Stockholms universitet

Statistiska institutionen

Tre av tio har avgått

En överlevnadsstudie av tiden till avgång för kommunfullmäktigeledamöter i Stockholms län.

Three in ten has resigned

A survival analysis of time to resignation for city council politicians in the region of Stockholm.

Av: Maria Frykhult och Fredrik Engleryd

15-högskolepoäng, Statistik III, HT 2011
Handledare: Göran Rundqvist

ABSTRACT

This study is a survival analysis which aims to examine the time to resignation during mandate terms, not resignation due to election, for City council politicians in the Stockholm region, based on several covariates like; age, sex, party, party belonging to right or left wing, position in City council, middle income group, experience from previous work in City council. The twenty-six municipalities in the region of Stockholm were divided into three groups according to the middle-income. Nine of the municipalities are included in the study, three were chosen from each income group. The study period extends from the election 2002 until March 2010, and a total of 740 City council politicians have been observed during this period. The material for the study has been provided for by the municipalities and has been supplemented with information from different internet databases when needed.

Several of the common AFT models and the Cox proportional hazard model are being evaluated, with the purpose to find the probability distribution that provides a good description of the data, and to find a survival model including covariates that affect the time to resignation. Three models proved to fit the data reasonably well. Two of the AFT models, the Weibull and the exponential distribution and an extended Cox model. Due to a suspicious violation of the proportionality assumption in the Cox model for one of the covariates, an extended Cox model with a time dependent covariate was preferable. The risk of resignation during mandate terms, or in between elections, seems to be almost constant over the period. The covariates that have an influence of the time to resignation in this study are age, experience from previous work in the City council, whether the politician act in opposition or not and the middle income group of the municipality that the politician represent.

INNEHÅLL

1. INLEDNING

Det som skänker legitimitet till den politiska processen, beslutsfattande i våra demokratiska organ såsom fullmäktigeförsamlingar och parlament, är folkets möjlighet att kontrollera de förtroendevalde vid demokratiska val. (Gilljam & Hermansson; 296)
Har därför våra förtroendevalda ett ansvar att slutföra de uppdrag som folket ger dem, och gör de det?

Uppsatsen tar sin utgångspunkt i problemet att förtroendevalda väljer att hoppa av sina uppdrag innan mandatperioden är slut. Avhopp bland förtroendevalda i kommunfullmäktige är intresseområdet för undersökning, med Stockholms län som målpopulation. Uppsatsen syftar till att undersöka vad som påverkar avhopp innan mandatperioden är slut bland kommunfullmäktigeledamöter i Stockholms läns kommuner.

Under mandatperioden med start 2006 avgick till exempel 16,4 procent av alla kommunfullmäktigeledamöter i Sverige. Bland kommunerna inom Stockholms län är antalet avgångna ledamöter 18,5 procent, vilket är tre i topp vad gäller andel avgångar från kommunfullmäktige bland Sveriges län. (valmyndigheterna.se) Stockholms län är således ett intressant område att studera. Undersökningen kommer göras som en fallstudie av Stockholms län, där ett urval av Stockholms läns kommuner väljs ut och dess ledamöter undersöks. Teoretisk bakgrund som ligger till grund för de olika variabler som undersöks redovisas i materialavsnittet 3.1 val av variabler.

Den statistiska metoden överlevnadsanalys kommer användas för att undersöka tiden till avhopp från kommunfullmäktige för kommunfullmäktigeledamöterna. Överlevnadsanalys mäter tiden för observationerna från undersökningens start till dess att en viss händelse inträffar, eller till dess att mätperioden är slut. Överlevnadsanalys kan t.ex. mäta tiden från då individer får en diagnos till dess att de dör, eller alternativt till dess att de blir vad som kallas censurerade, vilket betyder att de lever då undersökningen är slut. En observation betraktas också som censurerad om den försvinner från undersökningen under mätperioden, och det inte går att följa upp om individen upplevt händelsen eller inte.

Det är viktigt att beakta att begreppet överlevnad inte ska tolkas i sin bokstavliga betydelse, utan används inom metoden som ett allmänt begrepp för tiden mellan mättidens start och till dess att händelsen inträffar. Det behöver inte vara individer som är observationerna, tid till dess att glödlampor slocknar kan till exempel vara målet för en överlevnadsanalys.

