052.
- Lilienfeld-Toal, U.v. & Ruenzi, S. (2014). CEO Ownership, Stock Market Performance, and Managerial Discretion. *The Journal of Finance*, 69(3), 1013–1050. doi: 10.1111/jofi.12139.
- López, V., Fernández, A., García, S., Palade, V. & Herrera, F. (2013). An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information Sciences*, 250, 113–141. doi: 10.1016/j.ins.2013.07.007.
- Mangena, M. & Chamisa, E. (2008). Corporate governance and incidences of listing suspension by the JSE Securities Exchange of South Africa: An empirical analysis. *The International Journal of Accounting*, 43(1), 28–44. doi: 10.1016/j.intacc.2008.01.002.
- Mansi, S.A., Maxwell, W.F. & Zhang, A.J. (2012). Bankruptcy Prediction Models and the Cost of Debt. *The Journal of Fixed Income*, 21(4), 25–42. doi: 10.3905/jfi.2012.21.4.025.
- Manzaneque, M., Priego, A.M. & Merino, E. (2016). Corporate governance effect on financial distress likelihood: Evidence from Spain. *Spanish Accounting Review*, 19(1), 111–121. doi: 10.1016/j.rcsar.2015.04.001.
- Nokeri, T.C. (2021). *Data Science Revealed: With Feature Engineering, Data Visualization, Pipeline Development, and Hyperparameter Tuning*. Berkeley, CA: Apress. doi: 10.1007/978-1-4842-6870-4.
- Nærings - og handelsdepartementet. (2012). *Små bedrifter – store verdier. Regjeringens strategi for små og store bedrifter* (2. utg.). Hentet fra <https://www.regjeringen.no/no/no/dokumenter/sma-bedrifter---store-verdier/id676379/>
- Næss, A.B., Wahlstrøm, R.R., Helland, F.F. & Kjærland, F. (2017). Konkursprediksjon

-
- for norske selskaper - En sammenligning av regresjonsmodeller og maskinlæringsteknikker. Bred og spiss! NTNU Handelshøyskolen 50 år: En vitenskapelig jubileumsantologi. *Fagbokforlaget*.
- Ohlson, J.A. (1980). Financial Ratios and the Probabilistic Prediction of Bankruptcy. *Journal of Accounting Research*, 18(1), 109–131. doi: 10.2307/2490395.
- Paraschiv, F., Schmid, M. & Wahlstrøm, R.R. (2021). Bankruptcy Prediction of Privately Held SMEs Using Feature Selection Methods. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3911490.
- Pearce, J.A. & Zahra, S.A. (1992). Board composition from a strategic contingency perspective. *Journal of management studies*, 29(4), 411-438. doi: 10.1111/j.1467-6486.1992.tb00672.x.
- Pelja, I. & Wahlstrøm, R.R. (2021). Hvordan påvirker bedriftens størrelse predikering av konkurs? *Magma - Tidsskrift for Økonomi Og Ledelse*, 7, 82–91.
- Platt, H. & Platt, M. (2012). Corporate board attributes and bankruptcy. *Journal of Business Research*, 65(8), 1139–1143. doi: 10.1016/j.jbusres.2011.08.003.
- Ravi Kumar, P. & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques – A review. *European Journal of Operational Research*, 180(1), 1–28. doi: 10.1016/j.ejor.2006.08.043.
- Regnskapsloven. (1999). Lov om årsregnskap m.v. (LOV-1998-07-17-56). Lovdata. Hentet 2022-04-23 fra <https://lovdata.no/dokument/NL/lov/1998-07-17-56>
- Schalck, C. & Yankol-Schalck, M. (2021). Predicting French SME failures: new evidence from machine learning techniques. *Applied Economics*, 53(51), 5948–5963. doi: 10.1080/00036846.2021.1934389.
- Shumway, T. (2001). Forecasting Bankruptcy More Accurately: A Simple Hazard Model. *The Journal of Business*, 74(1), 101–124. doi: 10.1086/209665.
- Skatteetaten. (2021). *Skatteinnkreving under koronakrisen*. Hentet 2022-04-24 fra <https://www.skatteetaten.no/presse/nyhetsrommet/skatteinnkreving-under-koronakrisen/>
- Smith, R.F. & Winakor, A.H. (1930). A test analysis of unsuccessful industrial companies.
-

-
- Bureau of Business Research, Bulletin 30. Urbana, IL: University of Illinois Press.*
- Smogeli, P.O. (1987). Dokumentasjonsnotat SEBRA. *Interne notater (statistisk sentralbyrå)*, 87, 53.
- Statistisk sentralbyrå. (2009a). *Internasjonal finanskriser - fra uro til krise*. Hentet 2022-04-25 fra <https://www.ssb.no/forskning/makrookonomi/makrookonomiske-analyser/internasjonal-finanskriser-fra-uro-til-krise>
- Statistisk sentralbyrå. (2009b). *Standard for næringsgruppering (SN)*. Hentet 2022-04-07 fra <https://www.ssb.no/klass/klassifikasjoner/6>
- Statistisk sentralbyrå. (2020a). *Koronakrisens innvirkning på antall konkurser*. Hentet 2022-04-24 fra <https://www.ssb.no/virksomheter-foretak-og-regnskap/artikler-og-publikasjoner/koronakrisens-innvirkning-pa-antall-konkurser>
- Statistisk sentralbyrå. (2020b). *Ukentlig statistikk over konkurser*. Hentet 2022-04-24 fra <https://www.ssb.no/virksomheter-foretak-og-regnskap/artikler-og-publikasjoner/ukentlig-statistikk-over-konkurser--416315>
- Sæther, P. & Larsen, K. (1999). Kredittrisiko knyttet til foretakssektoren. *Penger og Kreditt*, 99(4), 546–554.
- Tang, X., Li, S., Tan, M. & Shi, W. (2020). Incorporating textual and management factors into financial distress prediction: A comparative study of machine learning methods. *Journal of forecasting*, 39(5), 769–787. doi: 10.1002/for.2661.
- Tian, S., Yu, Y. & Guo, H. (2015). Variable selection and corporate bankruptcy forecasts. *Journal of Banking & Finance*, 52, 89–100. doi: 10.1016/j.jbankfin.2014.12.003.
- Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. doi: 10.1111/j.2517-6161.1996.tb02080.x.
- Wahlstrøm, R.R. (2022). Financial statements of companies in Norway. , doi: 10.48550/arXiv.2203.12842.
- Withers, M.C. & Fitza, M.A. (2017). Do board chairs matter? The influence of board chairs on firm performance. *Strategic Management Journal*, 38(6), 1343–1355. doi:
-

10.1002/smj.2587.

Zmijewski, M.E. (1984). Methodological Issues Related to the Estimation of Financial Distress Prediction Models. *Journal of Accounting Research*, 22, 59–82. doi: 10.2307/2490859.

Appendiks

Appendiks A: Variabler som inngår i SEBRA-basis og SEBRA-utvidet

Appendiks B: Korrelasjonsmatrise av ikke-finansielle variabler

Appendiks C: Manglende observasjoner for ikke-finansielle variabler

Appendiks D: LASSO-plott med tilhørende AUC-verdier

Appendiks E: Oppsummering av variabler utvalgt av LASSO

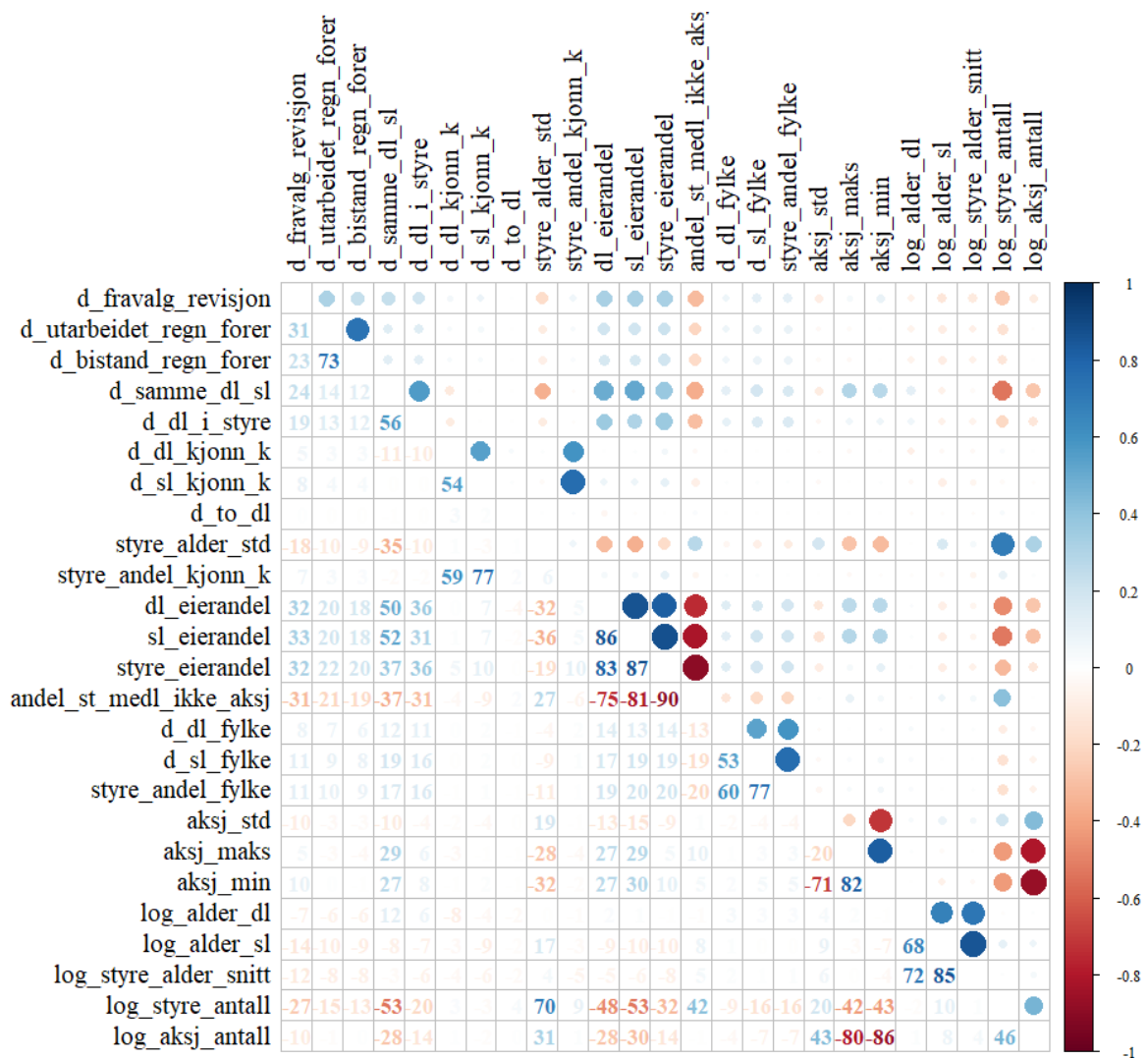
Appendiks A

Tabellen viser variabler som inngår i SEBRA-basis og SEBRA-utvidet og er hentet fra Bernhardsen & Larsen (2007). Variablene som inngår i SEBRA-basis er markert med grå bakgrunn, og SEBRA-utvidet er samtlige variabler i tabellen.

Variabeldefinisjon	Variabeltype	Variere over
Ordinært resultat før av- og nedskrivninger i prosent av total gjeld	Nøkkeltall	Foretak/år
	Gjennomsnitt	Bransje/år
	Standardavvik	Bransje/år
	Korrelasjon med Norgesporteføljen	Bransje
Egenkapital i prosent av total kapital	Nøkkeltall	Foretak/år
	Gjennomsnitt	Bransje/år
Bokført egenkapital mindre enn innskutt egenkapital	Indikator	Foretak/år
Likvider minus kortsiktig gjeld i prosent av omsetning	Nøkkeltall	Foretak/år
Alder (år) = 1,2,3....8	Indikatorer	Foretak/år
Sum eiendeler i faste kroner	Nøkkeltall	Foretak/år
Leverandørgjeld i prosent av total kapital	Nøkkeltall	Foretak/år
Skyldige offentlige avgifter i prosent av total kapital	Nøkkeltall	Foretak/år

Appendiks B

Figuren viser korrelasjonsmatrise for de ikke-finansielle variablene denne studien benytter. Korrelasjonsverdiene er i nedre halvdel av matrisen oppgitt i prosent. Visuelt er høy korrelasjon i matrisen markert med sterke farger.



Appendiks C

Tabellen viser antall manglende verdier for de ikke-finansielle variablene denne studien benytter fordelt på de ulike regnskapsårene. Røde områder indikerer at antallet manglende verdier er høyt, og grønne områder viser at andelen manglende verdier er lavt.

