



Stockholms
universitet

Legitimacy of newness and smallness - En studie i överlevnad för små och nya företag

Cecilia Söderberg

Kandidatuppsats 2015:11
Matematisk statistik
Juni 2015

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Legitimacy of newness and smallness - En studie i överlevnad för små och nya företag

Cecilia Söderberg*

Juni 2015

Sammanfattning

I denna uppsats har vi undersökt hur ett företags ålder och storlek påverkat dess risk för att gå i konkurs. Speciellt har vi undersökt om små företag är mer benägna att gå i konkurs än stora företag och om yngre företag har en större benägenhet att gå i konkurs än äldre. För att undersöka dessa frågeställningar har vi använt oss av logistik regression och överlevnadsanalys. Metoder inom överlevnadsanalys som har använts är Kaplan-Meier skattning och Log-rank test. Data som har använts för att göra den statistiska analysen är ett 10% urval från databasen Serrano. Serrano är en registerdatabas om svenska företag som bland annat innehåller information från företagens bokslut. I detta register finns information om företag som har lämnat in ett bokslut under tidsperioden 1998 till 2013. Det visar sig av den statistiska analysen ett företags storlek har en inverkan på företagets benägenhet att gå i konkurs. Från båda analysmetoderna visar det sig att företag av storlek 1-9 anställda är de företag som har störst risk att gå i konkurs. Då det upptäcktes att det förekommer vänstertrunkering för företagens ålder kan det inte göras några tolkningar om företagens ålder har en påverkan på benägenheten att gå i konkurs i den här uppsatsen.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: cissi_soderberg@hotmail.com. Handledare: Jan-Olov Persson, Martin Sköld.

Acknowledgments

Jag vill rikta ett stort tack till Per Weidenman på Bisnode för att jag har fått tillgång till data och för att jag fått bolla idéer med honom. Jag vill även tacka min handledare Jan-Olov Persson för bra litteraturtips och bra feedback som fört arbetet framåt. Även Martin Sköld har varit till en stor hjälp för att sammanställa detta arbete. Slutligen vill jag tacka mina kurskamrater för ett otroligt stöd och bollplank.

Innehåll

Sammanfattning	i
Acknowledgments	ii
1 Inledning	2
2 Data	4
2.1 Information om data	4
2.1.1 Olika typer av bolag	5
2.2 Beskrivning av data	6
2.3 Bearbetning av data	7
3 Metod	9
3.1 Censur	10
3.2 Trunkering	10
3.3 Begrepp och samband inom överlevnadsanalys	10
3.4 Kaplan-Meier skattning	12
3.4.1 Log rank test och Wilcoxon test	13
3.5 Cox-regression	15
3.6 Logistisk regression	16
3.6.1 Tolka statistikor	17
4 Resultat från statistisk analys och diskussion	18
4.1 Kaplan Meier	18
4.1.1 Log rank test och Wilcoxon test	21
4.2 Cox proportional hazard modell	22
4.3 Logistisk regression	22
4.4 Enmansföretagen	24
5 Slutsats	25
A Appendix	27
A.1 Definition av aktivt företag	27
A.2 Juridisk form	27
Litteraturförteckning	28

1. Inledning

Har små företag en större risk för att gå i konkurs än stora och har nystartade företag en större risk att gå i konkurs än äldre? I denna uppsats kommer vi försöka besvara dessa frågor. Som bakgrund till uppsatsen är kända begrepp inom organisationsteorin, *legitimacy of newness* och *legitimacy of smallness*. *Legitimacy of newness* innebär att äldre företag har större chanser att överleva än nystartade företag och *legitimacy of smallness* innebär att stora företag har en större chans att överleva än små företag. Enligt dessa teorier finns det flera möjliga förklaringar till varför det går bättre för äldre och stora företag. Ett nystartat företag stöter på många nya utmaningar. De måste skapa nya rutiner och relationer med intressenter. Eftersom de är nya måste de även skapa ett förtroende då intressenter inte kan bedöma företaget på grund av hur det har skött sig i tidigare samarbeten. En förklaring till varför stora företag klarar sig bättre på marknaden kan vara att stora företag anses vara mera pålitliga. Att de är stora kan av intressenter anses som att de tidigare varit framgångsrika och därmed är de mer pålitliga (Clegg et al., 2003).

Vi ska i den här uppsatsen undersöka om teorierna i föregående stycke går att applicera på svenska företag. Har ett företags storlek och ålder påverka på dess överlevnad? Som överlevnad i den här uppsatsen menas att ett företag inte går i konkurs. Frågeställningarna som ska undersökas i detta arbete är om små företag är mer benägna att gå i konkurs än stora och om företagets ålder påverkar benägenheten att gå i konkurs.

Vi använder Sveriges domstolar definition för att definiera vad det innebär att gå i konkurs. Konkurs innebär att en person eller ett företags tillgångar tas hand om för att användas för att betala av skulder som företaget eller personen har. En person eller ett företag sätts i konkurs om personen eller företaget inte kan betala sina skulder i tid och inte kommer kunna betala dem under en överskådlig framtid. Det är tingsrätten som fattar beslutet att ett företag eller en person ska gå i konkurs. Om det är ett företag som går i konkurs så kommer det efter avslutad konkurs bli upplöst (Sveriges domstolar).

För att undersöka företagens överlevnad kommer metoderna logistisk regression och överlevnadsanalys användas. De metoder som kommer användas inom överlevnadsanalysen är Kaplan-Meier skattning, Log rank test eller Wilcoxon's test och om vissa antaganden är uppfyllda, Cox proportional hazard modell. Samtliga analyser kommer göras i den statistiska programvaran SAS. Data som har använts kommer från databasen Serrano som innehåller information om svenska företag.

Uppsatsen har följande upplägg; I kapitel 2 kommer data som har använts för

att göra analysen beskrivas, i kapitel 3 kommer teorin för hur man gör överlevnadsanalys och logistisk regression beskrivas samt hur resultaten tolkas, i kapitel 4 kommer resultaten från den statistiska analysen redovisas och tolkas. Slutligen kommer resultaten sammanfattas och diskuteras i kapitel 5.

2. Data

2.1 INFORMATION OM DATA

För att undersöka hur överlevnaden för företag ser ut har data från databasen Serrano använts. Databasen Serrano är skapad av senior analytikern Per Weidenman på Bisnode och innehåller registerdata om företag. Databasen innehåller bland annat all information som finns i bokslut för svenska bolag och uppdateras i december och maj varje år. Informationen hämtas från Bolagsverket, SCBs företagsregister och från företaget Bisnodes register. Registret innehåller en post med information för varje år ett företag lämnar in ett bokslut från 1998 till och med det senaste bokslutet fram till år 2013. Detta innebär att om ett företag har blivit inregistrerat år N och avregistrerat år $N + i$ finns det en rad med information för var och en av de åren. Databasen har anpassats så att det alltid går att ta informationen årsvis vilket underlättar analysen om till exempel ett företag har ett brutet räkenskapsår, då har informationen räknats om till motsvarande värden för ett helår.

Ett urval på ca 10% av alla bolag i Serrano har valts ut till den här uppsatsen. Då datamängden i Serrano är väldigt stor har ett stickprov valts ut för att göra den statistiska analysen. Företag har valts ut genom att de organisationsnummer som är jämnt delbara med 10 har valts ut. Detta stickprov resulterar i 70 319 företag vilket motsvarar ca 10% av alla bolag i Sverige.

Totalt finns det 177 olika variabler i Serrano. Då alla inte är väsentliga för detta arbete beskriver vi endast följande variabler:

Konkurs, indikator - Antar värde 1 om ett företag har gått i konkurs det året och 0 annars. Det finns ett värde för variabeln för varje år som ett företag har bokslut för mellan åren 1998 och 2013. Om ett företag har gått i konkurs år n kommer denna variabel anta värde 0 fram till år n , då antar den värde 1.