I analysen med överlevnadsdata över kommunfullmäktigeledamöter är överlevnadstiden det antal veckor som ledamoten sitter i kommunfullmäktige innan han eller hon väljer att avgå vilket alltså är händelsen, eller försvinner från kommunfullmäktige i samband med val, vilket redovisas som att observationen blir censurerad. De ledamöter som har suttit över mätperiodens slut ses också som censurerade. Mätperioden tar sin början vid 2002-års val och sträcker sig drygt två mandatperioder framåt.

I avsnitt två specificeras uppsatsens syfte och frågeställning närmare. Därefter följer, i avsnitt tre, en redovisning av materialet och val av variabler. I avsnitt fyra beskrivs metoden; överlevnadsanalys. Fokus ligger på de parametriska modellerna och Cox proportionella hazardmodell. Icke parametriska metoder som Kaplain-Meier och Life-table kommer att beskrivas övergripande. Resultaten av överlevnadsanalysen visas i avsnitt fem. Slutligen kommer en analys för de statistiska resultaten i avsnitt sex, och en närmare diskussion av resultaten i anknytning till teori, samt förslag till utveckling av undersökningen i avsnitt sju.

2. SYFTE FRÅGESTÄLLNING OCH TILLVÄGAGÅNGSSÄTT
2.1 Syfte

Uppsatsens syfte är att undersöka överlevnadstiden för ledamöterna i kommunfullmäktige i Stockholms läns kommuner, mellan valet år 2002 och fram till april 2011, med utgångspunkt på vad som kan tänkas påverka överlevnadstiden. Studien syftar till att anpassa en modell med bra förklaringsvariabler, och undersöka hur den modellen bäst kan presenteras. Med studien avses att undersöka hur överlevnadsfunktionen ser ut för ledamöter som avgår i förtid, det vill säga innan mandatperioden är slut. Det avses alltså inte att undersöka överlevnadstiden för ledamöter som försvinner från kommunfullmäktig i samband med val.

2.2 Frågeställning

Hur påverkas tid till avgång från kommunfullmäktige för kommunfullmäktigledamöter i Stockholms läns kommuner av; ledamotens ålder, kön, partitillhörighet och tidigare erfarenhet av att sitta i kommunfullmäktige, samt kommunens resurser (mätt i kommuninvånarnas medelinkomst).

2.3 Avgränsningar

Samtliga avhopp i förtid är att betrakta som händelser eftersom vi inte vet den egentliga orsaken till avhopp. Både frivilliga och ofrivilliga orsaker till avhopp som till exempel dödsfall och sjukdom kommer alltså att räknas som avhopp. Censurering kan därför bara inträffa i samband med val.

2.4 Tillvägagångssätt

Initialt kommer datamaterialet att analyseras med hjälp av Kaplan-Meier och Life-table för att få en övergripande bild av hur överlevnadsdatat är fördelat samt indikation på vilka variabler som kan tänkas påverka.

Analysens huvudsakliga process består sedan i att med hjälp av Cox-proportionella hazard modell och parametriska modeller hitta en modell med signifikanta förklaringsvariabler.

Cox proportionella hazard modell kommer, på grund av sin lätthanterlighet, att skattas först. Antagandet som den modellen vilar på är proportionalitet mellan variablernas nivåer, vilket kommer att testas innan Cox-modellen skattas.

Parametriska modeller kan ge mer precisa skattningar, samt grafer för hazarden. För att kunna använda skattningar från parametriska modeller krävs det att överlevnadsdatat följer en känd sannolikhetsfördelning. Inom detta testas alltså om överlevnadsdatat följer någon fördelning, och därefter skattas modellen med hjälp av de parametriska modellerna.

Analysen avslutas med en jämförelse mellan de olika modellerna, parametriska skattningarna såväl som Cox.

3 MATERIAL

Målpopulationen för undersökningen består av kommunfullmäktigeledamöter i samtliga kommuner i Stockholms län, under tidsperioden 2002-10-31 till 2011-03-31. Urvalet består av nio kommuner. Stockolms tjugosex kommuner delades upp i 3 stratum beroende på medelinkomsten i kommunen (För stratumindelning samt medelinkomster i kommunerna se appendix B). I det resursstarkaste stratumet valdes Danderyd Lidingö och Nacka. I det resurssvagaste stratumet valdes Botkyrka Södertälje och Haninge. Från mittenstratumet slutligen valdes Järfälla Solna och Sigtuna. Eftersom vissa kommuner i det ursprungliga urvalet inte kunde tillhandahålla tillräcklig data för att genomföra undersökningen var det nödvändigt att göra en justering av urvalet. Vilket ledde till ett slags tillgänglighetsurval, allstå de kommuner som hade information att tillgå valdes ut.