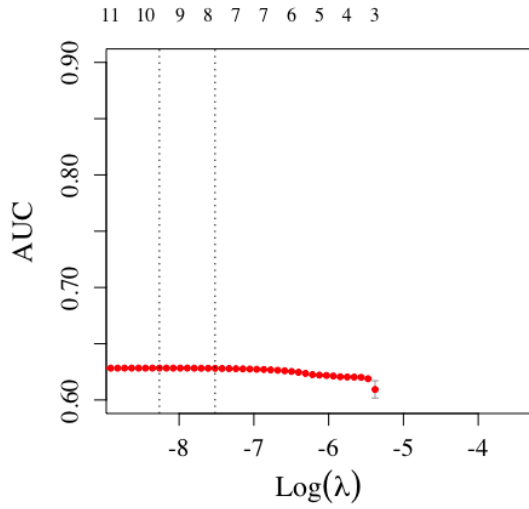
Variabler	2006	2007	2008	2009	2010	2011	2012	2013
alder_dl	49 779	47 921	44 290	40 029	35 862	32 141	28 595	24 757
alder_sl	53 029	50 249	45 823	40 913	36 459	32 015	28 148	23 806
d_samme_dl_sl	52 279	47 535	41 783	36 532	32 032	27 997	24 748	21 518
d_dl_i_styre	50 181	45 469	39 765	34 591	30 165	26 286	23 179	20 154
d_dl_kjonn_k	39 540	35 296	30 458	26 191	22 637	19 583	17 220	14 920
d_sl_kjonn_k	36 377	31 183	26 031	21 861	18 522	15 556	13 209	10 910
d_to_dl	39 540	35 296	30 458	26 191	22 637	19 583	17 220	14 920
styre_alder_snitt	43 361	38 420	32 988	28 164	23 992	20 181	16 993	13 484
styre_alder_std	43 361	38 420	32 988	28 164	23 992	20 181	16 993	13 484
styre_antall	33 301	28 132	23 033	18 963	15 666	12 855	10 613	8 441
styre_andel_kjonn_k	33 301	28 132	23 033	18 963	15 666	12 855	10 613	8 441
d_dl_fylke	56 790	56 960	54 572	50 764	46 956	43 614	40 434	36 610
d_sl_fylke	64 958	64 516	61 191	56 534	52 229	47 964	44 250	39 564
styre_andel_fylke	68 376	70 184	68 612	65 088	61 730	58 454	55 446	51 401
d_fravalg_revisjon	97 536	103 985	106 810	107 369	108 822	99 495	19 285	10 658
d_utarbeidet_regn_forer	97 536	103 985	106 810	107 369	108 823	99 497	19 430	10 273
d_bistand_regn_forer	97 536	103 985	106 810	107 369	108 824	99 498	19 601	9 989
aksj_antall	97 536	103 985	106 812	107 371	108 846	111 607	116 211	120 621
aksj_std	97 536	103 985	106 812	107 371	108 846	111 607	116 211	120 621
aksj_maks	97 536	103 985	106 812	107 371	108 846	111 607	116 211	120 621
aksj_min	97 536	103 985	106 812	107 371	108 846	111 607	116 211	120 621
dl_eierandel	97 536	103 985	106 812	107 371	108 846	111 607	116 211	120 621
sl_eierandel	97 536	103 985	106 812	107 371	108 846	111 607	116 211	120 621
styre_eierandel	97 536	103 985	106 812	107 371	108 846	111 607	116 211	120 621
andel_st_medl_ikke_aksj	97 536	103 985	106 812	107 371	108 846	111 607	116 211	120 621

Variabler	2014	2015	2016	2017	2018	2019	2020
alder_dl	20 968	17 641	14 754	12 112	9 492	8 745	8 946
alder_sl	19 568	15 978	12 185	7 787	3 574	1 578	715
d_samme_dl_sl	18 468	15 583	13 276	11 444	10 307	9 427	8 904
d_dl_i_styre	17 342	14 700	12 760	11 242	10 572	10 394	10 420
d_dl_kjonn_k	12 783	10 735	9 335	8 464	8 310	8 438	8 704
d_sl_kjonn_k	8 907	7 088	5 395	3 768	2 584	1 495	706
d_to_dl	12 783	10 735	9 335	8 464	8 310	8 438	8 704
styre_alder_snitt	10 378	7 944	5 813	3 366	1 519	585	149
styre_alder_std	10 378	7 944	5 813	3 366	1 519	585	149
styre_antall	6 600	4 951	3 528	2 113	1 219	551	147
styre_andel_kjonn_k	6 600	4 951	3 528	2 113	1 219	551	147
d_dl_fylke	32 051	27 290	22 744	18 688	15 661	13 515	12 376
d_sl_fylke	34 145	28 610	22 693	16 599	12 184	8 941	6 725
styre_andel_fylke	46 261	40 882	34 567	27 520	20 412	14 587	10 344
d_fravalg_revisjon	12 672	12 850	12 580	6 543	5 883	4 923	4 049
d_utarbeidet_regn_forer	12 433	10 256	9 689	6 302	5 972	5 055	4 515
d_bistand_regn_forer	12 446	10 387	9 783	6 405	6 058	5 213	4 658
aksj_antall	2 692	2 410	1 193	999	1 047	1 183	1 187
aksj_std	2 692	2 410	1 193	999	1 047	1 183	1 187
aksj_maks	2 692	2 410	1 193	999	1 047	1 183	1 187
aksj_min	2 692	2 410	1 193	999	1 047	1 183	1 187
dl_eierandel	15 084	12 832	10 380	9 350	9 259	9 526	9 786
sl_eierandel	11 447	9 385	6 540	4 735	3 598	2 655	1 864
styre_eierandel	9 179	7 281	4 691	3 096	2 245	1 720	1 312
andel_st_medl_ikke_aksj	9 179	7 281	4 691	3 096	2 245	1 720	1 312

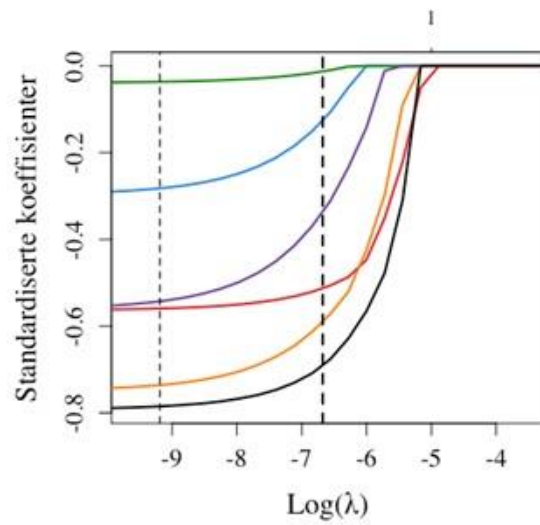
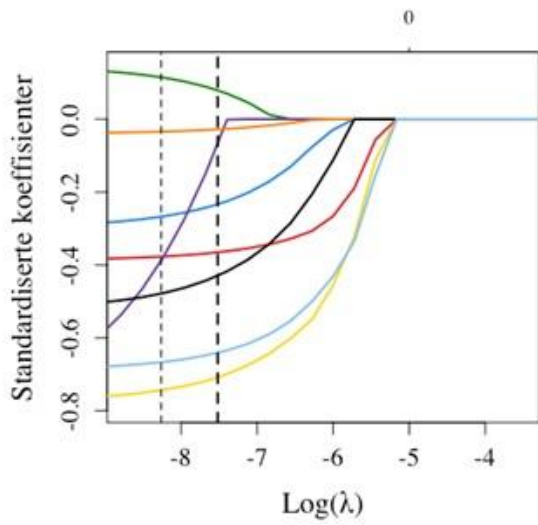
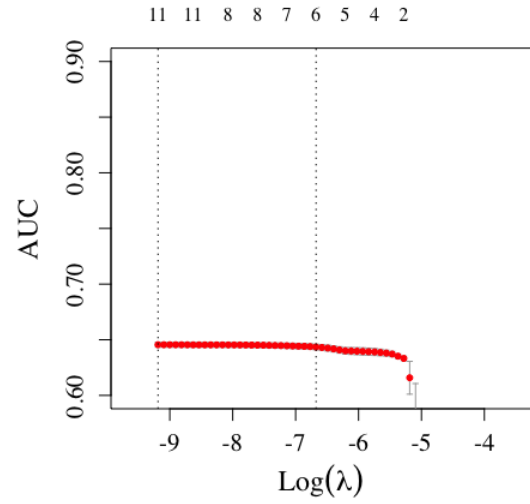
Appendiks D

Det vil i det følgende bli presentert LASSO-plot for hvert regnskapsår for de 8 ulike modellene. LASSO-plottene viser hvilke variabler som blir valgt ut til å inngå i de ulike modellene. Den venstre stiplede vertikale linjen angir beste gjennomsnittlige verdien for *in-sample* AUC blant alle verdier av λ som er inkludert i grid-søket identifiseres. Verdien av λ som så velges er den verdien som gir en *in-sample* AUC som ligger ett standardavvik fra den beste. Hvordan AUC-verdiene endres med antall variabler inkludert er vist i tilhørende graf.

Modell 1A – testår 2010



Modell 1A – testår 2011



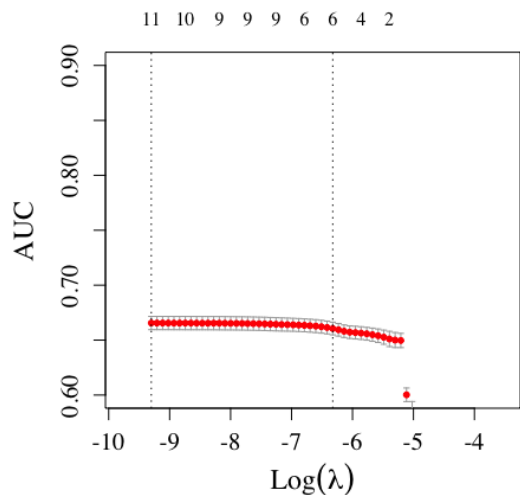
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- d_dl_kjonn_k
- d_to_dl
- styre_alder_std
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt

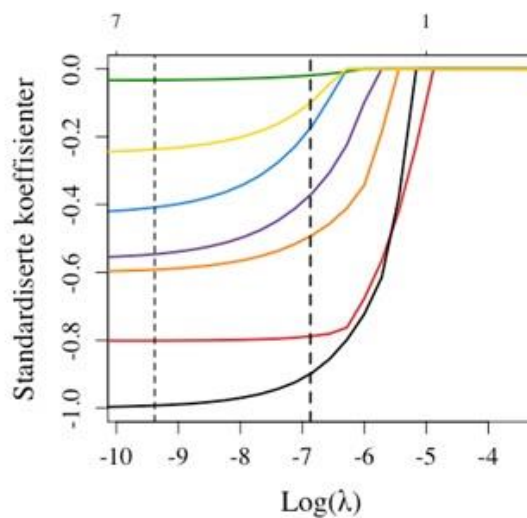
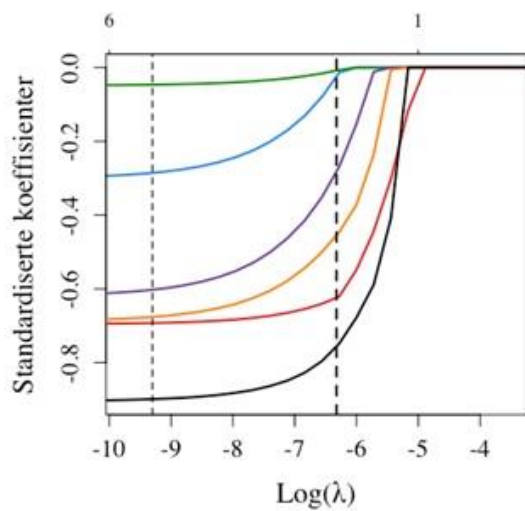
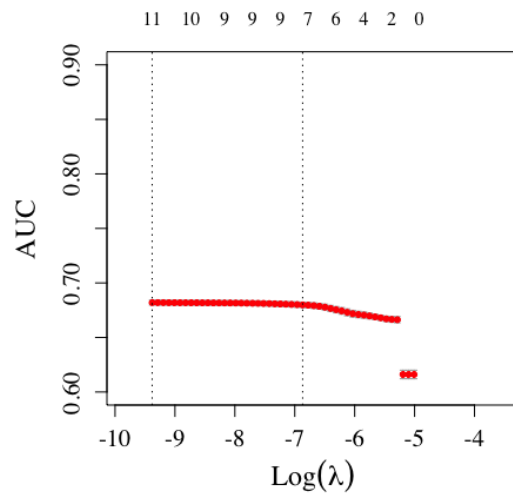
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- styre_alder_std
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt

Modell 1A – testår 2012



Modell 1A – testår 2013



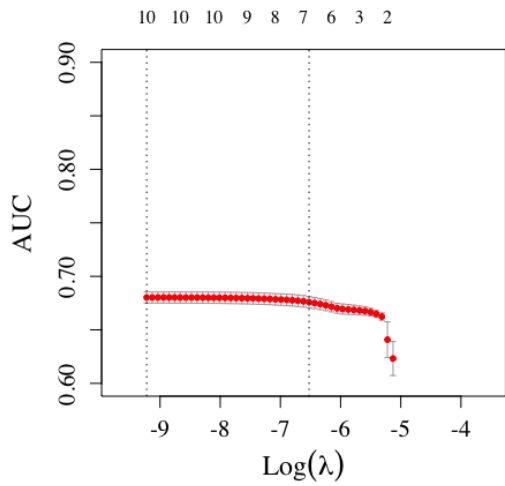
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- styre_alder_std
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt

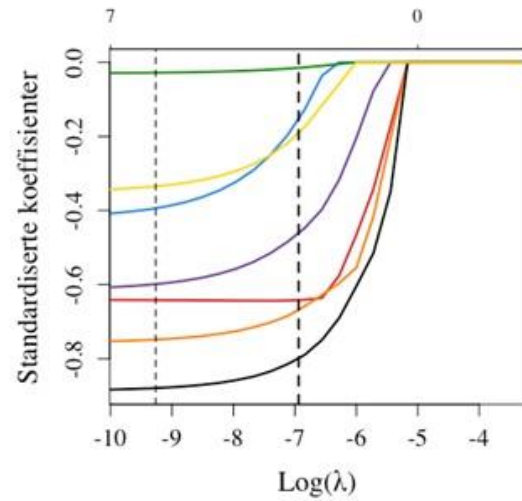
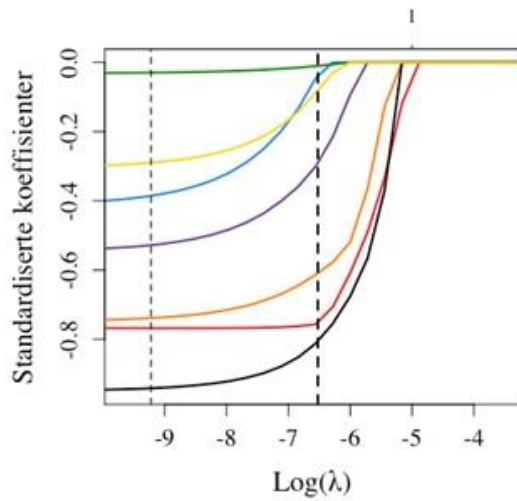
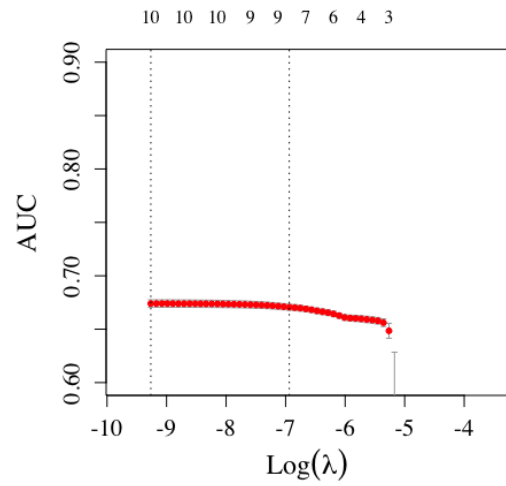
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- styre_alder_std
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

Modell 1A – testår 2014



Modell 1A – testår 2015



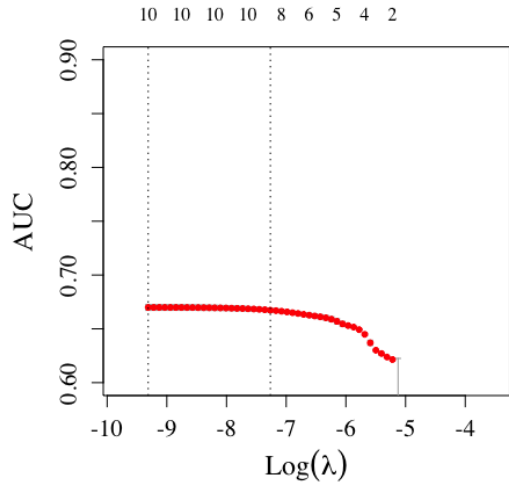
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- styre_alder_std
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

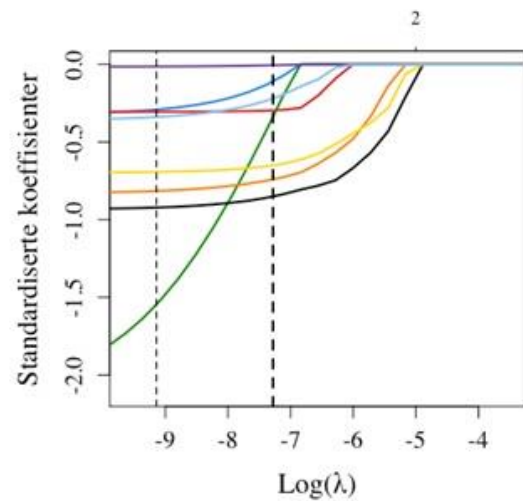
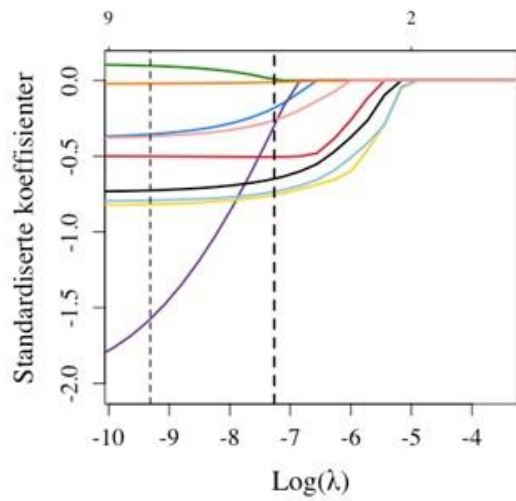
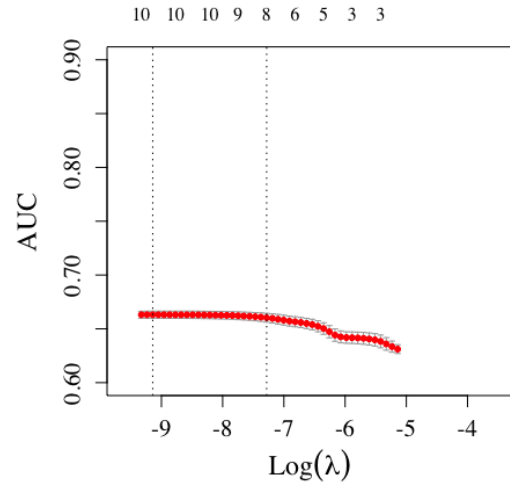
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- styre_alder_std
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

Modell 1A – testår 2016



Modell 1A – testår 2017



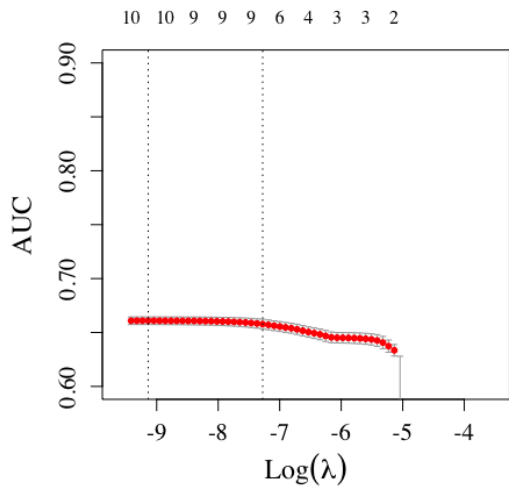
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- d_dl_kjonn_k
- d_to_dl
- styre_alder_std
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

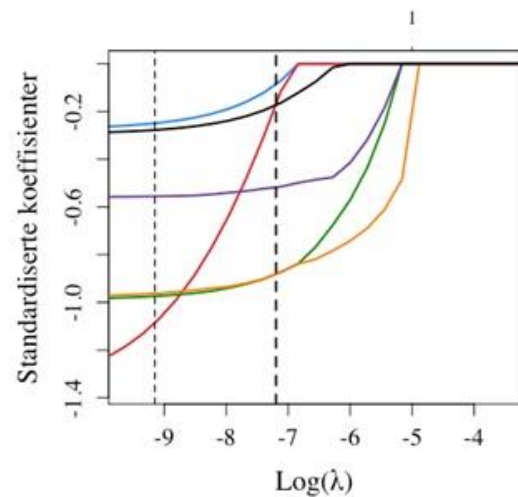
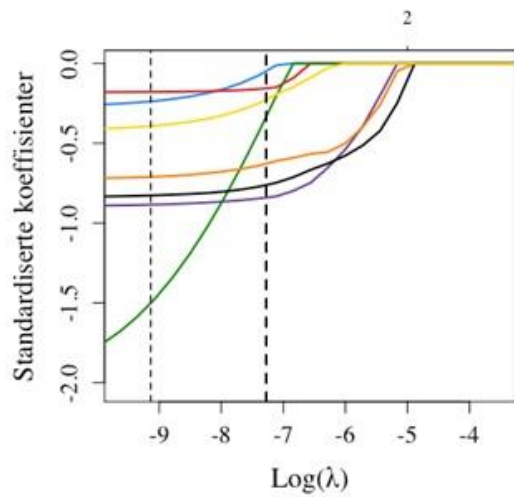
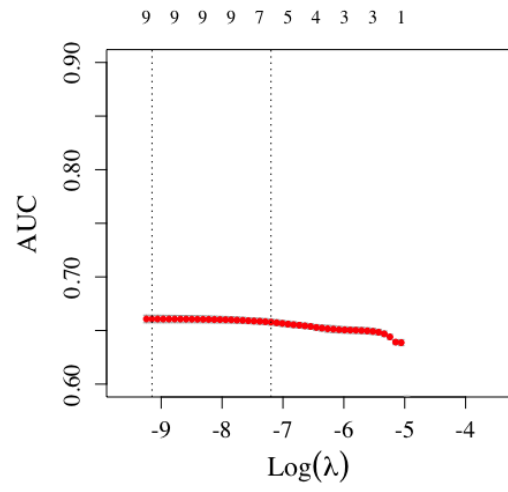
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- d_to_dl
- styre_alder_std
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

Modell 1A – testår 2018



Modell 1A – testår 2019



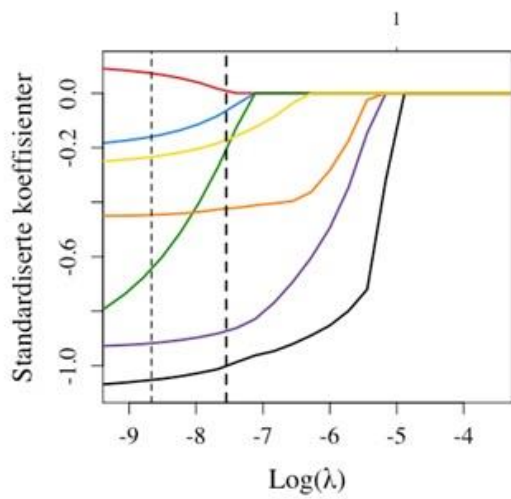
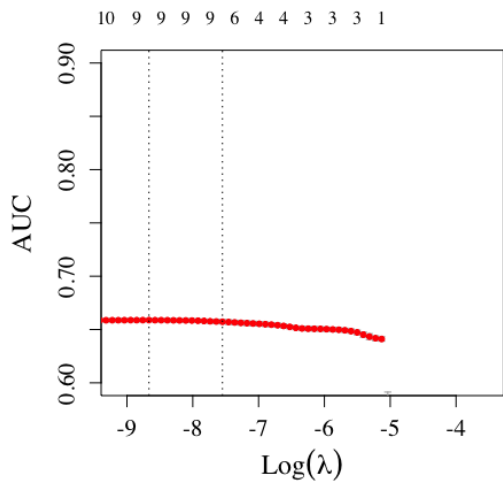
Variabler valgt av LASSO

- d_samme_dl_sl
- d_dl_i_styre
- d_to_dl
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

Variabler valgt av LASSO

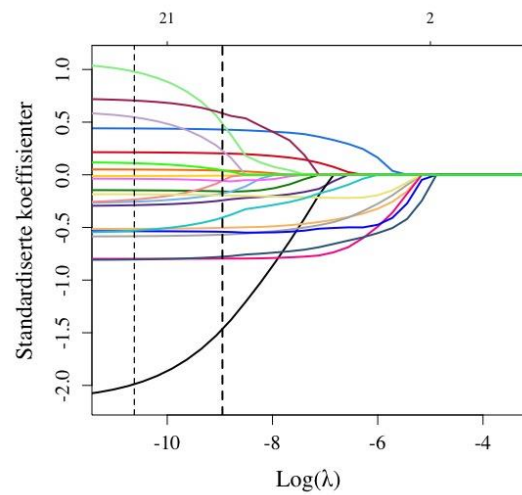
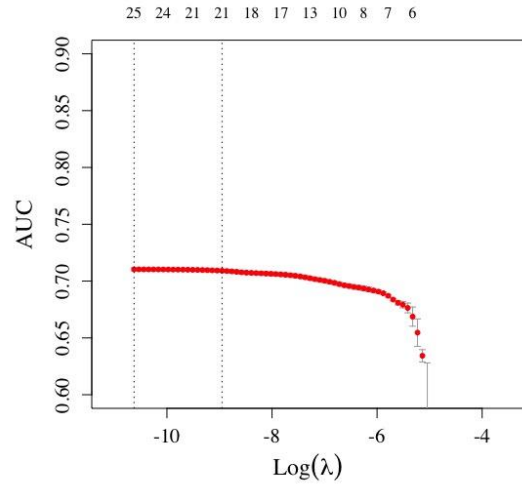
- d_dl_i_styre
- d_to_dl
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

Modell 1A – testår 2020



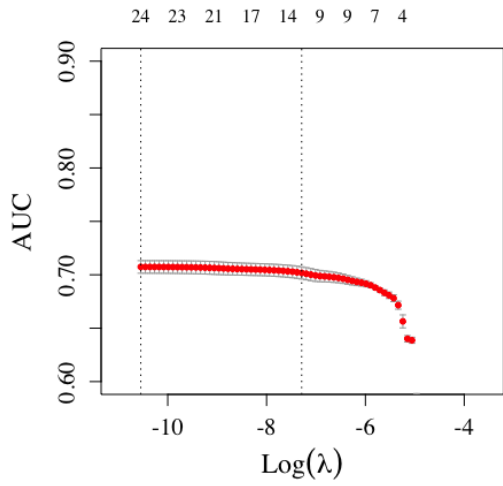
- Variabler valgt av LASSO**
- d_dl_i_styre
 - d_dl_kjonn_k
 - d_to_dl
 - log_alder_dl
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall

Modell 1B – testår 2018

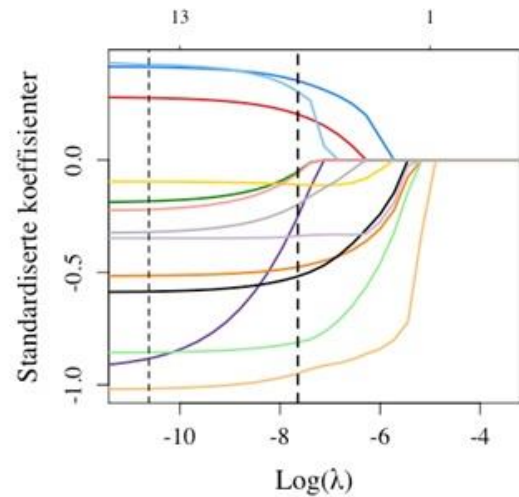
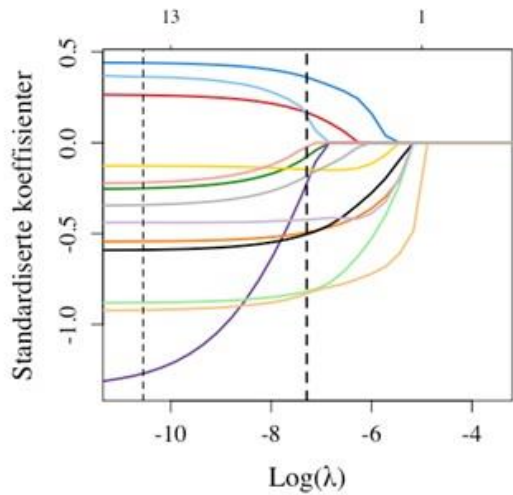
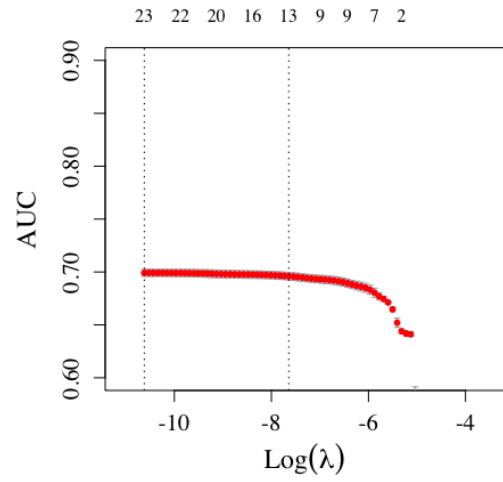


- Variabler valgt av LASSO**
- d_fravalg_revisjon
 - d_bistand_regn_forer
 - d_samme_dl_sl
 - d_dl_in_styre
 - d_dl_kjonn_k
 - d_to_dl
 - styre_alder_std
 - dl_eierandel
 - sl_eierandel
 - styre_eierandel
 - andel_st_medl_ikke_aksj
 - d_dl_fylke
 - d_sl_fylke
 - styre_andel_fylke
 - aksj_std_slutt
 - aksj_min_slutt
 - log_alder_dl
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall
 - log_aksj_antall_slutt

Modell 1B – testår 2019



Modell 1B – testår 2020



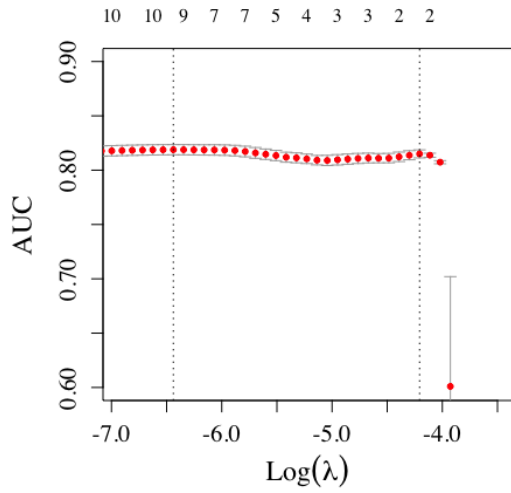
Variabler valgt av LASSO

- d_fravalg_revisjon
- d_bistand_regn_forer
- d_dl_i_styre
- d_to_dl
- d_dl_fylke
- d_sl_fylke
- styre_andel_fylke
- aksj_std
- aksj_min
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

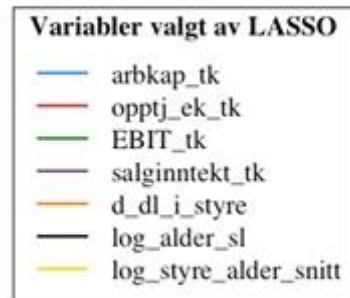
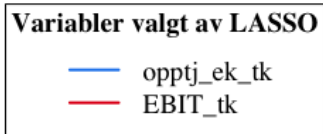
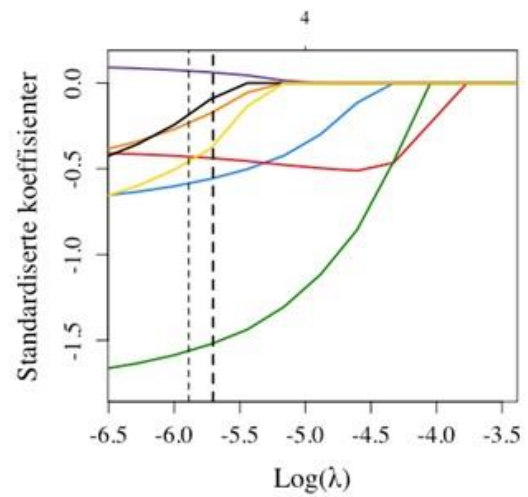
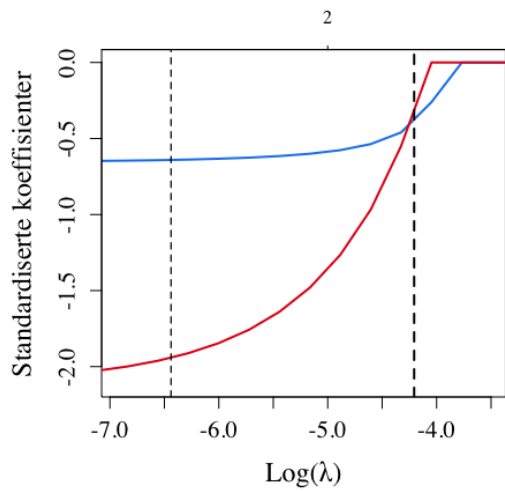
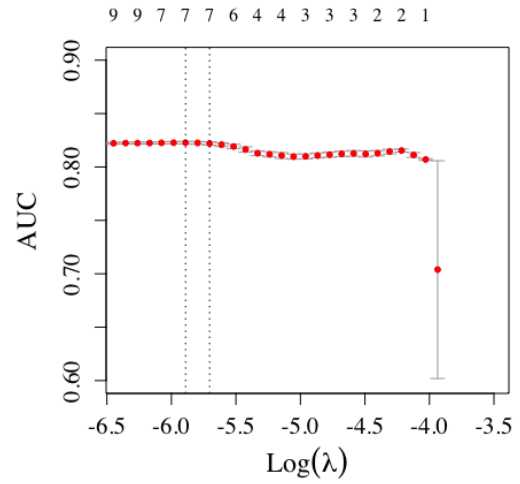
Variabler valgt av LASSO

- d_fravalg_revisjon
- d_bistand_regn_forer
- d_dl_i_styre
- d_to_dl
- d_dl_fylke
- d_sl_fylke
- styre_andel_fylke
- aksj_std
- aksj_min
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

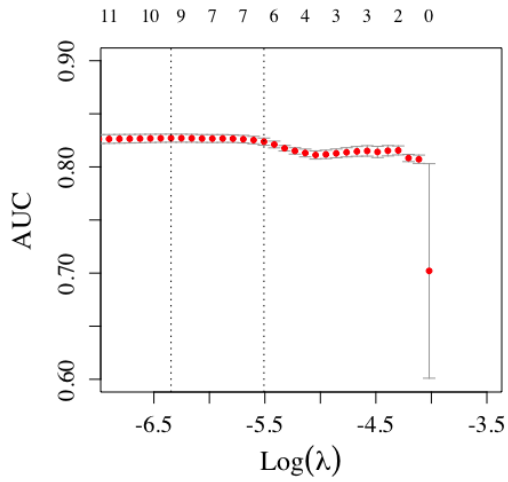
Modell 2B – testår 2010



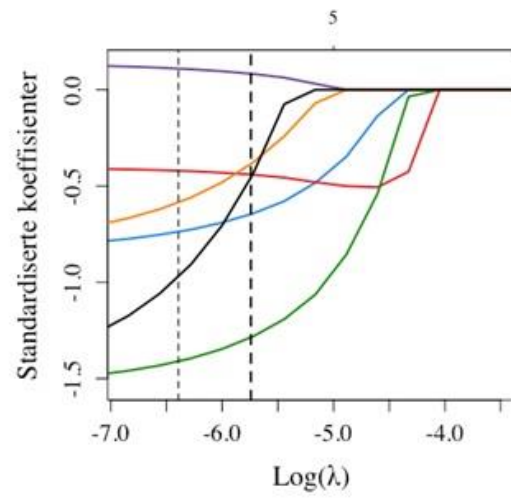
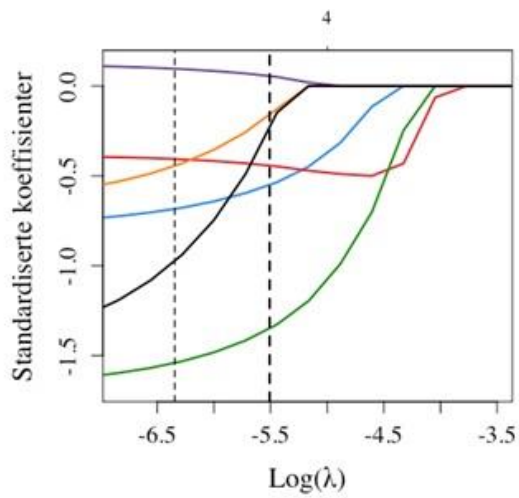
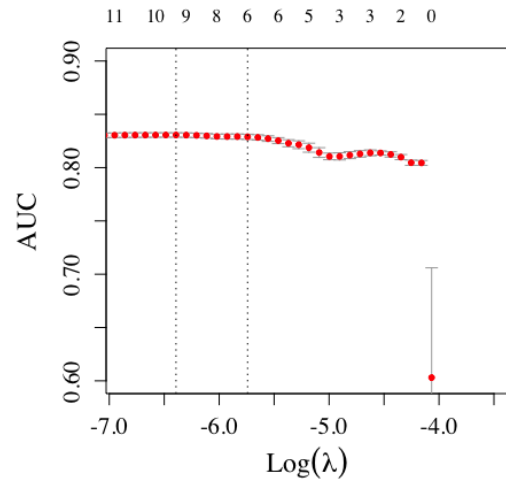
Modell 2B – testår 2011



Modell 2B – testår 2012



Modell 2B – testår 2013



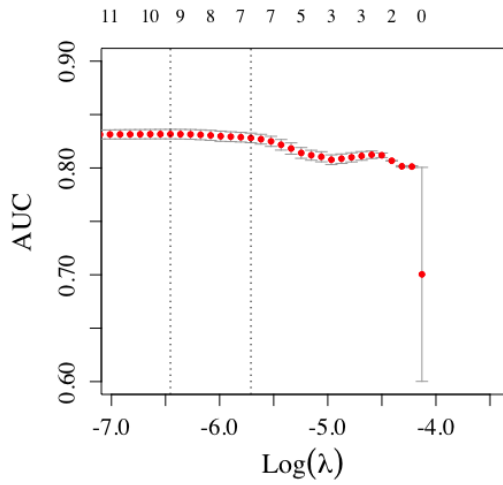
Variabler valgt av LASSO

- arbkap_tk
- opptj_ek_tk
- EBIT_tk
- salginntekt_tk
- d_dl_i_styre
- log_styre_alder_snitt

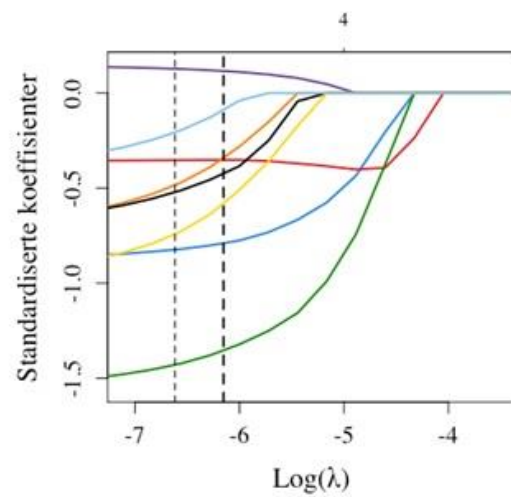
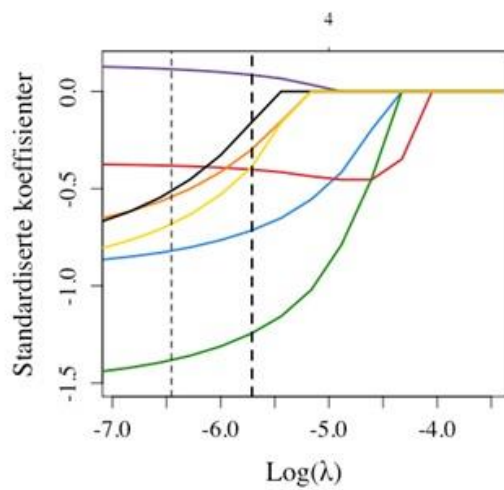
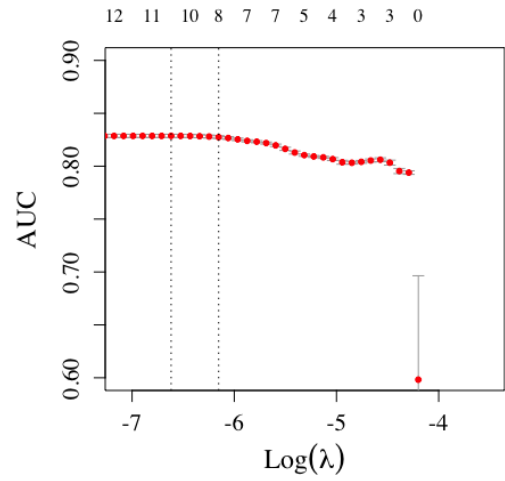
Variabler valgt av LASSO