Juridisk form - Anger vilken bolagsform ett företag tillhör. Se Appendix A för de olika bolagsformerna.

Aktiv - Anger om ett företag är aktivt eller inaktivt. Se Appendix A för definition av ett aktivt företag.

Registreringsdatum - Det datum då företaget registrerades hos Bolagsverket.

Bokslutsår - En värde för varje år mellan 1998 och 2013 som ett företag har lämnat in ett bokslut. Datum för bokslut är den sista december varje år.

Ålder - Ett företags ålder räknas som:

Senaste bokslutsdatumet för företaget - registreringsdatumet.

Storleksklass - Antalet anställda på ett företag är sekretessbelagt. Därför anges antalet anställda i intervall. Se Tabell 2.2.1.

För att förtydliga hur variablerna kan se ut för två företag, se Tabell 2.1.1.

För varje år ett bolag har lämnat in ett bokslut under tidsperioden 1998 till 2013 finns en rad med information om bolaget. Organisationsnumret, ålder och registreringsdatumet är konstanta och förändras inte för de olika åren medan de andra variablerna kan förändras. Åldern har räknats fram som skillnaden mellan registreringsdatumet och 31/12 det sista året som det finns ett bokslutet för företaget. Vi ser att det första företaget, med organisationsnummer 5562035500, bara har information till år 2004. Anledningen till detta är för att det fusionerades upp i ett annat bolag år 2004. Det andra bolaget, som har organisationsnummer 5562035500, har gått i konkurs. Vi ser att bolaget har en nolla för varje år fram tills det år då det går i konkurs. Då visas istället en etta i konkursfältet.

Tabell 2.1.1: Exempel på hur de olika variablerna ser ut för två bolag.

Org.nr.	Bolagsform	Bokslutsår	Konkurs	Storleksklass	Registreringsdatum	Aktiv	Ålder
5562035500	49	1998	0	1	1979-11-30	1	25.1
5562035500	49	1999	0	1	1979-11-30	1	25.1
5562035500	49	2000	0	1	1979-11-30	1	25.1
5562035500	49	2001	0	1	1979-11-30	1	25.1
5562035500	49	2002	0	1	1979-11-30	1	25.1
5562035500	49	2003	0	1	1979-11-30	1	25.1
5562035500	49	2004	0	1	1979-11-30	1	25.1
5562038900	49	1998	0	1	1979-11-20	1	25.13
5562038900	49	1999	0	1	1979-11-20	1	25.13
5562038900	49	2000	0	1	1979-11-20	1	25.13
5562038900	49	2001	0	1	1979-11-20	1	25.13
5562038900	49	2002	0	0	1979-11-20	1	25.13
5562038900	49	2003	0	0	1979-11-20	1	25.13
5562038900	49	2004	1	0	1979-11-20	0	25.13

2.1.1 OLIKA TYPER AV BOLAG

I Serrano finns data för flera olika bolagsformer men inte för bolagsformen enskilda firmor. Detta är på grund utav att ekonomin för privatpersonen och företaget inte är åtskild. I denna typ av bolag är ägaren personligt ansvarig för ett företags ekonomi och sätts i personlig konkurs om företaget går i konkurs. Denna typ av information är känslig och enligt Kreditupplysningslagen får vi inte ta ut den information om vi inte har ett legitimt skäl.

Av alla de olika bolagsformerna som finns i Serrano har vi i denna uppsats bara utfört analysen på aktiebolag och handelsbolag. Detta eftersom andra typer av företag har andra förutsättningar än aktiebolag eller handelsbolag som påverkar deras benägenhet att gå i konkurs. Till exempel kan kommuner inte gå i konkurs och bankaktiebolag kan bli räddade av staten om de skulle riskera att gå i konkurs. Då de övriga bolagsformerna tas bort återstår 9774 bolag.

Följande definitioner av aktiebolag och handelsbolag är tagna från Bolagsverket. Enligt Bolagsverket definieras ett aktiebolag som ett företag där ägarna inte har något personligt ansvar för företagets skulder. För att starta ett aktiebolag behövs en kapitalinsats. Ett handelsbolag definieras som ett bolag som ägs av två eller fler personer och där ägarna är ansvariga för företagets skulder. Ett handelsbolag räknas som en juridisk person vilket innebär att det kan teckna avtal och anställa personer. Ett kommanditbolag är en form av handelsbolag och ingår i samma kategori som handelsbolag i Serrano. Skillnaden mellan handelsbolag och kommanditbolag är hur ägandet är uppdelat. Mer information om detta finns på Bolagsverkets hemsida (Bolagsverket).

2.2 BESKRIVNING AV DATA

Hur många företag som finns i varje storleksklass, hur många av dessa som har gått i konkurs samt vilken procentandel det utgör av företagen i den kategorin ses i Tabell 2.2.1. När vi har bearbetat data, se kapitel 2.3, finns det totalt 6427 olika företag och av dessa har 660 gått i konkurs. Vi ser att de flesta företagen tillhör de mindre storleksklasserna, 0 och 1. Procentandelen som går i konkurs är störst för de företag som tillhör storleksklasserna 1, 2 och 3. Vi noterar även att inga företag har gått i konkurs i storleksklass 7.

Tabell 2.2.1: Hur många företag som finns i varje storleksklass och hur många som gått i konkurs per storleksklass.

Storleksklass	Antal anställda	Antal företag	Antal som gått i konkurs	Procent som gått i konkurs
0	0	2415	116	4.80
1	1-4	3058	427	13.96
2	5-9	498	71	14.26
3	10-19	241	30	12.45
4	20-49	125	11	8.80
5	50-99	37	2	5.41
6	100-199	17	1	5.88
7	≥ 200	9	0	0
9	Data saknas	27	2	0.07

Vi studerar företag under en tidsperiod, mellan år 1998 och 2013. Alla bolag registrerades inte under denna period utan år 1998 kommer det in företag som redan varit registrerade i ett visst antal år. Innan 1998 har det även funnits företag som inte varit framgångsrika och gått i konkurs. Vi märker inte av dessa företag utan de företag som vi börjar observera 1998 är de som redan varit framgångsrika nog att klara sig till början av år 1998. Detta innebär att det blir svårt att tolka åldern på företagen som en variabel som har en påverkan på konkurs.

2.3 BEARBETNING AV DATA

Då det finns en observation för varje år för varje företag från det är registrerat tills det går i konkurs eller tills det senaste bokslutet slås data ihop. Då data slås ihop uppstår problemet att data för olika år kan skilja sig åt. Ett exempel på detta är storleksklasserna som från ett år till ett annat kan vara olika. För enkelhetens skull tas medelvärdet av antal anställda under de åren ett företag har varit aktiv.

En del korrigeringar av data har gjorts i arbetet för att underlätta den statistiska analysen; företag som saknar information som är väsentlig för analysen tas bort och några av storleksklasserna har slagits ihop.

Företag som har försvunnit på andra sätt än genom att ha gått i konkurs tas bort. Detta innebär att företag som fusionerats, aldrig varit aktiva eller har avvecklats under perioden 1998 till 2013 tas bort. Fusionerade bolag tas bort då det inte går att säga något om överlevnaden för dessa, det kan finnas flera möjliga skäl till varför bolaget har fusionerats. Bolag som aldrig varit aktiva är inte intressanta för analysen och tas därför bort. Avvecklade företag tas bort då den inte behöver betyda att företaget går dåligt utan kan även betyda att ägarna för bolaget inte vill bedriva verksamheten av något annat skäl.

De 27 företag som tillhör storleksklass 9 tas bort. Storleksklass 9 innebär att det inte finns någon information om hur många anställda ett företag har. Då vi i denna uppsats vill undersöka just hur ett företags storlek påverkar dess benägenhet att gå i konkurs tas dessa bolag bort.