Tidsperioden för undersökningen har begränsats till att innefatta valet 2002 och fram till april 2011. Ledamöterna som ingår i studien har begränsats till dem som valt in i kommunfullmäktige i 2002 och 2006 års val. Ledamöter som kommit in som ersättare för avhopparna under mandatperioderna har inte tagit med i datamaterialet. Den sammanlagda tiden för undersökningen är 438 veckor.

Datamaterialet har sammanställts med information från flera olika källor. Information om vilka som suttit i kommunfullmäktige, deras ålder, kön, partitillhörighet, samt när personen avgått har efterfrågats från ansvariga för kommunernas register över förtroendevalda. Dessa register har i några fall funnits tillgängliga på kommunens hemsida, och har i så fall använts. Information om ledamöternas ålder har ibland inte gått att få tag på via kommunerna och har i så fall inhämtats från bland annat valsedlar på valmyndighetens hemsida. Åldersvariabeln innehåller en del missing values, sexton stycken, varav samtliga för censurerade observationer.

Datamaterialet består av 741 observationer. De ledamöter som avgått vid val är censurerade. De ledamöter som sitter kvar i kommunfullmäktige vid studieperiodens slut är också censurerade. Det är cirka 71 procent (525 stycken) censurerade observationer i datamaterialet. Cirka 29 procent har upplevt händelsen, avhopp från kommunfullmäktige i förtid, vilket är 216 ledamöter.

Av ledamöterna i urvalet är 43 procent kvinnor och 57 procent män, 55 procent har ingen tidigare erfarenhet av tidigare kommunfullmäktige arbete emedan 45 procent har det, och 46 procent av ledamöterna satt i opposition, vid tidpunkten då de avgick alternativt blev censurerade.

59 procent av ledamöterna tillhör högerblocket och 41 procent vänsterblocket. Moderaterna och socialdemokraterna är de största partierna, 31 procent av ledamöterna är moderater och 29 procent är socialdemokrater. 12 procent av ledamöterna är folkpartister. Andel ledamöter i datamaterialet som tillhör någon av de övriga partierna, eller partigruppen övriga, är mellan 3 och 7 procent.

Fördelningen av ledamöter mellan de tre kommunstratumen är relativt jämn, kommungrupp ett består av 31 procent av ledamöterna, kommungrupp två består av 34 procent och kommungrupp tre består följaktligen av 34 procent av ledamöterna i datamaterialet.

3.1 Val av variabler.

Variablerna har valts och skapats, med olika grad av teoretisk utgångspunkt, utifrån idéer om vad som påverkar tiden som ledamöterna sitter i kommunfullmäktige.

Enligt tidigare forskning om avhopp från riksdagen (Albäck m.fl. ”Exit riksdagen”) är ålder en stark faktor. Det är sannolikare för avhopp vid högre ålder. Kön, tidigare erfarenhet från riksdagsarbete och parlamentarisk position är andra variabler som påverkar ledamotens benägenhet att hoppa av i förtid. Orsaker för avhopp för riksdagsledamöter har fått fungera som en vägledning inför val av variabler till undersökningen om kommunfullmäktigeledamöters orsaker till avhopp. Tidigare forskning om förhållanden i riksdagen fungerar alltså som en teoretisk utgångspunkt för uppsatsen.

Variablerna i undersökningen är; ålder, kön, tidigare erfarenhet av arbete i kommunfullmäktige, partitillhörighet, höger eller vänsterblock, status i kommunfullmäktige (om ledamoten sitter i opposition eller inte) och kommungrupp.

Ålder är mätt i ledamöternas ålder vid inträde i studien. Variabeln tidigare erfarenhet mäter om ledamöterna har tidigare erfarenhet av att sitta i kommunfullmäktige. Tidigare erfarenhet har bara räknats en mandatperiod bakåt från inträdet i undersökningen. Ledamöter kan alltså ha suttit i fullmäktige perioder långt bakåt i tiden utan att anses ha tidigare erfarenhet i den här studien. Tidigare erfarenhet av uppdrag i kommunens andra styrelser eller nämnder räknas inte med som tidigare erfarenhet. Avgränsningen har varit nödvändig av tidsskäl. Variabeln opposition mäter helt enkelt om ledamoten satt i opposition eller inte vid avhopp (eller censurering).

Vad gäller variabeln partitillhörighet har Sverigedemokrater och lokala partier sammanförts till gruppen övriga partier. Partierna är även indelade i block. Högerblocket består av m c kd fp och gruppen övriga. Vänsterblocket består av v mp och s.