- arbkap_tk
- opptj_ek_tk
- EBIT_tk
- salginntekt_tk
- d_dl_i_styre
- log_styre_alder_snitt

Modell 2B – testår 2014



Modell 1B – testår 2015



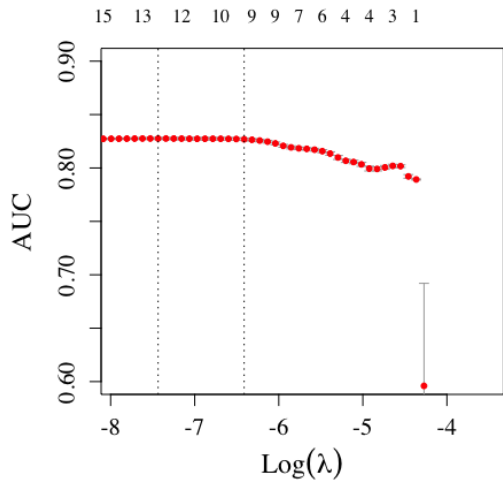
Variabler valgt av LASSO

- arbkap_tk
- opptj_ek_tk
- EBIT_tk
- salginntekt_tk
- d_dl_i_styre
- log_alder_sl
- log_styre_alder_snitt

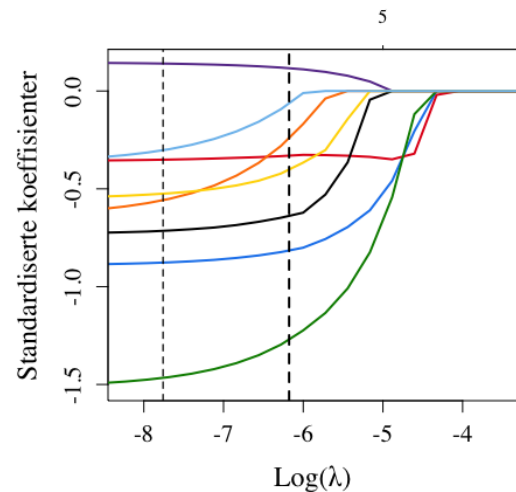
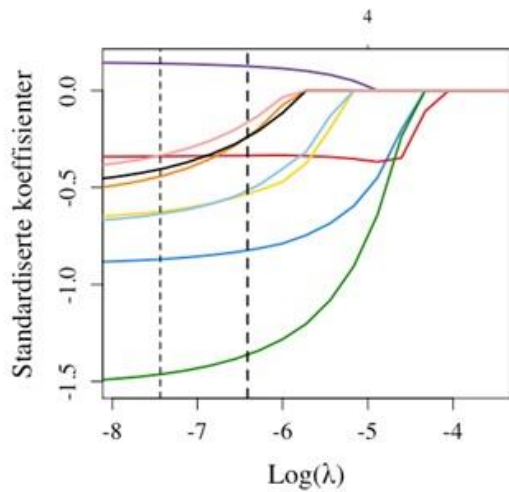
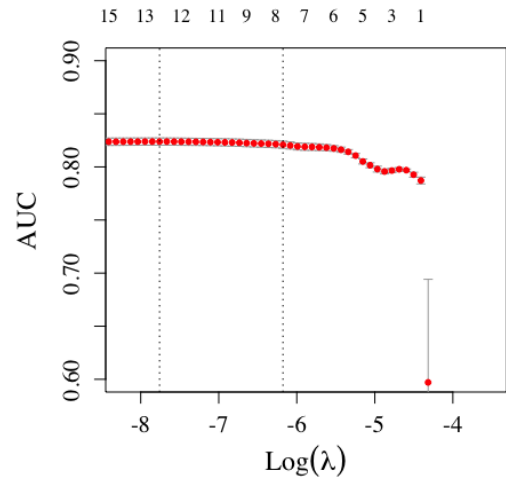
Variabler valgt av LASSO

- arbkap_tk
- opptj_ek_tk
- EBIT_tk
- salginntekt_tk
- d_dl_i_styre
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

Modell 2B – testår 2016



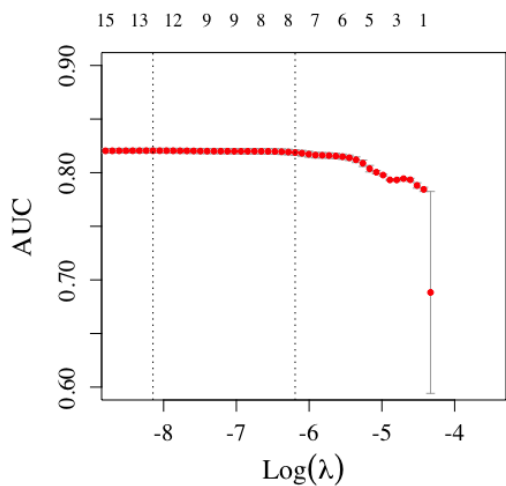
Modell 2B – testår 2017



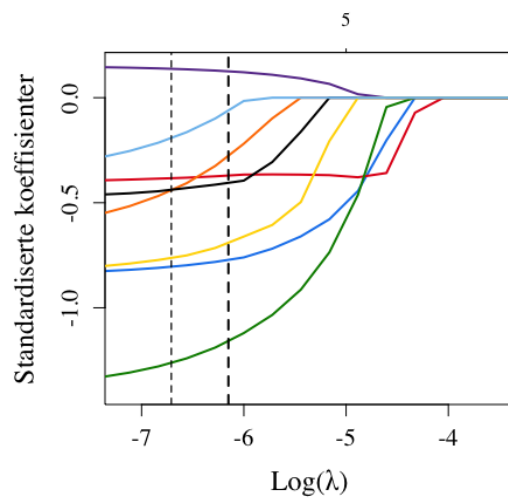
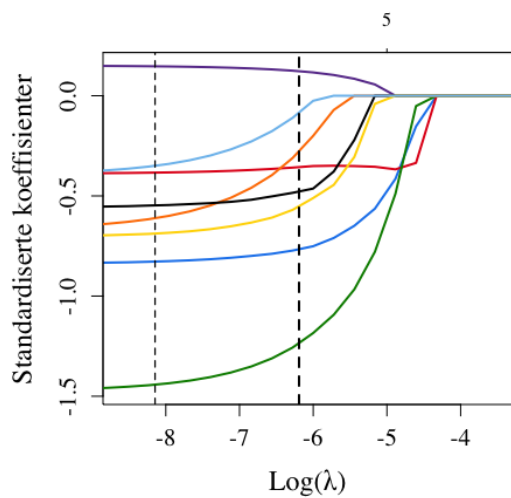
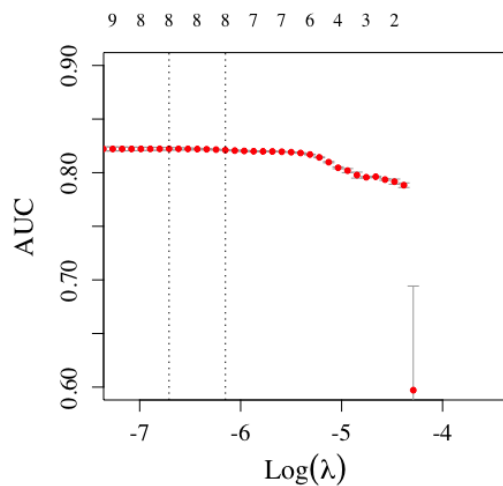
- Variabler valgt av LASSO**
- arbkap_tk
 - opptj_ek_tk
 - EBIT_tk
 - salginntekt_tk
 - d_dl_i_styre
 - log_alder_dl
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall

- Variabler valgt av LASSO**
- arbkap_tk
 - opptj_ek_tk
 - EBIT_tk
 - salginntekt_tk
 - log_alder_dl
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall

Modell 2B – testår 2018



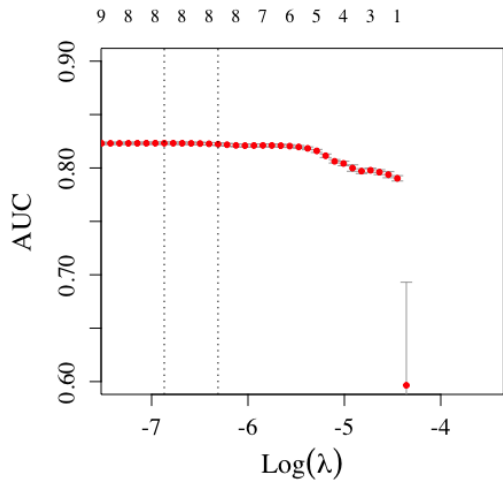
Modell 2B – testår 2019



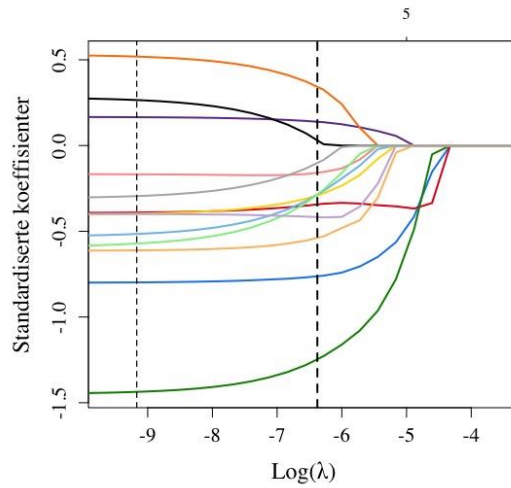
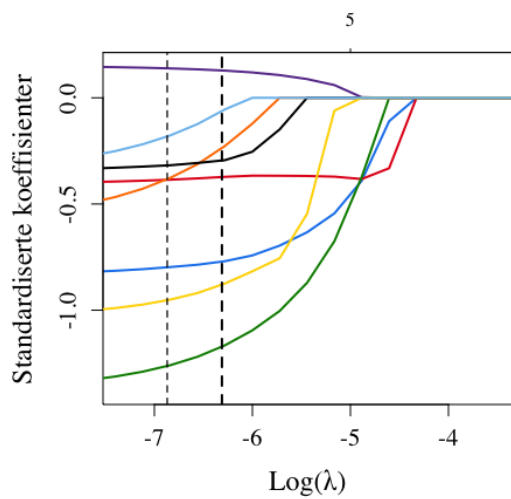
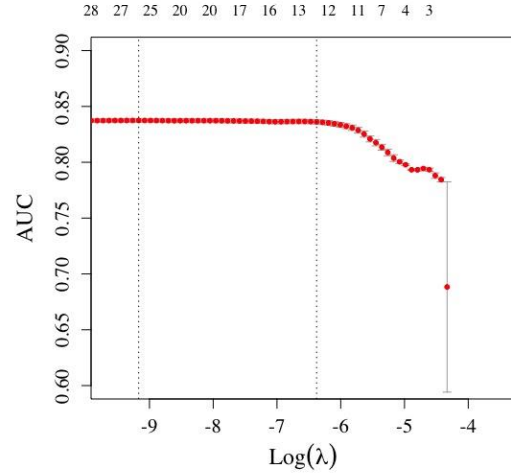
- Variabler valgt av LASSO**
- arbkap_tk
 - opptj_ek_tk
 - EBIT_tk
 - salginntekt_tk
 - log_alder_dl
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall

- Variabler valgt av LASSO**
- arbkap_tk
 - opptj_ek_tk
 - EBIT_tk
 - salginntekt_tk
 - log_alder_dl
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall

Modell 2B – testår 2020



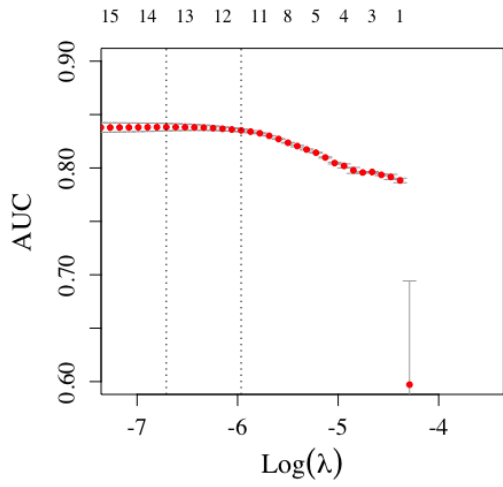
Modell 2C – testår 2018



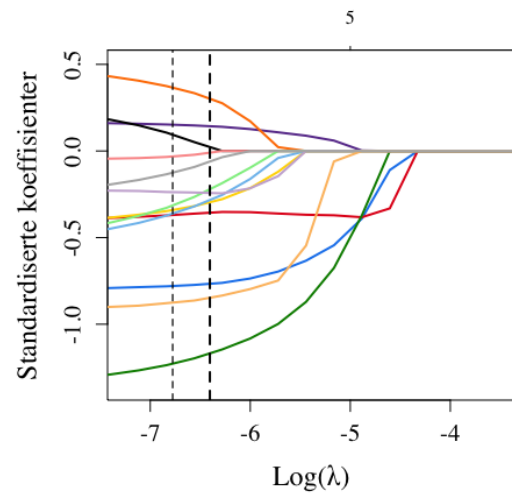
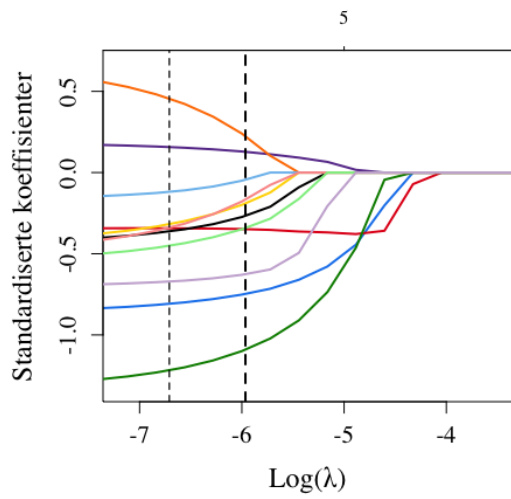
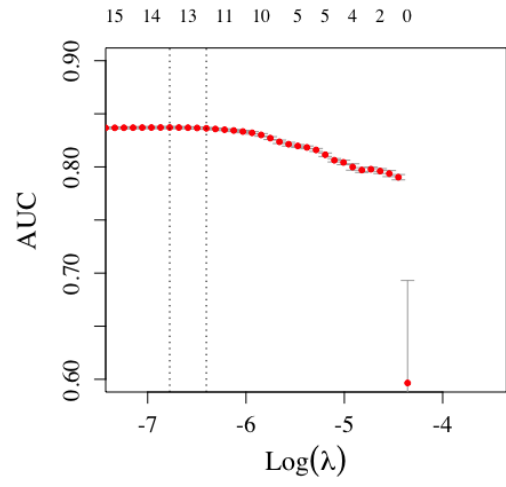
- Variabler valgt av LASSO**
- arbkap_tk
 - opptj_ek_tk
 - EBIT_tk
 - salginntekt_tk
 - log_alder_dl
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall

- Variabler valgt av LASSO**
- arbkap_tk
 - opptj_ek_tk
 - EBIT_tk
 - salginntekt_tk
 - d_fravalg_revisjon
 - styre_eierandel
 - d_dl_fylke
 - d_sl_fylke
 - styre_andel_fylke
 - log_alder_dl
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall

Modell 2C – testår 2019



Modell 2C – testår 2020



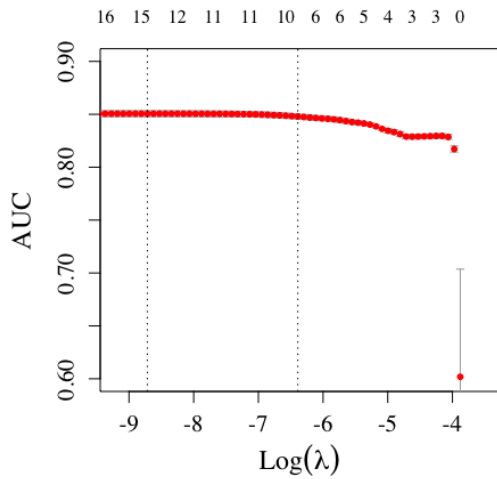
Variabler valgt av LASSO

- arbkap_tk
- opptj_ek_tk
- EBIT_tk
- salginntekt_tk
- d_fravalg_revisjon
- d_dl_fylke
- d_sl_fylke
- styre_andel_fylke
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt

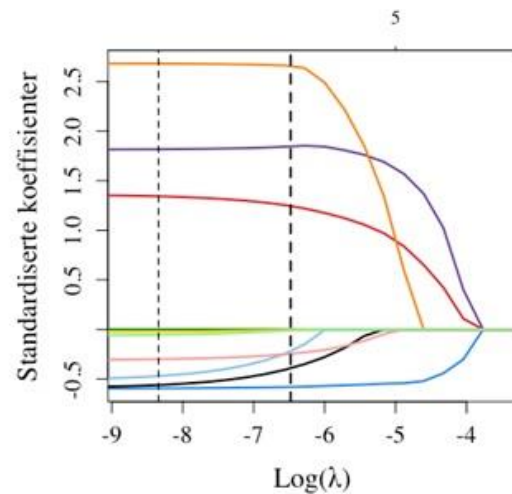
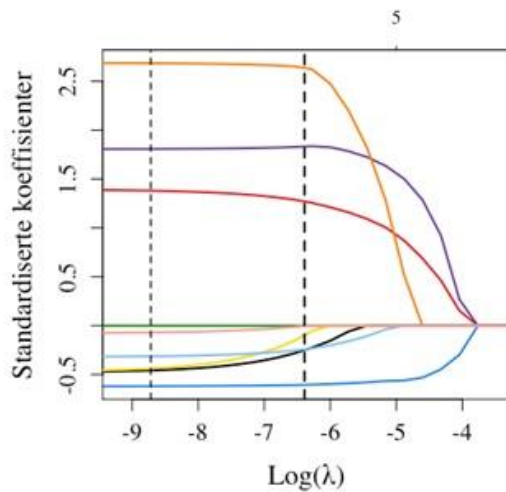
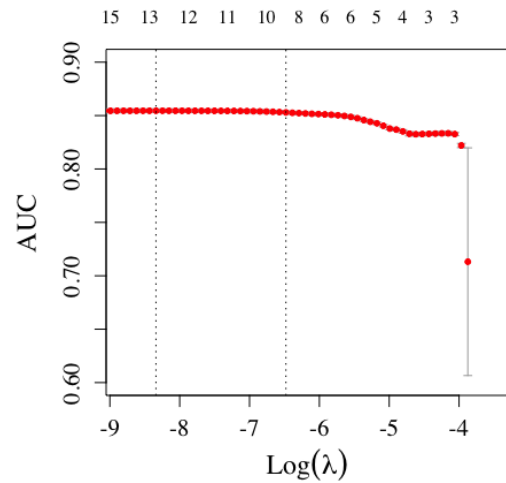
Variabler valgt av LASSO

- arbkap_tk
- opptj_ek_tk
- EBIT_tk
- salginntekt_tk
- d_fravalg_revisjon
- styre_eierandel
- d_dl_fylke
- d_sl_fylke
- styre_andel_fylke
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall

Modell 3B – testår 2010



Modell 3B – testår 2011



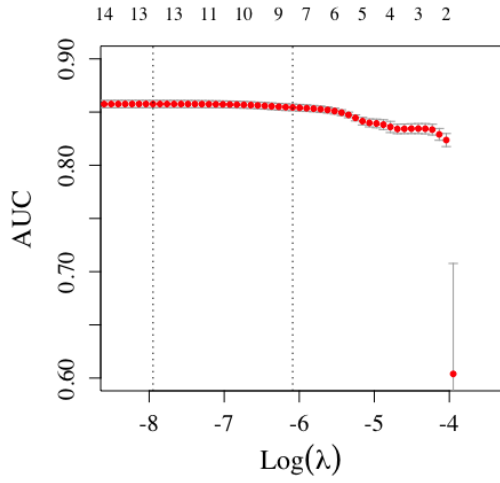
Variabler valgt av LASSO

- ek_tk
- d_bek_iek
- likv_kg_oms
- levgjeld_tk
- avg_tk
- d_dl_i_styre
- log_styre_alder_snitt
- log_alder
- log_tk

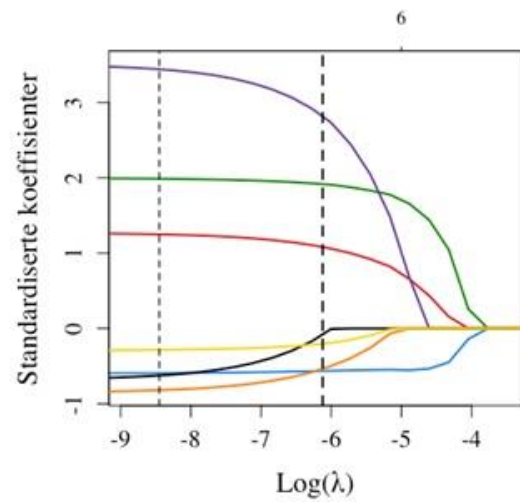
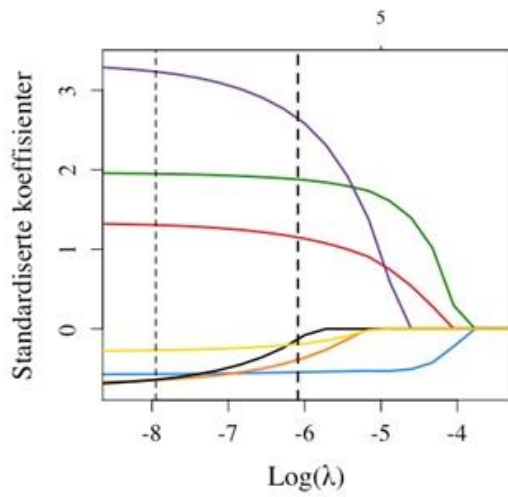
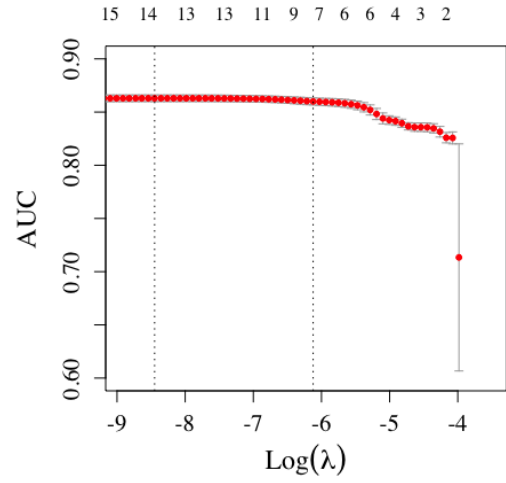
Variabler valgt av LASSO

- ek_tk
- d_bek_iek
- likv_kg_oms
- levgjeld_tk
- avg_tk
- d_dl_i_styre
- styre_alder_std
- log_styre_alder_snitt
- log_alder
- log_tk

Modell 3B – testår 2012



Modell 3B – testår 2013



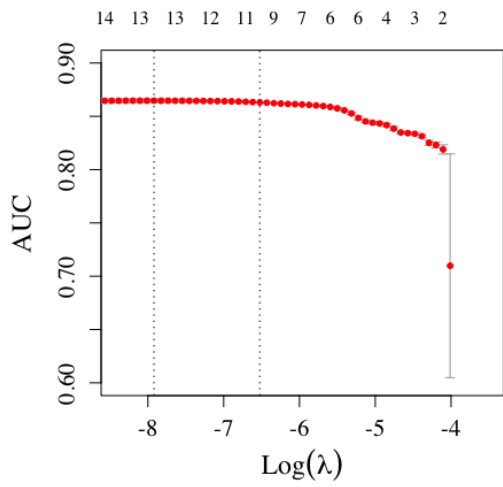
Variabler valgt av LASSO

- ek_tk
- d_bek_iek
- levgjeld_tk
- avg_tk
- d_dl_i_styre
- log_styre_alder_snitt
- log_alder

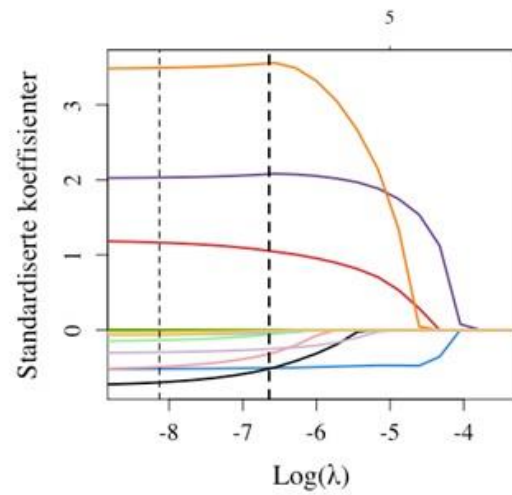
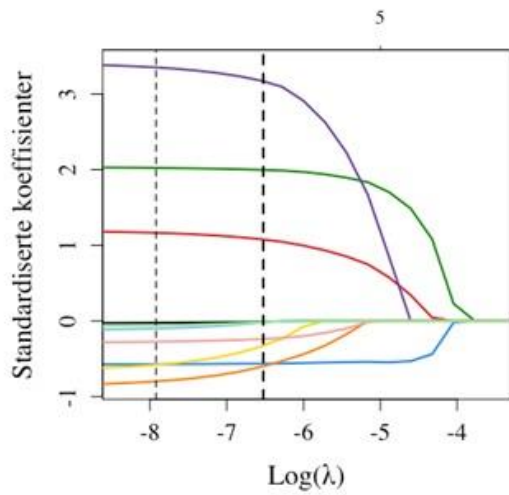
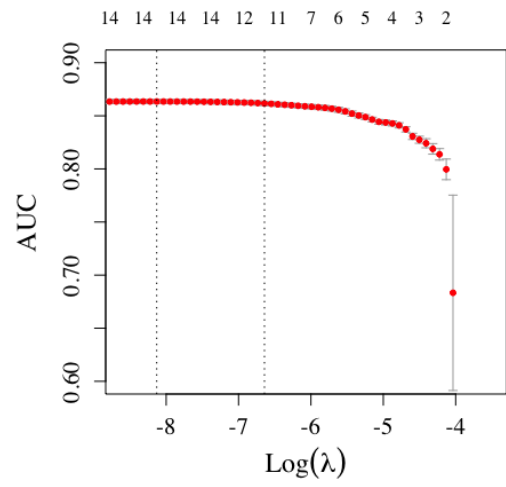
Variabler valgt av LASSO