Då det finns få observationer i de större storleksklasserna sammanfogar vi några av storleksklasserna. Som utgångspunkt för de nya kategorierna tar vi följande definitioner från Ekonomifakta och Bolagsverket:

- Ett enmansföretag har 0 anställda
- Ett mikroföretag har 1 till 9 anställda
- Små företag är de företag som har 10 till 49 anställda
- Ett företag räknas som ett större företag om de har mer än 50 anställda.

Hur många företag som finns i varje storlekskategori och vilken procentandel som har gått i konkurs när vi har bildat de nya kategorierna ses i Tabell 2.3.1.

Tabell 2.3.1: Hur många företag som tillhör de nya storleksklasserna och procentandel som har gått i konkurs per storlekskategori.

Kategori	Antal anställda	Antal företag	Procent som gått i konkurs
Enmansföretag	0	2415	4.80
Mikroföretag	1-9	3556	14.00
Små företag	10-49	366	11.20
Stora företag	≥ 50	63	4.76

3. Metod

I den här uppsatsen kommer analys av hur ett företags storlek och ålder påverkar företags benägenhet att gå i konkurs göras med logistisk regression och överlevnadsanalys. Metoder som kommer användas inom överlevnadsanalys är Kaplan-Meier skattning, Log rank test och Wilcoxon test och Cox proportional hazard modell, om antagandet om proportional hazard är uppfyllt.

Det finns en skillnad på hur vi ser på företag som går i konkurs när vi gör logistisk regression mot hur vi ser på dem när vi gör överlevnadsanalys. Inom logistisk regression ser vi sannolikheten för att ett företag ska gå i konkurs och inom överlevnadsanalysen ser vi det som tiden till en händelse. Låt oss anta att det finns totalt n företag som vi observerar under en bestämd tidsperiod. Av dessa n företag har x gått i konkurs. Inom logistisk regression ser vi då det som att x företag har gått i konkurs och sannolikheten att ett företag ska gå i konkurs är p . De resterande $n - x$ företagen har inte gått i konkurs och sannolikheten att ett företag inte kommer att gå i konkurs är $1 - p$. I överlevnadsanalysen å andra sidan ser vi det som att vi inte har observerat dessa $n - x$ företag gå i konkurs ännu. Dessa företag är högercensurerade.

Innan mer detaljerade beskrivningar av hur modellerna ser ut ska vissa begrepp och samband inom överlevnadsanalys förklaras. Inom överlevnadsanalys studerar vi och observerar om en viss händelse har inträffat inom ett tidsintervall. Då överlevnadsanalys har sitt ursprung från biostatistik är den vanliga händelsen att en patient dör, därav namnet överlevnadsanalys. Dock behöver inte händelsen vara att en individ dör utan i den här uppsatsen är händelsen att ett företag går i konkurs. Tiden till en händelse i överlevnadsanalys har oftast inte en symmetrisk fördelning, ofta har fördelningen en längre svans till höger. Detta innebär att data inte kan approximeras med en normalfördelning. Ett annat problem som kan uppstå i överlevnadsanalys är censurering, att vi inte har fullständig information om hur det kommer att gå för alla individer som vi studerar (Collett, 2003).

Vi börjar med att förklara vad censur och trunkering innebär innan vi går igenom begrepp och samband inom överlevnadsanalys. Vi kommer sedan beskriva metoden bakom Kaplan-Meier skattning, Cox proportional hazard modell och logistisk regression.

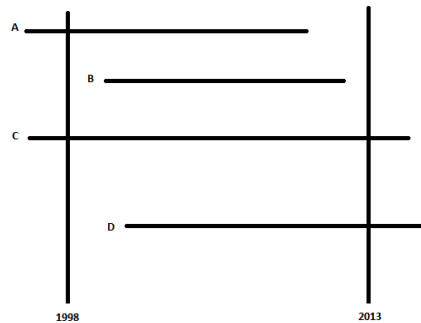
Såvida inte någon annan källa nämns kommer informationen i metoddelen vara hämtad från (Collett, 2003).

3.1 CENSUR

Ett problem som ofta uppstår inom överlevnadsanalys är censurering. Censur innebär att vi inte ännu har observerat en händelse inträffa för alla de individer som vi observerar. Vi observerar om en händelse inträffar under ett tidsintervall men vi vet inte hur det ser ut efter tidsintervallet tagit slut. Ett företag som är aktivt vid slutet av tidsintervallet kan gå i konkurs dagen efter. Detta kallas högercensurering (Collett, 2003). I figur 3.2.1 är företag C och D censurerade, vi har inte observerat slutpunkten för dessa företag.

3.2 TRUNKERING

Trunkering innebär att vi bara observerar de individer som dör inom ett tidsintervall och vi har ingen information om de individer som ligger utanför det intervallet. Vi studerar ett begränsat tidsintervall, 1998-2013. När studien börjar 1998 finns det företag som har startat flera år innan. Vi observerar dessa företag men inte de som startade samtidigt men som gick i konkurs innan 1998. Detta innebär att vi har vänstertrunkering (Klein and Moeschberger, 2003). I Figur 3.2.1 är företag A och C vänstertrunkerade, de startade innan vi började observera dem. Under samma tidperiod som de startade, startade även andra företag som har hunnit gå i konkurs innan 1998.



Figur 3.2.1: Censurerade och trunkerade företag. C och D är censurerade, A och C är trunkerade företag.

3.3 BEGREPP OCH SAMBAND INOM ÖVERLEVNADSANALYS

Det finns tre grundläggande begrepp inom överlevnadsanalys som måste definieras; livstid, överlevnadstid och hazard rate. Genom metoddelen används vi följande termer; de som har utsätts för en viss typ av händelse ”dör” och de som ännu inte blivit utsatta för händelsen ”lever”.

Vi betecknar överlevnadstiden för en individ som t . Där t är ett utfall av T som är en stokastisk variabel som kan anta positiva tal. Sannolikheten att en individ lever mindre än till tidpunkten t är definierad som:

$$F(t) = P(T < t) = \int_0^t f(u)du, \quad t \geq 0 \quad (3.3.1)$$

Där $f(t)$ är täthetsfunktionen för T .

Överlevnadsfunktionen är sannolikheten att en individ överlever längre än till tidpunkten t och är definierad som:

$$S(t) = P(T \geq t) = 1 - F(t) \quad (3.3.2)$$

Hazard funnktionen, $h(t)$, är den momentana risken att en individ dör vid tidpunkten t betingat på att individen har överlevt tiden fram till t . Formellt definieras den momentana risken som:

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t | T \geq t)}{\delta t} \quad (3.3.3)$$

Där täljaren är sannolikheten att en individ dör mellan t och $t + \delta t$ givet att $T > t$ (Collett, 2003).

Från 3.3.3 kan vi härleda samband mellan överlevnadsfunktionen och hazard funktionen. För att visa detta måste vi använda följande samband: $P(B|A) = \frac{P(A \cap B)}{P(A)}$. Som i ord kan uttryckas som sannolikheten att händelsen B inträffar givet att händelsen A har inträffat är lika med sannolikheten att både A och B inträffar dividerat med sannolikheten att händelse A inträffar (Alm and Britton, 2008). Vi utnyttjar detta samband:

$$P(t \leq T \leq t + \delta t | T \geq t) = \frac{P(t \leq T \leq t + \delta t)}{P(T \geq t)} \quad (3.3.4)$$

Vi utnyttjar 3.3.2 och 3.3.1 och kan därmed skriva om 3.3.4 till:

$$\frac{F(t + \delta t) - F(t)}{S(t)} \quad (3.3.5)$$

När vi sätter in 3.3.5 i 3.3.3 får vi följande :

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{F(t + \delta t) - F(t)}{\delta t} \frac{1}{S(t)} = \frac{f(t)}{S(t)} \quad (3.3.6)$$

Den momentana risken att en individ dör vid tidpunkt t är täthetsfunktionen för livstid dividerat med överlevnadsfunktionen.