Tiden till avhopp och censurering är mätt i veckor.

4. METOD
4.1 Introduktion till överlevnadsanalys.

Överlevnadsanalys har fördelar gentemot vanlig regressionsanalys. En fördel är att överlevnadsanalys kan hantera censurerade observationer. En observation räknas som censurerad om den har försvunnit från undersökningen utan att händelsen har inträffat, och det inte finns någon möjlighet att ta reda på om händelsen kommer att inträffa för observationen eller inte. Tidpunkten för censurering är dock känd. Det finns flera typer av censurering. Så kallad vänstercensurering innebär att observationen försvinner från undersökningen innan startpunkten för mätningen tar sin början. Högercensurering, som är vanligare, innebär att observationen försvinner från undersökningen innan slutpunkten för studien. Om det endast är känt att observationen blivit censurerad någon gång mellan två tidpunkter har man intervallcensurering. I datamaterialet för undersökningen förekommer endast högercensurering.

Ett centralt antagande som överlevnadsanalys bygger på är att censureringen är icke-informativ. (Collett 2003) Med ickeinformativ censurering menas att censurering sker slumpmässigt och oberoende av några av de mekanismer som påverkar händelserna.

En annan fördel är att överlevnadsanalys tar hänsyn till att överlevnadsdata vanligtvis inte är symmetriskt fördelat, utan har en positiv skevhet. Överlevnadsdata är så att säga fördelat med en längre högersvans.

De centrala byggstenarna i överlevnadsanalys är överlevnadsfunktioner och hazard funktioner. Överlevnadsfunktioner estimerar sannolikheten för en individ att överleva längre än tidpunkt *t*. Funktionen skattas enligt; $S(t) = P(T > t)$. Varje observations överlevnadstid *t* kan också ses som ett värde för en slumpmässig variabel T. Där T kan anta alla positiva värden sådana att $0<T<\infty$. T antas ha en fördelningsfunktion F(t) och en täthetsfunktion f(t), från vilka överlevnadsfunktionen kan bestämmas.

$$S(t) = P(T > t) = 1 - F(t) = 1 - P(T < t) = 1 - \int_0^t f(u)du \qquad (1)$$

Hazardfunktioner skattar risken för en individ att uppleva händelsen vid tidpunkt *t* givet att individen har överlevt fram till den tidpunkten, $P(t \le T < t + \Delta / T \ge t)$, där delta är en tidsenhet. Risken att uppleva händelsen, hazarden, kan redovisas både som den kumulativa hazarden, H(t), och hazardintensiteten, h(t).

$$h(t) = \lim_{\Delta \to 0} \left\{ \frac{P(t \le T < t + \Delta / T \ge t)}{\Delta} \right\} = \frac{f(t)}{S(t)} \qquad (2)$$

Formel 2 ger hazardfunktionen. Genom att dividera sannolikheten för att uppleva händelsen vid varje tidpunkt med tidsenheten ges risken för att uppleva händelsen vid tidpunkt *t*. Observera att hazardfunktionen alltså ger hazarden för den specifika tidpunkten. En liknelse för en mätning av en specifik tidpunkt (Kleinbaum 2005) är hastighetsmätaren på bilen. Mätaren visar 50 km i timmen, om du fortsätter i samma hastighet kommer du ha kört 50 km om en timma. Men det är fullt troligt att du kommer stanna eller ändra hastighet, liksom det är möjligt att hazarden kommer ändras över tid. Hazarden, och hastighetsmätaren, kan med andra ord bara ses som en punktskattning och säger ingenting om resten av kurvan.

Vad som vidare bör beaktas angående hazarden är att den anger en kvot och inte en sannolikhet. Hazarden kan anta ett värde mellan noll och oändligheten (inte noll och ett som en sannolikhet), beroende på vilken enhet tiden mäts i. (Kleinbaum 2005)

Hazarden kan också redovisas som den kumulativa hazarden. Sambandet mellan hazarden och den kumulativa hazarden är detsamma som mellan sannolikhetsfördelningen och dess täthetsfunktion.

$$H(t) = \int_0^t h(t)dt \tag{3}$$

Sambandet mellan överlevnadsfunktionen och den kumulativa hazardfunktionen kan skrivas som formel (4) alternativt formel (5).

$$H(t) = -\log S(t) \tag{4}$$

$$S(t) = \exp\{-H(t)\} \tag{5}$$

Överlevnads- och hazardfunktioner kan skattas både då fördelningen från vilket överlevnadsdatat kommer ifrån är känd, och då fördelningen inte är känd.