- ek_tk
- d_bek_iek
- levgjeld_tk
- avg_tk
- d_dl_i_styre
- log_styre_alder_snitt
- log_alder

Modell 3B – testår 2014



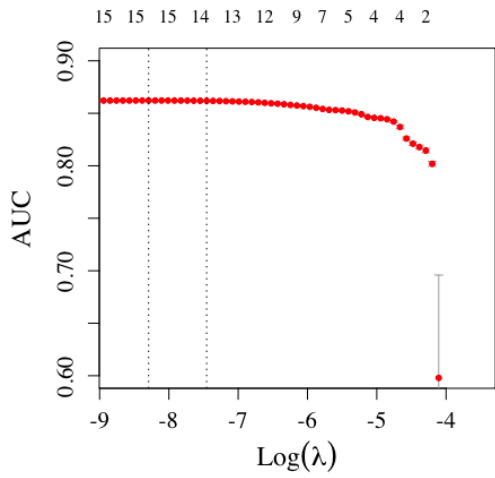
Modell 3B – testår 2015



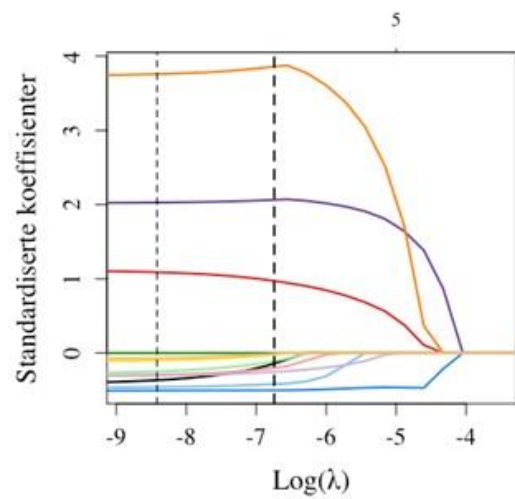
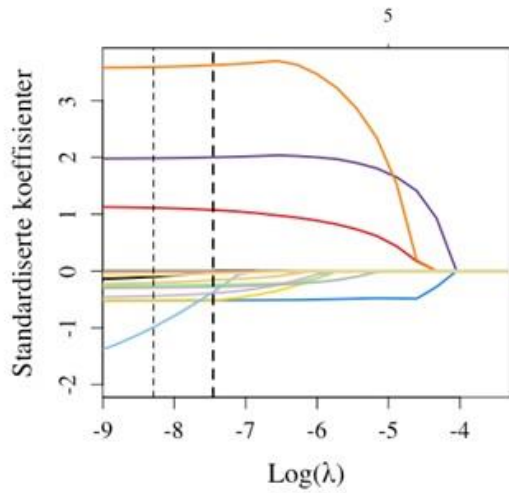
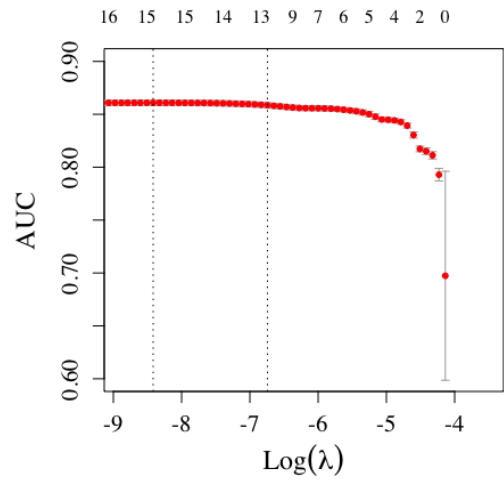
- Variabler valgt av LASSO**
- ek_tk
 - d_bek_iek
 - levgjeld_tk
 - avg_tk
 - d_dl_i_styre
 - styre_alder_std
 - log_styre_alder_snitt
 - log_styre_antall
 - log_alder
 - log_tk

- Variabler valgt av LASSO**
- ek_tk
 - d_bek_iek
 - likv_kg_oms
 - levgjeld_tk
 - avg_tk
 - d_dl_i_styre
 - styre_alder_std
 - log_alder_sl
 - log_styre_alder_snitt
 - log_styre_antall
 - log_alder
 - log_tk

Modell 3B – testår 2016



Modell 3B – testår 2017



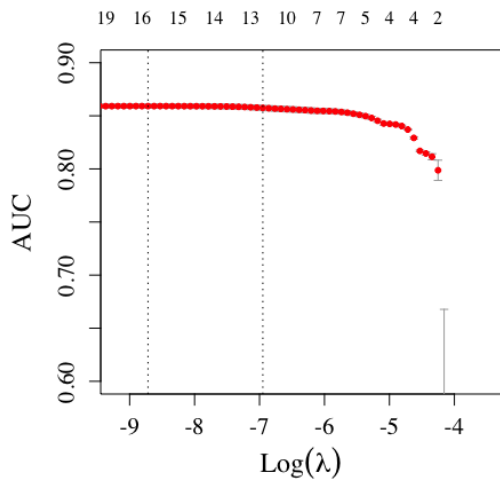
Variabler valgt av LASSO



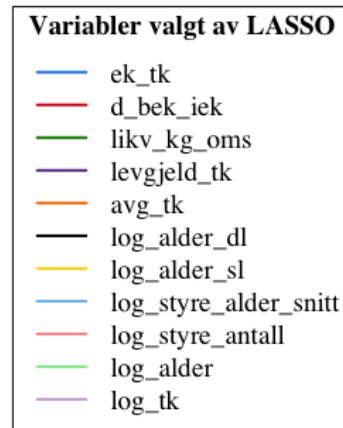
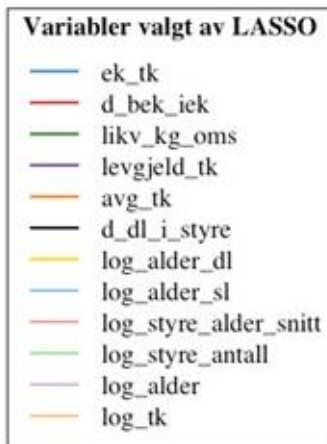
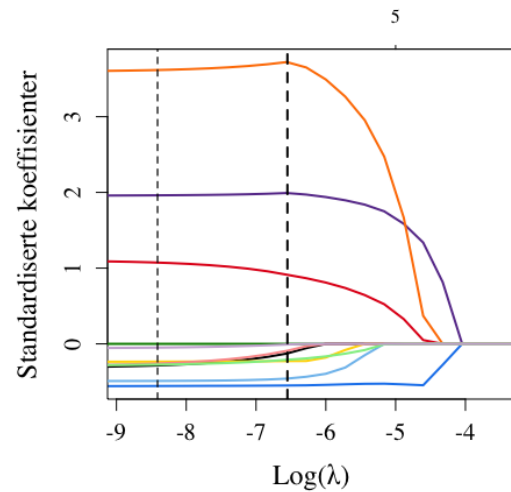
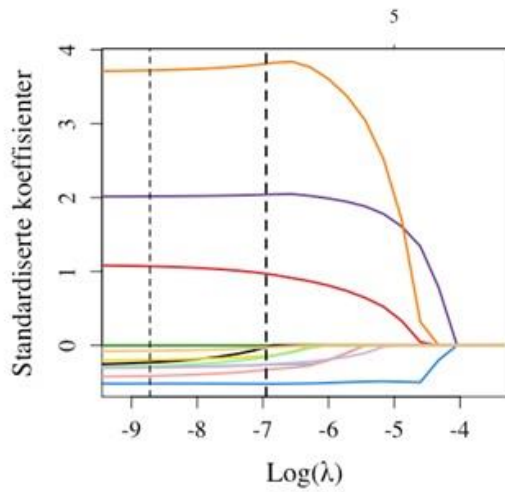
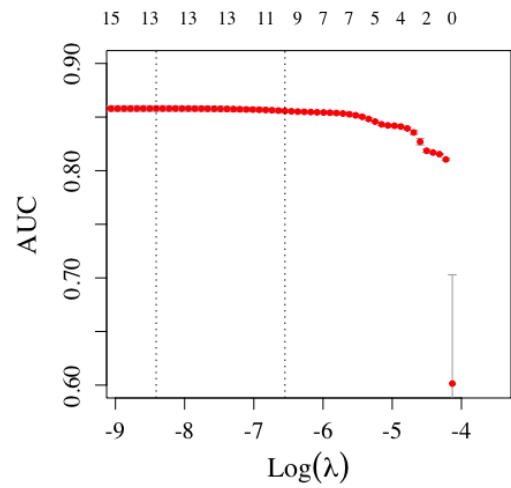
Variabler valgt av LASSO



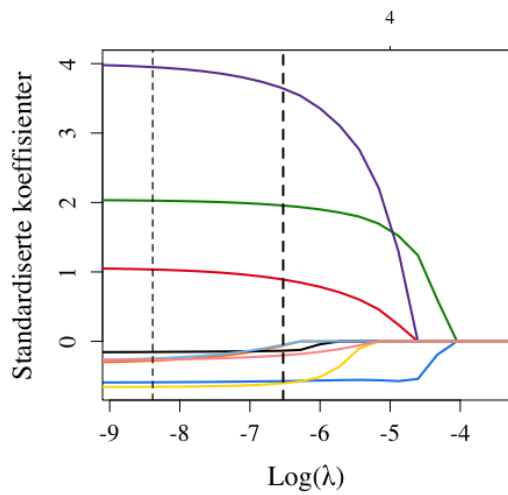
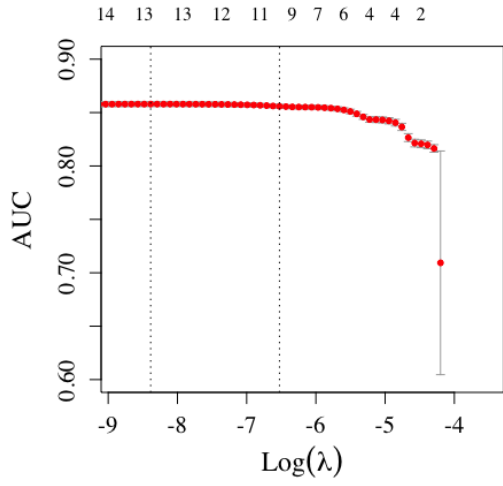
Modell 3B – testår 2018



Modell 3B – testår 2019



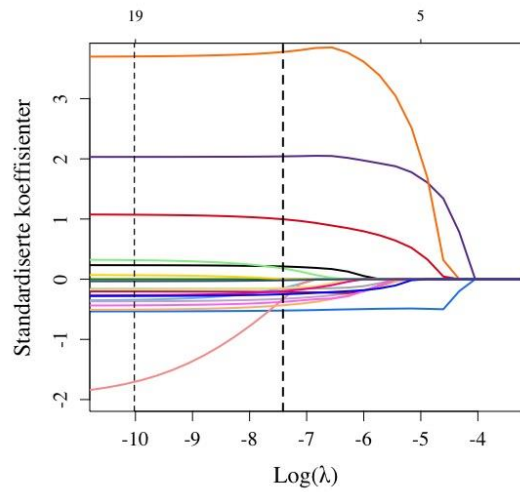
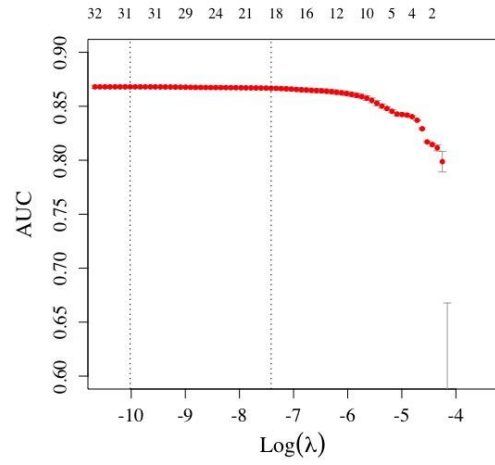
Modell 3B – testår 2020



Variabler valgt av LASSO

- ek_tk
- d_bek_iek
- levgjeld_tk
- avg_tk
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall
- log_alder