Det följer även att:

$$h(t) = \frac{-d}{dt} \{\log(S(t))\} \Rightarrow S(t) = e^{-H(t)} \quad (3.3.7)$$

Där $H(t) = \int_0^t h(u)du$ och benämns kumulativ hazard (Collett, 2003).

3.4 KAPLAN-MEIER SKATTNING

Kaplan-Meier skattaren är en icke-parametrisk metod som används för att skatta överlevnadsfunktionen $S(t)$. Vi skapar k olika tidsintervall där $t_1 < t < t_k$. Mellan tidpunkterna t_j och t_{j+1} finns det n_j individer som lever vid början av tidpunkt t_j och som riskerar att dö. De individer som dör mellan tidpunkterna t_j och t_{j+1} betecknas d_j . Sannolikheten att överleva tidsintervallet mellan t_j och t_{j+1} ges utav $\frac{n_j - d_j}{n_j}$. Vi antar att dödsfall mellan individer sker oberoende av varandra vilket gör det möjligt att multiplicera sannolikheter med varandra. Då är Kaplan-Meier skattaren definierad som:

$$\hat{S}(t) = \prod_{j=i}^k \left(1 - \frac{d_j}{n_j}\right). \quad (3.4.1)$$

I ord kan vi beskriva 3.4.1 som: $\frac{n_j - d_j}{n_j}$ kan tolkas som den betingade sannolikheten att överleva fram tills tiden t_{j+1} givet att man har överlevt fram tills t_j . Denna sannolikhet beräknas för var och en av tidpunkterna fram till tidpunkt t . Vi multiplicerar sedan ihop sannolikheterna (Collett, 2003).

Eftersom vi har vänstertrunkerad data uppstår ett problem med att beräkna ålder på företagen. Det finns två olika sätt att hantera vänstertrunkerad data när vi ska skatta Kaplan-Meier. Vi kan antingen räkna ålder på de företag som startade innan 1998 som år från studien har börjat tills de har gått i konkurs eller blivit censurerade, eller så kan vi hantera ålder genom att säga att vi ser de individer som har funnits innan studiens början inte befinner sig i risk för att dö förrän vi börjar observera dem. Det sistnämnda är vårt valda alternativ och det innebär att de företag som startade innan 1998 inte var i risk för att gå i konkurs förrän år 1998. Detta innebär att vi i 3.4.3 har kvar samma antal observerade dödsfall, d_j , men de individer som riskerar att dö vid tidpunkten, n_j , måste anpassas. Vi definierar L_j som ålder när studien börjar, i vårt fall hur gammalt ett företag var år 1998, och T_j som tidpunkten då individen antingen dör eller blir censurerad. Nu är n_j antalet individer som har $L_j < t_j < T_j$ (Klein and Moeschberger, 2003).

Skattningen av överlevnadsfunktionen presenteras grafiskt med Kaplan-Meier kurvor. Där kommer tiden visas på x-axeln och $\hat{S}(t)$ på y-axeln. Kurvan skär y-axeln i 1, i det första tidsintervallet har inga individer hunnit dö, för att sedan stegvis röra sig nedåt för varje nytt tidsintervall. I varje tidsintervall antas den

skattade överlevnadsfunktionen vara konstant och vid varje nytt tidsintervall kommer kurvan hoppa nedåt till det nya värdet på $\hat{S}(t)$.

Inom överlevnadsanalys är det ofta intressant att veta när en viss andel av det totala antalet individer som observerades beräknats ha dött. Vanliga mått är då medianen eller andra percentiler. Vi betecknar percentilen $p(t)$ och $\hat{S}(\hat{t}(p)) = 1 - \frac{p}{100}$. Eftersom funktionen är stegvis går det inte alltid att få exakta värden utan vi tar den lägsta observerade överlevnadstiden sådan att $1 - \frac{p}{100}$ är uppfyllt vilket innebär att vi kan skriva $\hat{S}(\hat{t}(p)) < 1 - \frac{p}{100}$. Runt percentilerna kan vi även räkna ut konfidensintervall.

För att bilda ett konfidensintervall för det sanna värdet av överlevnadsfunktionen vid en given tidpunkt t antas det skattade värdet för överlevnadsfunktionen vara normalfördelat med väntevärde $S(t)$ och varians $\hat{S}(t)^2 \sum_{j=1}^k \frac{d_j}{n_j(n_j - d_j)}$. Se s. 24 Collett (2003) för detaljer. Vi använder oss utav normalfördelningens kvantiler för att bilda ett konfidensintervall. Ett $100(1 - \alpha)\%$ konfidensintervall för överlevnadstiden ges av:

$$\hat{S}(t) \pm z_{\alpha/2} se(\hat{S}(t))$$

(Collett, 2003).

3.4.1 LOG RANK TEST OCH WILCOXON TEST

För att testa om det finns en skillnad mellan överlevnadskurvor för olika grupper kan vi använda Log rank test eller Wilcoxon test. Eftersom Wilcoxon test är en viktad form av Log rank test börjar vi med att beskriva hur Log rank test utförs. Vi beskriver hur testen utförs för att jämföra två grupper. Det fall då fler än två grupper jämförs bygger på samma princip men är lite mer komplicerat och då grundidén för hur testen utförs fås i för jämförelse av två grupper väljer vi att beskriva det. Den intresserade läsaren hänvisas till s. 48 (Collett, 2003) för beskrivning av hur testen utförs när 3 eller fler grupper jämförs.

För att använda Log rank testet för att testa skillnader mellan överlevnadsfunktioner för två grupper delas data först in i två grupper. För grupp 1:

$$U_{LR} = \sum_{j=1}^r (d_{1j} - e_{1j}) \quad (3.4.2)$$

Där d_{1j} är det observerade antalet dödsfall i grupp 1 vid tid j och e_{1j} är det förväntade antalet dödsfall vid tid j för grupp 1.

$$e_{1j} = \frac{n_{1j}d_j}{n_j} \quad (3.4.3)$$

Där n_j är totala antalet som riskerar att dö vid tid j , d_j är det totala antalet som dör vid tidpunkt j för båda grupperna (Allison, 2010).

För att få fram själva teststatistikan utnyttjar vi det faktum att U_{LR} är approximativt normalfördelad om stickprovet är stort och att en normalfördelad variabel dividerat med sin varians är standard normalfördelad. Vilket innebär att kvadraten på denna är χ^2 -fördelad. Detta innebär att Log rank teststatistikan ges av:

$$\frac{U_{LR}^2}{Var(U_{LR})} \sim \chi_1^2$$

där

$$Var(U_{LR}) = \sum_{j=1}^r \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}.$$

Detta innebär att om vi får ett värde på teststatistikan som överstiger värdet vi får från tabell över χ^2 -fördelningen förkastar vi nollhypotesen. För signifikansnivå 5% är χ^2 -statistikan 3.841. Om värdet på Log Rank testet är större än 3.841 förkastar vi nollhypotesen att överlevnadskurvorna ser likadana ut för alla kategorier.

Om överlevnadskurvorna som ges av Kaplan-Meier korsar varandra är antagandet om proportional hazard inte är uppfyllt och då är det inte lämpligt att använda Log rank test. Då bör vi istället använda oss utav Wilcoxon test. Wilcoxon test testar precis som Log rank testet om det är skillnad mellan överlevnadstider för två olika grupper. Skillnaden mot Log rank testet är att Wilcoxon viktat skillnaden mellan de som dör mot det förväntade antalet som skulle dö, med det totala antalet som riskerade att dö vid tillfället. Detta skrivs som:

$$U_W = \sum_{j=1}^r n_j(d_{1j} - e_{1j}) \tag{3.4.4}$$

Variansen för U_W ges av

$$Var(U_W) = \sum_{j=1}^r \frac{n_{1j}n_{2j}d_j(n_j - d_j)}{(n_j - 1)}.$$

Återigen använder vi oss utav att kvadraten av en standard normalfördelad variabel är χ^2 -fördelad och får teststatistikan:

$$\frac{U_W^2}{Var(U_W)} \sim \chi_1^2$$

(Collett, 2003).