4.2 Icke-parametriska metoder.

Icke-parametriska metoder har fördelen att de är fördelningsfria, överlevnads- och hazard funktioner kan skattas utan någon kunskap om huruvida observationerna följer någon känd fördelning för T. Den enklaste icke-parametriska och modellfria skattningen av överlevnadsfunktionen, ”*the empirical survivor function*”, kan användas om datamaterialet inte innehåller några censurerade observationer. Den empiriska fördelningsfunktionen, $\hat{F}(t)$ skattas då som ration av det totala antalet individer som ej upplevt händelsen gentemot det totala antalet individer i studien.

$$\hat{S}(t) = 1 - \hat{F}(t) \quad \hat{S}(t) = \frac{\text{individer med kortare överlevnadstid än t}}{\text{individer i hela datasetet}} \tag{6}$$

Eftersom datamaterialet med kommunfullmäktigeledamöters överlevnadstid innehåller censurerade observationer kommer de två icke-parametriska metoderna Kaplan-Meier och Life-Table att användas i analysen. Båda metoderna har fördelar vad gäller presentation av överlevnadsdata.

4.2.1 Kaplan-Meier metoden.

Kaplan-Meiers skattade överlevnadsfunktion är en generalisering av den empiriska överlevnadsfunktionen; överlevnadsfunktionen beräknas som överlevande till tidpunkt *t* i förhållande till samtliga observationer i undersökningen, men hänsyn tas även till censurerade observationer.

Observationerna delas upp i j= 1,2,3, … , m intervall, där den exakta tidpunkten för varje *händelse* innebär början på ett nytt intervall. Om flera händelser sker vid samma tidpunkt innehåller intervallet flera händelser. Inget nytt intervall påbörjas vid tidpunkterna för de censurerade observationerna. De censurerade observationerna ingår i intervallet för den händelse de föregås av.

Kaplan-Meiers överlevnadsfunktion skattas enligt följande:

$$\hat{S}(t) = \prod_{j=1}^{k} \left(\frac{n_j - d_j}{n_j} \right) \tag{7}$$

Där n_j = antal ind. utsatta för risk i intervall j, dvs $n_j = n - ((d_1 + \ldots + d_j) + (c_1 + \ldots + c_{j-1}))$
d_j= antal händelser i intervall j.

c_j= antal censurerade observationer i intervall *j*.

$\left(\frac{n_j - d_j}{n_j}\right) = \hat{p}_j$ där $\hat{p}_j$ är sannolikheten för att överleva intervall *j*, proportionen av överlevare med andra ord.

Standardavvikelsen för skattningarna fås enligt följande.

$$se\{\hat{S}(t)\} \approx \hat{S}(t)\left\{\sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d)}\right\}^{1/2} \tag{8}$$

4.2.2 Life-table metoden

Life-Table metoden delar in observationerna i intervall som, till skillnad från Kaplan-Meier metoden inte är beroende av tidpunkterna för händelserna. Intervallens längd bestäms a priori. Det *j:*te intervallet innehåller alla händelser som inträffat mellan intervallets start och slutpunkt, och *j*= 1,2,3, … ,m. När överlevnadsfunktionerna skattas förlorar man informationen om den exakta tidpunkten för händelsen, alla händelser antas ha inträffat i början av intervallet. De censurerade observationerna inträffar likformigt över hela intervallet, vilket ger ett risk set, n'_j där: $n'_j = n_j - c_j/2$ (Collett 2003).

Överlevnadsfunktionen för Life-Table metoden skattas enligt samma formel som för Kaplan-Meier metoden. Den enda skillnaden är risktsetet som här är n'_j istället för n_j.

$$S^*(t) = \prod_{j=1}^{k} \left(\frac{n'_j - d_j}{n'_j}\right) \tag{9}$$

Med standardavvikelsen:

$$se\{S^*(t)\} \approx S^*(t)\left\{\sum_{j=1}^{k} \frac{d_j}{n'_j(n'_j - d)}\right\}^{1/2} \tag{10}$$

Hazarden för Life-table metoden skattas enligt formel 11.

$$h^*(t) = \frac{d_j}{(n'_j - d_j)\tau_j} \tag{11}$$

där τ_j är längden på intervall *j*.
Standardavvikelserna för hazardskattningen ges enligt formel 12.

$$se\{h^*(t)\} \approx \frac{h^*(t)\sqrt{1 - \left[h^*(t)\tau_j/2\right]^2}}{\sqrt{(d_j)}} \tag{12}$$

4.2.3 Test för jämförelse mellan gruppers överlevnadskurvor.

De två icke-parametriska testen, wilcoxon och log-rank, kommer att användas för att jämföra gruppers överlevnadskurvor med varandra. Testen har, med sina olika teststatistikor, olika fördelar, och ger lite skilda resultat. Wilcoxon och log-rank testar hypotesen att det inte är skillnad mellan två eller flera gruppers överlevnadskurvor. Teststatistikan är chi-två fördelad med *q*, antal grupper minus en, frihetsgrader.