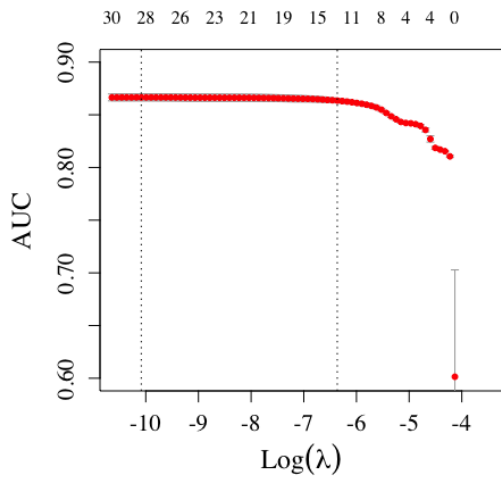
Modell 3C – testår 2018



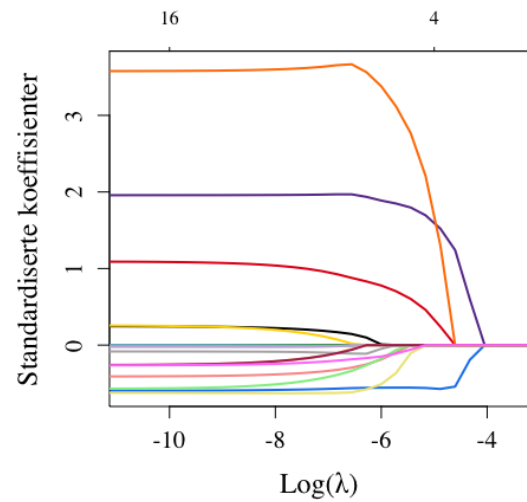
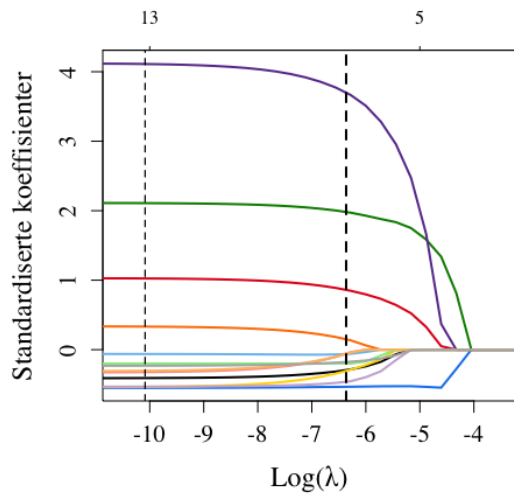
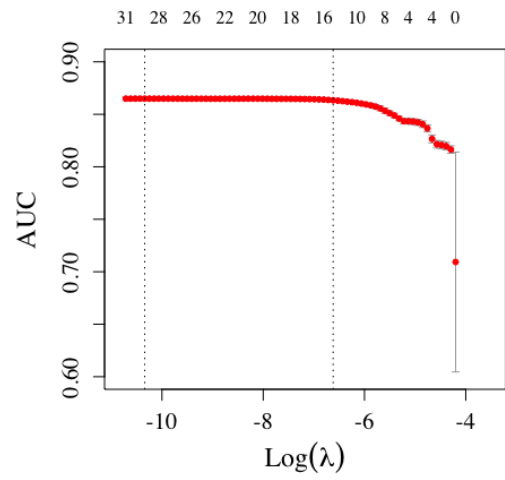
Variabler valgt av LASSO

- ek_tk
- d_bek_iek
- likv_kg_oms
- levgjeld_tk
- avg_tk
- d_fravalg_revisjon
- d_bistand_regn_forer
- d_dl_in_styre
- d_to_dl
- styre_eierandel
- d_dl_fylke
- d_sl_fylke
- styre_andel_fylke
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall
- log_alder
- log_tk

Modell 3C – testår 2019



Modell 3C – testår 2020



Variabler valgt av LASSO

- ek_tk
- d_bek_iek
- levgjeld_tk
- avg_tk
- d_fravalg_revisjon
- d_dl_fylke
- d_sl_fylke
- styre_andel_fylke
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall
- log_alder

Variabler valgt av LASSO

- ek_tk
- d_bek_iek
- likv_kg_oms
- levgjeld_tk
- avg_tk
- d_fravalg_revisjon
- styre_eierandel
- andel_st_medl_ikke_aksj
- d_dl_fylke
- d_sl_fylke
- styre_andel_fylke
- log_alder_dl
- log_alder_sl
- log_styre_alder_snitt
- log_styre_antall
- log_alder

Appendiks E

I de følgende seks tabellene vil det foreligge en presentasjon av rekkefølgen de ulike variablene blir valgt ut av LASSO for de ulike årene. Variablene som blir valgt ut først blir ansett for å være de viktigste variablene i modellene. For å skille på finansielle og ikke-finansielle variabler blir de ikke-finansielle variablene markert med grå bakgrunn. Det er ikke benyttet LASSO til modell 2A og 3A, og disse modellene vil følges ikke bli nevnt.

Modell 1A

	2010	2011	2012	2013	2014
1	d_dl_i_styre	d_dl_i_styre	d_dl_i_styre	d_dl_i_styre	d_dl_i_styre
2	log_alder_sl	log_alder_sl	log_alder_sl	log_styre_alder_snitt	log_alder_sl
3	log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt	log_alder_sl	log_styre_alder_snitt
4	d_samme_dl_sl	log_alder_dl	log_alder_dl	log_alder_dl	log_alder_dl
5	log_alder_dl	d_samme_dl_sl	d_samme_dl_sl	styre_alder_std	styre_alder_std
6	styre_alder_std	styre_alder_std	styre_alder_std	d_samme_dl_sl	log_styre_antall
7	d_dl_kjonn_k			log_styre_antall	d_samme_dl_sl
	d_to_dl				

	2015	2016	2017	2018	2019	2020
1	d_dl_i_styre	log_alder_sl	log_alder_sl	log_alder_sl	log_styre_alder_snitt	log_styre_alder_snitt
2	log_alder_sl	log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt	log_alder_dl	log_alder_dl
3	log_styre_alder_snitt	log_alder_dl	log_alder_dl	log_alder_dl	log_alder_sl	log_alder_sl
4	log_alder_dl	d_dl_i_styre	d_dl_i_styre	log_styre_antall	log_styre_antall	log_styre_antall
5	styre_alder_std	log_styre_antall	log_styre_antall	d_dl_i_styre	d_dl_i_styre	d_dl_i_styre
6	log_styre_antall	styre_alder_std	styre_alder_std	d_samme_dl_sl	d_to_dl	d_to_dl
7	d_samme_dl_sl	d_samme_dl_sl	d_samme_dl_sl	d_to_dl		d_dl_kjonn_k
8		d_to_dl	d_to_dl			
9		d_dl_kjonn_k				

Modell 1B

	2018	2019	2020
1	log_alder_sl	log_styre_alder_snitt	log_styre_alder_snitt
2	log_styre_alder_snitt	d_dl_fylke	d_dl_fylke
3	d_dl_fylke	d_sl_fylke	log_alder_dl
4	d_sl_fylke	log_alder_dl	log_alder_sl
5	styre_andel_fylke	log_alder_sl	d_sl_fylke
6	log_alder_dl	d_fravalg_revisjon	d_fravalg_revisjon
7	d_fravalg_revisjon	styre_andel_fylke	styre_andel_fylke
8	log_styre_antall	d_bistand_regn_forer	d_bistand_regn_forer
9	d_bistand_regn_forer	log_styre_antall	log_styre_antall
10	d_dl_i_styre	d_dl_i_styre	aksj_std
11	d_to_dl	d_to_dl	d_dl_i_styre
12	d_samme_dl_sl	aksj_std	d_to_dl
13	styre_eierandel	aksj_min	aksj_min
14	aksj_std		
15	styre_alder_std		
16	aksj_min		
17	d_dl_kjonn_k		
18	dl_eierandel		
19	andel_st_medl_ikke_aksj		
20	log_aksj_antall		
21	sl_eierandel		

Modell 2B

	2010	2011	2012	2013	2014
1	opptj_ek_tk	opptj_ek_tk	opptj_ek_tk	opptj_ek_tk	opptj_ek_tk
2	EBIT_tk	EBIT_tk	EBIT_tk	EBIT_tk	arbkap_tk
3		arbkap_tk	arbkap_tk	arbkap_tk	EBIT_tk
4		salgsinntekt_tk	salgsinntekt_tk	salgsinntekt_tk	salgsinntekt_tk
5		d_dl_i_styre	d_dl_i_styre	d_dl_i_styre	d_dl_i_styre
6		log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt
7		log_alder_sl			log_alder_sl

	2015	2016	2017	2018	2019	2020
1	opptj_ek_tk	opptj_ek_tk	opptj_ek_tk	arbkap_tk	opptj_ek_tk	arbkap_tk
2	arbkap_tk	arbkap_tk	arbkap_tk	opptj_ek_tk	arbkap_tk	opptj_ek_tk
3	EBIT_tk	EBIT_tk	EBIT_tk	EBIT_tk	EBIT_tk	EBIT_tk
4	salgsinntekt_tk	salgsinntekt_tk	salgsinntekt_tk	salgsinntekt_tk	salgsinntekt_tk	salgsinntekt_tk
5	log_alder_sl	log_alder_sl	log_alder_sl	log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt
6	log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt	log_alder_sl	log_alder_sl	log_alder_sl
7	d_dl_i_styre	d_dl_i_styre	log_alder_dl	log_alder_dl	log_alder_dl	log_alder_dl
8	log_styre_antall	log_alder_dl	log_styre_antall	log_styre_antall	log_styre_antall	log_styre_antall
9		log_styre_antall				

Modell 2C

	2018	2019	2020
1	arbkap_tk	opptj_ek_tk	arbkap_tk
2	opptj_ek_tk	arbkap_tk	opptj_ek_tk
3	EBIT_tk	EBIT_tk	EBIT_tk
4	salgsinntekt_tk	salgsinntekt_tk	salgsinntekt_tk
5	log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt
6	d_dl_fylke	d_dl_fylke	d_fravalg_revisjon
7	d_sl_fylke	log_alder_sl	d_dl_fylke
8	log_alder_sl	d_fravalg_revisjon	d_sl_fylke
9	d_fravalg_revisjon	d_sl_fylke	log_alder_sl
10	styre_andel_fylke	log_alder_dl	log_alder_dl
11	log_alder_dl	styre_andel_fylke	log_styre_antall
12	log_styre_antall		styre_eierandel
13	styre_eierandel		styre_andel_fylke

Modell 3B

	2010	2011	2012	2013	2014	2015
1		ek_tk	ek_tk	ek_tk	ek_tk	ek_tk
2		d_bek_iek	d_bek_iek	levgjeld_tk	levgjeld_tk	levgjeld_tk
3		levgjeld_tk	levgjeld_tk	d_bek_iek	d_bek_iek	d_bek_iek
4		avg_tk	avg_tk	avg_tk	avg_tk	avg_tk
5		log_alder	log_alder	log_alder	d_dl_in_styre	log_alder
6		d_dl_in_styre	d_dl_in_styre	d_dl_in_styre	log_alder	d_dl_in_styre
7		likv_kg_oms	log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt	log_styre_alder_snitt
8		log_styre_alder_snitt	likv_kg_oms			log_styre_alder_std
9		log_tk	styre_alder_std			log_styre_antall
10			log_tk			log_tk
	2016	2017	2018	2019	2020	
1	levgjeld_tk	ek_tk	ek_tk	ek_tk	ek_tk	ek_tk
2	ek_tk	levgjeld_tk	levgjeld_tk	levgjeld_tk	levgjeld_tk	levgjeld_tk
3	d_bek_iek	d_bek_iek	d_bek_iek	d_bek_iek	d_bek_iek	d_bek_iek
4	avg_tk	avg_tk	avg_tk	avg_tk	avg_tk	avg_tk
5	log_alder	log_alder	log_alder	log_alder	log_alder	log_styre_alder_snitt
6	d_dl_in_styre	d_dl_in_styre	log_alder_sl	log_alder_sl	log_styre_alder_snitt	log_alder
7	log_styre_alder_snitt	log_alder_sl	log_styre_alder_snitt	log_styre_alder_snitt	log_alder_sl	log_alder_sl
8	styre_alder_std	log_styre_alder_snitt	log_styre_antall	log_alder_dl	log_alder_dl	log_alder_dl
9	log_alder_sl	log_styre_antall	d_dl_in_styre	log_styre_antall	log_styre_antall	log_styre_antall
10	log_styre_antall	likv_kg_oms	log_alder_dl	log_tk	likv_kg_oms	
11	likv_kg_oms	styre_alder_std	log_tk	likv_kg_oms	log_tk	
12	log_tk	log_tk	likv_kg_oms	d_dl_in_styre		
13		d_to_dl				
14		d_samme_dl_sl				

Modell 3C

	2018	2019	2020
1	ek_tk	ek_tk	ek_tk
2	levgjeld_tk	levgjeld_tk	levgjeld_tk
3	d_bek_iek	d_bek_iek	d_bek_iek
4	avg_tk	avg_tk	avg_tk
5	log_alder	log_alder	log_styre_alder_snitt
6	d_dl_fylke	d_dl_fylke	log_alder
7	d_sl_fylke	d_sl_fylke	d_dl_fylke
8	styre_andel_fylke	log_styre_alder_snitt	d_sl_fylke
9	log_alder_sl	log_alder_sl	d_fravalg_revisjon
10	log_styre_alder_snitt	d_fravalg_revisjon	log_alder_sl
11	d_fravalg_revisjon	styre_andel_fylke	styre_eierandel
12	log_styre_antall	log_alder_dl	andel_st_medl_ikke_aksj
13	styre_eierandel	log_styre_antall	styre_andel_fylke
14	log_alder_dl		log_alder_dl
15	log_tk		log_styre_antall
16	likv_kg_oms		likv_kg_oms
17	d_dl_i_styre		
18	d_to_dl		
19	d_bistand_regn_forer		