3.5 COX-REGRESSION

Cox proportional hazard modellen används för att bestämma hazard, den momentana risken, för en individ samt hazard ratio, hur den momentana risken förhåller sig mellan två grupper, förutsatt att allt annat hålls konstant. I denna typ av regression är den momentana risken, $h_i(t)$ responsvariabel. Den momentana risken för individ i vid tidpunkt t ges av:

$$h_i(t) = h_0(t)e^{\beta_1 x_{i1} + \dots + \beta_k x_{ik}} \quad (3.5.1)$$

där $h_0(t)$ kallas baseline hazard funktionen. När alla k kovariater är 0 blir den momentana risken detta värde. Om vi logariterar 3.5.1 får vi:

$$\log(h_i(t)) = \alpha(t) + \beta_1 x_{i1} + \dots + \beta_k x_{ik} \quad (3.5.2)$$

där $\alpha(t) = \log(h_0(t))$.

Den momentana risken för en individ i Cox-regression är en fix proportion av den momentana risken för vilken som helst annan individ, det vill säga om log av den momentana risken plottas mot tiden kommer hazard funktionerna vara parallella med varandra. Att förklaringsvariablerna inte samspelar med tiden är ett viktigt antagande i modeller med proportional hazard. Detta är en egenskap som måste kontrolleras för i analysen. Detta ses genom att ta kvoten mellan individ i och individ j :

$$\frac{h_i(t)}{h_j(t)} = e^{\beta_1(x_{i1}-x_{j1}) + \dots + \beta_p(x_{ik}-x_{jk})} \quad (3.5.3)$$

Som vi ser ej beror på tiden.

För att skatta β -värdena i Cox-regression används maximum partial likelihood. När vi skattar värdena behöver vi inte skatta baseline hazard-funktionen utan maximum partial likelihoodfunktionen skattas bara för β -värdena. Maximum partial likelihood för k betecknas L_k och definieras som sannolikheten att det var just individ k som dog vid tiden t och inte någon av de andra n individerna. Vi dividerar den momentana risken för individ k vid tidpunkten t med summan av de momentana riskerna för alla de individer som riskerade att dö vid tidpunkt t :

$$L_k = \frac{h_k(t)}{h_k(t) + h_{k+1}(t) + \dots + h_n(t)} \quad (3.5.4)$$

Vi behöver inte ta med baseline-funktionen då den förekommer i både täljare och nämnare i likelihoodfunktionen i 3.5.4 och därför divideras bort. Vid varje ny tidpunkt tas de momentana riskerna som tillhör individer som dött eller blivit censurerade i föregående tidpunkter bort från nämnaren (Allison, 2010).

I SAS görs Cox-regression med hjälp av Proc Phreg.

3.6 LOGISTISK REGRESSION

Logistisk regression används när responsvariabeln, Y , är binär. Vi vill räkna ut sannolikheten för att en viss händelse ska inträffa givet våra förklaringsvariabler: $P(Y = 1|x)$. Då det är svårt anpassa en regression för sannolikheten att en händelse ska inträffa då den bara antar värden mellan 0 och 1 anpassar vi istället kurvan efter logaritmen av oddsen för att en händelse ska inträffa. Oddsen för att en händelse ska inträffa ges utav $\frac{p}{1-p}$. Där p är sannolikheten att en händelse inträffar och $1 - p$ sannolikheten att den inte inträffar. Vi antar att vi har k förklaringsvariabler, x_1, x_2, \dots, x_k . För att kunna utföra regression används $\text{logit}(p)$ som definieras som:

$$\log \frac{p}{1-p} = \alpha + \beta_1 x_1 + \dots + \beta_k x_k. \quad (3.6.1)$$

Där β_1, \dots, β_k är logoddsen för skattningen.

Om vi löser ut p ur 3.6.1 får vi:

$$p = \frac{e^{\alpha + \beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{\alpha + \beta_1 x_1 + \dots + \beta_k x_k}}. \quad (3.6.2)$$

e^β -värdena i 3.6.2 kallas oddskvoter. Oddskvoten säger hur oddsen mellan två grupper skiljer sig åt. Oddskvot för två olika grupper i och j definieras som:

$$\frac{\text{Odds för händelse för grupp } i}{\text{Odds för händelse för grupp } j}$$

Om kvoten är lika med 1 så har båda grupperna lika stora odds för att en händelse ska inträffa. Om oddskvoten är större än 1 så har grupp i större odds för att en händelse ska inträffa och om kvoten är mindre än 1 så har grupp j större odds för att en händelse ska inträffa.

Om ett konfidensintervall för oddskvoten innehåller 1 innebär det att oddsen är densamma för de båda grupperna. För att räkna ut en konfidensintervall för oddskvoten behöver vi först räkna ut konfidensintervallet för logoddsen, $\hat{\beta}$. Vi utnyttjar Wald-statistikan och det faktum att den är χ^2 -fördelad. En χ^2 -fördelad variabel med k frihetsgrader är en summa av kvadrater av k stycken standardiserade normalfördelade variabler. Wald statistikan ges utav:

$$\left(\frac{(\hat{\beta} - \beta_0)}{SE} \right)^2 \leq \lambda_{\alpha/2}^2 \quad (3.6.3)$$

Där $\hat{\beta}$ är skattningen som vi har räknat ut för logoddsen och β_0 är det värde som logoddsen antar under nollhypotesen, här vill vi testa om $\beta_0 = 0$. SE är standardavvikelsen för logoddsen. Från 3.6.3 får vi konfidensintervallet för β :

$$\hat{\beta} \pm \lambda_{\alpha/2} SE = (LL, UL) \quad (3.6.4)$$

Där är LL den undre gränsen och UL den övre gränsen för konfidensintervallet. För att få fram konfidensintervallets gränser för oddskvoten tar vi exponenten av övre och undre gränsen: (e^{LL}, e^{UL}) . Om intervallet inte innehåller 1 är skattningen för oddskvoten signifikant (Agresti, 2002).

3.6.1 TOLKA STATISTIKOR

Det finns tre teststatistikor som brukar användas för att bedöma om förklaringsvariablerna bidrar till att förklara responsvariabeln; Likelihoodkvot, Wald-test och Score-test. Wald test, Score test och likelihoodkvot test ger asymptotiskt samma resultat och då vi har ett relativt stort stickprov väljer vi att endast använda likelihoodkvoten.

I arbetet kommer signifikansnivån 5 % användas om inget annat anges.

LIKELIHOODKVOTTEST

Likelihoodkvottestet mäter bland annat till vilken grad en modell passar in med data.

Vi säger att vi vill testa en nollhypotes:

$$H_0 : \theta = \theta_0.$$

Där θ_0 är skattningen under nollhypotesen. Vi räknar ut en skattning av θ med hjälp utav Maximum-likelihood. Likelihoodfunktionen, $L(\theta; x)$, definieras som $\prod_{i=1}^n f(x_i; \theta)$ i det kontinuerliga fallet och som $\prod_{i=1}^n p(x_i; \theta)$ i det diskreta fallet. Då är θ_{ML} det värde där $L(\theta; x)$ antar sitt största värde (Alm and Britton, 2008).

Likelihoodkvotstatistikan är definierad som:

$$T_{LR}(x) = -2 \log \frac{L(\theta_0; x)}{L(\hat{\theta}_{ML}; x)} \sim \chi_1^2$$

Där $L(\theta_0; x)$ är likelihoodfunktionen med θ_0 och $L(\hat{\theta}_{ML}; x)$ är Likelihoodfunktionen med θ_{ML} .

Vi kan utvidga till det flerdimensionella fallet till en modell med q parametrar. $\hat{\beta}$ är maximum likelihood skattningen som kan anta värden över hela parameterområdet medan β_0 begränsas av värdena under nollhypotesen. Teststatistikan blir då

$$-2 \log \frac{L(\beta_0)}{L(\hat{\beta})} \sim \chi_q^2$$

(Agresti, 2002).

4. Resultat från statistisk analys och diskussion

Vi kommer att börja den statistiska analysen med överlevnadsanalys genom Kaplan-Meier skattning och test av om överlevnadskurvorna är likadana för de olika företagsstorlekarna. Om antagandet om proportional hazard är uppfyllt kommer vi även att göra regression med Cox proportional hazard. Vi kommer att avsluta den statistiska analysen med att göra logistisk regression.

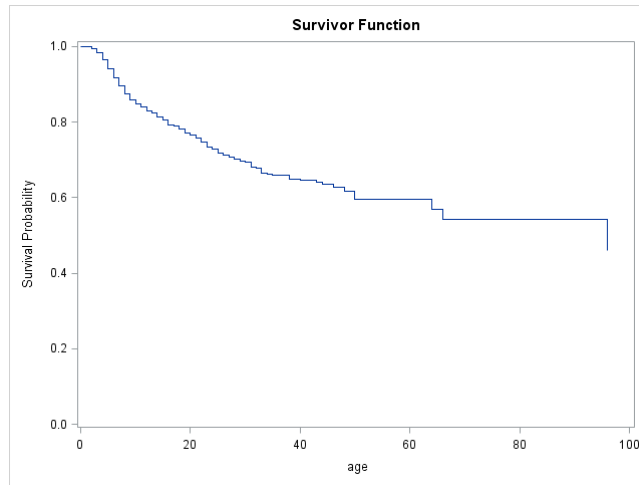
Analysen kommer att göras på de 6400 företag som är kvar efter att vi har modifierat data.

Vi kommer undersöka hur överlevnaden påverkas av företagens storlek men inte hur den påverkas av dess ålder. Då vi har problem med vänstertrunkering kommer vi inte att kunna göra någon analys av hur överlevnaden påverkas av ålder.

4.1 KAPLAN MEIER

Med hjälp utav Kaplan-Meier får vi skattningar av hur lång tid det tar innan ett företag har gått i konkurs, överlevnadstiden för ett företag. För vårt data är n_j de företag som ej gått i konkurs vid början av år j och d_j de företag som gått i konkurs mellan år j och år $j + 1$. I och med vänstertrunkeringen ser vi att företag inte befinner sig i risk för att gå i konkurs innan 1998 även om de har startat flera år innan.

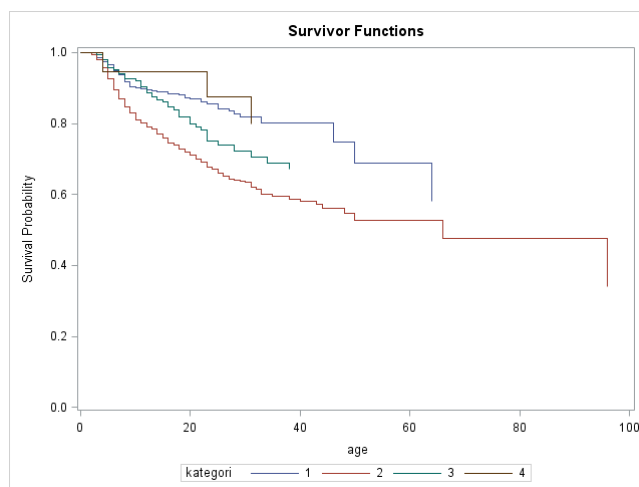
Vi börjar med att undersöka hur överlevnaden ser ut för alla företag. Vi får då skattningen av överlevnadskurvan med hjälp utav Kaplan-Meier enligt Figur 4.1.1.



Figur 4.1.1: Kaplan-Meier kurva för alla bolag

För att undersöka hur överlevnaden ser ut använder vi oss utav percentiler. 25% percentilen är den tidpunkt då 25% av alla företag beräknats ha gått i konkurs. Från SAS får vi att 25% percentilen är 27 år med skattad sannolikhet 0.7 och konfidensintervall (0.69,0.73). Detta innebär att drygt 25% av alla företag beräknats ha gått i konkurs vid 27 års ålder. Medianvärdet, den tidpunkt då 50% av företagen beräknats ha gått i konkurs, är vid 96 år. Då är sannolikheten för att överleva till den åldern eller längre 0.46 med konfidensintervall (0.32, 0.6683).

Vi vill nu testa om överlevnadskurvorna ser likadana ut för de olika företagsstorlekarna. Vi delar nu upp överlevnadskurvorna per storlekskategori och får Kaplan-Meier-kurvor enligt Figur 4.1.2.



Figur 4.1.2: Kaplan-Meier uppdelat per storlekskategori

Vi studerar kurvorna uppdelade efter storlek. I Figur 4.1.2 svarar den blå kurvan mot enmansföretag, den ljusbruna mot mikroföretag, den gröna mot små företag och den mörkbruna mot stora företag. Vi ser sannolikheten för överlevnad är minst för mikroföretag och högst för stora företag. Vi noterar att överlevnadskurvorna för enmansföretag och små företag korsar varandra. När kurvan gör lite större hopp nedåt som den gör för mikroföretagen när de är äldre än 90 år, betyder det att det finns få företag som är äldre än 90 år och därmed blir hoppet större. Hoppen blir större eftersom det finns färre företag kvar, för varje nytt företag som går i konkurs då påverkar det sannolikheten för överlevnad mer än då det finns fler företag.

Vi undersöker hur lång tid det tar tills 25% av de företagen i de olika storlekskategorierna har gått i konkurs, se Tabell 4.1.1. Från kurvorna i Kaplan-Meier kunde vi se att mikroföretag är den typ av företag som det går sämst för. Vi ser även att 25% percentilen är vid 16 år för dessa företag. Detta innebär att det tar 16 år innan 25% av företagen i denna storleksklass beräknas ha gått i konkurs. För enmansföretagen tar det 46 år innan 25% av företagen uppskattas ha gått i konkurs. 25% percentilen för små företag är 25 år vilket innebär att det tar ca 25 år innan 25% av bolagen i den kategorin beräknas ha gått i konkurs. För kategorin stora företag finns det ingen 25% percentil.

Tabell 4.1.1: 25 % percentilen för de olika företagsstorlekarna

Företagsstorlek	Ålder	$\hat{S}(t)$	95% konfidensintervall
Enmansföretag	46	0.747	(0.64, 0.87)
Mikroföretag	16	0.744	(0.72, 0.77)
Små företag	25	0.739	(0.67, 0.82)
Stora företag	-	-	-

Då vi inte kunde jämföra 25% percentilen för alla de olika företagsstorlekarna väljer vi istället att jämföra när sannolikheten att ett företag inte gått i konkurs vid en viss tidpunkt eller längre är ungefär 0.9. Det vill säga den tidpunkt då ungefär 10% av alla företag beräknats ha gått i konkurs, Se Tabell 4.1.2.

Tabell 4.1.2: 10% percentilen för de olika företagsstorlekarna

Företagsstorlek	Ålder	$\hat{S}(t)$	95% konfidensintervall
Enmansföretag	11	0.897	(0.88, 0.92)
Mikroföretag	6	0.895	(0.88, 0.91)
Små företag	12	0.885	(0.84, 0.94)
Stora företag	23	0.876	(0.73, 1)

Vi ser att det tar ca 11 år innan 10% av enmansföretagen beräknas ha gått i konkurs, 6 år för mikroföretagen och 12 år för de små företagen. Nu får vi även en skattning för de stora företagen. Det beräknas ta ca 23 år innan 10% av de företagen beräknats ha gått i konkurs.

Vi kan dra slutsatsen från analys med Kaplan-Meier skattning att företagens storlek har en inverkan på sannolikheten att gå i konkurs. Vi kan också dra slutsatsen att den storlek på företagen som har störst benägenhet att gå i konkurs är mikroföretagen.

4.1.1 LOG RANK TEST OCH WILCOXON TEST

För att testa om överlevnadskurvorna för de olika företagsstorlekarna inte skiljer sig åt kan vi använda oss utav någon av teststatistikorna Log rank eller Wilcoxon. Då överlevnadskurvorna i Kaplan-Meier korsar varandra blir teststatistikan för Log rank testet ej giltigt. Då är det mer lämpligt att använda Wilcoxon test.

Då vi använder en alternativ metod för att räkna ut överlevnaden för företag då det förekommer vänstertrunkering, ges ett score test som är asymptotiskt lika med log rank testet. Detta ger värdet 41.3273 med p-värde < 0.0001 . Testet pekar på att det finns en signifikant skillnad mellan överlevnaden för de olika företagsstorlekarna men då överlevnadskurvorna korsar varandra är det osäkert hur giltigt detta test är.

4.2 COX PROPORTIONAL HAZARD MODELL

Om vi vill fortsätta vår överlevnadsanalys med Cox-regression måste proportional hazard gälla, att hazardkvoten är oberoende av tiden. Detta innebär att överlevnadskurvor inte kan korsa varandra för då skulle det innebära att överlevnaden för de olika grupperna inte är proportionella mot varandra utan förändras med tiden. Vi ser att överlevnadskurvorna för mikroföretagen och den kurvan som representerar kategorin små företag korsar varandra och därför bryter mot antagandet om proportional hazard. Vi kan därmed inte göra regression genom Cox proportional hazard.

4.3 LOGISTISK REGRESSION

Inom logistisk regression betraktar vi konkurs som ett utfall, antingen går ett företag i konkurs eller inte. Vi betecknar denna variabel Y .

$$Y = \begin{cases} 1 & \text{om företaget har gått i konkurs,} \\ 0 & \text{om ett företag inte har gått i konkurs.} \end{cases}$$

Vi är intresserade av att veta hur sambandet mellan sannolikheten att ett företag går i konkurs och företagsstorlekarna ser ut. Vi kan inte använda logistisk regression för att avgöra om ålder har en påverkan på benägenheten att gå i konkurs då åldern är vänstertrunkerad. Vi får ekvation enligt 4.3.1

$$P(Y = 1|x) = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3}} \quad (4.3.1)$$

Där:

$$X_1 = \begin{cases} 1 & \text{om ett företag tillhör kategorin mikroföretag,} \\ 0 & \text{annars.} \end{cases}$$

$$X_2 = \begin{cases} 1 & \text{om ett företag tillhör kategorin små företag,} \\ 0 & \text{annars.} \end{cases}$$

$$X_3 = \begin{cases} 1 & \text{om företaget tillhör kategorin stora företag,} \\ 0 & \text{annars.} \end{cases}$$

Referenskategori är enmansföretag.

Vi skriver nu om 4.3.1 så att vi har logodds för att ett företag ska gå i konkurs som responsvariabel och som förklaringsvariabler har vi indikatorvariablerna för företagsstorlekarna:

$$\log(\text{Odds för Konkurs}) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.$$

Odds för att ett företag ska gå i konkurs ges av:

$$\frac{P(Y = 1|x)}{P(Y = 0|x)}$$

Den logistiska regressionen ger resultatet i Tabell 4.3.1

Tabell 4.3.1: Resultat från logistisk regression

Variablel	Skattning	p-värde	Oddsquot	95% Konfidensintervall
Intercept	-2.9864	<0.0001	-	-
Enmansföretag	0	-	1	-
Mikroföretag	1.1715	<0.0001	3.227	(2.618, 3.978)
Små företag	0.9162	<0.0001	2.5	(1.719, 3.635)
Stora företag	-0.00908	0.9879	0.991	(0.306,3.207)

Från Tabell 4.3.1 kan vi avläsa skattningar för logodds (under "Skattning") och oddskvoter. Logoddsen och oddskvoterna för de olika företagsstorlekarna är i jämförelse med kategorin enmansföretag.

Oddskvoterna säger hur oddsen för de olika storlekskategorierna förhåller sig till storlekskategori enmansföretag. Det finns två sätt att kontrollera om en skattning är signifikant. Antingen kontrollera att konfidensintervallet inte innehåller 1 eller att se om p-värdet för skattningen av logoddset är mindre än 0.05. Den skattade oddskvoten för ett företag som är ett mikroföretag, 3.227, säger att oddsen att gå i konkurs är 3.227 gånger högre för ett mikroföretag än oddsen för ett företag som är ett enmansföretag att gå i konkurs. Då p-värdet för skattningen av logoddset är signifikant är även skattningen för oddskvoten signifikant. För små företag skattas oddskvoten till 2.5. Vilket innebär att odds- sen att gå i konkurs är 2.5 större om ett företag tillhör kategorin små företag än om ett företag tillhör kategorin enmansföretag. Även för denna skattning är skattningen för logoddset signifikant. 0.991 som är skattningen för oddskvoten för företag som tillhör kategorin stora företag säger att oddsen att gå i konkurs är 8.9% lägre för ett företag som tillhör den kategorin än ett företag som är ett enmansföretag. Dock framgår det att denna skillnad inte är signifikant, p-värdet för skattningen av logoddset är större än 0.05.

För att testa om benägenheten att gå i konkurs påverkas av hur stort företaget är använder vi likelihoodkvottest. Resultatet av detta test är 147.4156 med p-värde <0.0001. Detta innebär att testet är signifikant och därmed kan det konstateras att ett företags storlek har en effekt på benägenheten för att ett företag går i konkurs.

Från den logistiska regressionen kan vi dra slutsatsen att vilken storlek ett företag tillhör har en påverkan på företagets benägenhet att gå i konkurs. Oddsquo-

terna för storleken för ett företag anger att oddsen att ett företag går i konkurs är störst för mikroföretag i jämförelse med enmansföretag och verkar minska med ökad storlek. Det går dock inte att uttala sig om den största företagsstorleken, stora företag, då resultatet för den skattningen inte var signifikant.

4.4 ENMANSFÖRETAGEN

Från analysen med Kaplan-Meier skattning såg vi att antagandet om proportional hazard bröts eftersom överlevnadskurvan för enmansföretagen korsade överlevnadskurvan för små företag. Vi kunde även se i analysen att den här typen av företag skiljde sig åt genom att den tid det tar innan 25% av alla företag hade gått i konkurs var högre än för mikroföretagen och kategorin små företag. Även i den logistiska regressionen såg vi att den här typen av företag betedde sig annorlunda mot de övriga kategorierna då alla andra kategorier av företagsstorlekar hade större odds att gå i konkurs än denna. Frågan är om denna typ av företag skiljer sig så markant att den inte borde jämföras med de andra företagen.

Storleksklassen 0 anställda innebär att företagets ägare är den enda som jobbar i företaget. Det kan hända att denna kategori skiljer sig mot de övriga kategorierna av företagsstorlekar. Om ett företag har anställda kommer det med ökade kostnader. Dels för att ett företag måste betala ut löner och sociala avgifter varje månad men det kan även dyka upp oväntade extra utgifter då de kan behöva ta in en vikarie eller rekrytera ny personal vid föräldraledighet och sjukskrivning. I ett företag som inte har några anställda kan ägaren ta ut vinst vid de tidpunkter som det går bra för företaget och låta bli att ta ut pengar de månader då det går sämre. Det finns även skillnader i vad som händer i ett större bolag om ägaren går i pension eller slutar. Det är mer troligt att någon annan tar över bolaget när ägaren går i pension eller vill avsluta sitt engagemang i företaget av någon annan anledning i ett större bolag än i ett där ägaren är den enda anställda. Detta gör att det kanske är svårare att tolka Kaplan-Meier kurvan efter till exempel 50 år.

5. Slutsats

Syftet med den här uppsatsen var att ta reda på om företagens ålder och storlek påverkar deras överlevnad. Vi kunde konstatera att ett företags storlek påverkar benägenheten att gå i konkurs men upptäckte även ett problem i analysen då vi har en vänstertrunkering för företagets ålder och vi kunde inte med metoder som användes i detta arbete undersöka hur benägenheten att gå i konkurs påverkas av ett företags ålder.

Från den logistiska regressionen såg vi att företagens storlek har en signifikant effekt på benägenheten att gå i konkurs. Vi såg att mikroföretag och små företag hade större odds att gå i konkurs än enmansföretag. Att företagens storlek hade en effekt på överlevnaden bekräftades även genom överlevnadsanalysen där Wilcoxon test visade på att överlevnaden för de olika företagsstorlekarna skiljer sig åt. Kaplan-Meier kurvorna visade även de att de olika överlevnadskurvorna för företagsstorlekarna skiljde sig åt och att den grupp som har störst benägenhet att gå i konkurs är mikroföretagen. Där visade det sig till skillnad mot den logistiska regressionen att de företag som tillhör kategorin stora företag var den kategori som hade störst sannolikhet att överleva.

Vi kunde observera att kategorin enmansföretagen skiljer sig från de övriga storlekarna på företagen. Denna kategori har en högre överlevnad än mikroföretagen och de företagen som tillhör kategorin små företag. Överlevnadskurvan för denna kategori skar överlevnadskurvan för små företag två gånger i analysen med Kaplan-Meier kurvor. Om denna grupp skulle ha analyserats för sig själv och inte i jämförelse med de andra företagen, då andra förhållanden och regler kan gälla för den här typen av företag, är en fråga som lämnas öppen i detta arbete.

Om vi bortser från enmansföretagen skulle vi kunna bekräfta att teorin *legitimacy of smallness* även gäller för svenska företag. Vi skulle behöva andra typer av analyser för att undersöka vår andra frågeställning, om teorin *legitimacy of newness* är giltig för svenska företag.

Det finns flera möjliga felkällor i denna analys, de främsta är approximationer av de olika variablerna och bias när det gäller företagens ålder. När vi slog ihop data så att vi fick en observation per företag istället för en observation per företag och år var vi tvungna att approximera de variabler som kunde förändras mellan de olika åren. Till exempel approximerade vi storleksklassen med medelvärdet. Denna approximation kan säkert i vissa fall ha gett en missvisande bild. Den andra felkällan, att data är snedvridet är ett allvarigare fel. Vi studerar företag under en tidsperiod, mellan år 1998 och 2013. I början av 1998 kommer företag in i studien som redan har varit registrerade och aktiva. Innan 1998 har

det även funnits företag som inte varit framgångsrika och gått i konkurs. Vi märker inte av dessa företag utan de företag som vi börjar observera 1998 är de som redan varit framgångsrika nog att klara sig till början av år 1998. Detta innebär att det blir svårt att tolka åldern på företagen som en variabel som har en påverkan på konkurs. För att komma runt detta skulle åtgärder behöva vidtas, till exempel skulle vi kunna välja att bara studera åldern på de företag som har blivit nystartade under tidsperioden eller använda någon annan typ av modell.

Enligt rapporten *Smått om små företag* från Svenskt näringsliv finns det flera faktorer som påverkar hur de små företagen klarar sig. Enligt rapporten kan möjliga hinder för små företag vara; krångliga regelverk, marknadens efterfrågan, höga personalkostnader samt höga avgifter och skatter. Troligen innebär de flesta av dessa hinder mest ett hinder för att ett företag ska expandera och anställa fler personer än en avgörande faktor för om ett företag kommer att gå i konkurs men det är möjligt att dessa anledningar är bidragande orsaker till konkurs som inte fångas upp i den här uppsatsen.

Förutom det som nämns i studien i stycket ovan kan det kan finnas fler anledningar till varför företag går i konkurs som inte fångas upp i den här uppsatsen. Små företag kanske inte samma resurser som stora företag att lägga på marknadsföring eller för att locka och rekrytera personal. För att fånga upp andra möjliga orsaker skulle man behöva enkäter eller andra sätt för att undersöka de egentliga orsakerna till varför ett företag gick i konkurs. Det är även möjligt att konjunkturen påverkar överlevnaden för företag. Det kan tänkas att små företag är mer känsliga för om det är lågkonjunktur eller högkonjunktur. För att fånga upp om detta har en påverkan på företagens benägenhet att gå i konkurs skulle en studie med konkurser per år och även uppdelat per företagsstorlek vara en intressant studie. En annan möjlig faktor som kan påverka benägenheten att gå i konkurs är vilken bransch företagen tillhör. Enligt Svenskt näringsliv studie är ungefär hälften av alla små företag tjänsteföretag. Denna bransch är kanske mer känslig för svängningar i konjunkturen och skiftande regelverk, till exempel har den fått ett uppsving med RUT-avdraget. Tänkbara utvidgningar av denna studie vore alltså att undersöka hur konkurser skiljer sig åt mellan år, speciellt kontrollera mot de år som det har varit högkonjunktur och lågkonjunktur samt att undersöka hur benägenheten att gå i konkurs ser ut för de olika branscherna.

A. Appendix

Följande information är hämtat från dokumentation till databasen Serrano.

A.1 DEFINITION AV AKTIVT FÖRETAG

Ett företag defineras som aktiv om:
Nettomsättning > 10 000 eller
Övriga rörelseintäkter > 10 000 eller
Finansiella intäkter > 10 000 eller
Finansiella utgifter < -10 000 eller
Utdelningsbeloppet > 10 000 eller
Totala tillgångar > 500 000.

A.2 JURIDISK FORM

22 = 'Partrederier'
23 = 'Värdepappersfonder'
31 = 'Handelsbolag/Kommanditbolag'
41 = 'Bankaktiebolag'
42 = 'Försäkringsaktiebolag'
49 = 'Aktiebolag'
51 = 'Ekonomiska föreningar'
81 = 'Statliga enheter'
82 = 'Primärkommuner'
83 = 'Kommunalförbund'
84 = 'Landsting'
85 = 'Allmänna försäkringskassor'
87 = 'Offentliga korporationer och anstalter'
88 = 'Hypoteksföreningar'
89 = 'Statliga regionala myndigheter'
92 = 'Ömsesidiga försäkringsbolag'
93 = 'Sparbanker'
96 = 'Utländska juridiska personer'

Litteraturförteckning

- Bolagsverket. URL <http://www.bolagsverket.se/ff/foretagsformer>. Besökt 2015-03-31.
- Ekonomifakta. URL <http://www.ekonomifakta.se/sv/Fakta/Foretagande/Naringslivet/Naringslivets-struktur/>. Besökt 2015-04-15.
- Sveriges domstolar. URL <http://www.domstol.se/Skuld--konkurs/Om-konkurs/>. Besökt 2015-04-29.
- Alan Agresti. *Categorical Data Analysis*. John Wiley Sons, Inc., Hoboken, New Jersey, second edition, 2002.
- Paul D. Allison. *Survival Analysis Using SAS*. SAS Institute Inc., Cary, North Carolina, second edition, 2010.
- Sven Erik Alm and Tom Britton. *Stokastik*. Liber AB, Stockholm, first edition, 2008.
- S.R. Clegg, C. Hardy, and W.R. Nord. *Handbook of Organization Studies*. SAGE Publications, first edition, 2003.
- David Collett. *Modelling Survival Data in Medical Research*. Chapman & Hall/CRC, Boca Raton, Florida, second edition, 2003.
- John P Klein and Melvin L. Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data*. Springer, New York, second edition, 2003.
- Svenskt näringsliv. Smått om små företag. URL http://www.svensktnaringsliv.se/material/rapporter/smatt-om-sma-foretag_571248.html. Besökt 2015-05-01.