Log-rank ochWilcoxon testar följande nollhypotes:
$H_0 : S_1(t) = S_2(t)$
Mot alternativhypotesen:
$H_1 : S_1(t) \neq S_2(t)$

I undersökningen kommer grupperna att jämföras med stratifierade test. Stratifierade test kombinerar effekterna av skillnaderna i varje gruppmitt. Om antal grupper som ska jämföras är ≥ 2 så summeras i teststatistikans täljare över *k* gruppers värden U_{Lk} där k = 1, 2, … ,g-1. I nämnaren summeras över variansen.

Teststatistikorna för de båda testen skiljer sig lite åt, wilcoxon lägger mer vikt vid skillnader som uppstår i tidigt skede av perioden än vad log-rank testet gör. I wilcoxon testet multipliceras i teststistikans täljare och nämnare vikterna med n_j, antal individer utsatta för risk vid tidpunkt *j*. Antal individer som är utsatta för risk, det vill säga de observationer som ännu ej har blivit censurerade eller upplevt händelsen är ju störst i början av studieperioden för att sedan minska.

Log-rank testets teststatistika har följande komponenter;

$$U_{Lk} = \sum_{j=1}^{r} \left(d_{kj} - \frac{n_{kj} d_j}{n_j} \right) \tag{13}$$

$$V_{Lkk'} = \sum_{j=1}^{r} \frac{n_{kj} d_j (n_j - d_j)}{n_j (n_j - 1)} \left(\delta_{kk'} - \frac{n_{k'j}}{n_j} \right) \tag{14}$$

där $\delta_{kk'} =$ 1 om k = k´ och 0 annars.

Teststatistika:

$$\frac{\left(\sum_{k=1}^{S} U_{Lk} \right)^2}{\sum_{k=1}^{S} V_{Lk}} \sim \chi_{(q)}^{2} \tag{15}$$

Wilcoxontestet har en teststatistika med följande komponenter;

$$U_{Wk} = \sum_{j=1}^{r} n_j \left(d_{kj} - \frac{n_{kj} d_j}{n_j} \right) \tag{16}$$

$$V_{Wkk'} = \sum_{j=1}^{r} n_j^2 \frac{n_{kj} d_j (n_j - d_j)}{n_j (n_j - 1)} \left(\delta_{kk'} - \frac{n_{k'j}}{n_j} \right) \tag{17}$$

Den stratifierade wilcoxon teststatistikan är således;

$$\frac{\left(\sum_{k=1}^{S} U_{Wk} \right)^2}{\sum_{k=1}^{S} V_{Wk}} \sim \chi_{(q)}^{2} \tag{18}$$

En intressant aspekt av teststatisktikornas olika uppbyggnad är att de, vid olika resultat, ger en indikation på att skillnaden mellan gruppernas överlevnadskurvor inte är konstant över tiden. Detta kan tyda på oproportionallitet, vilket kommer att utredas mer i nästa avsnitt.

4.3 Cox proportionella hazard modell, en semiparametrisk modell.

Ett primärt syfte i vår undersökning, såsom vanligtvis i överlevnadsanalys (Collett 2003) är att utforma en modell med de oberoende variabler som bäst förklarar hazarden för händelsen. Modellbyggandet möjliggör dessutom estimering av en specifik individs hazardfunktion utifrån förklaringsvariablerna.

Cox proportionella hazard modell är en semiparametrisk modell, koefficienterna för kovariaten skattas med hjälp av likelihoodskattning, men den så kallade baslinjen för modellen lämnas ospecificerad. Cox modellen kan skattas utan kännedom om datamaterialets sannolikhetsfördelning. Resultaten som Cox modellen ger är nära de resultat en rätt specificerad parametrisk modell skulle ge. Det gör den till ett pålitligt modellval då det kan vara svårt att med säkerhet hitta rätt parametrisk modell.

Den generella proportionella hazard modellen har följande form:

$$h_i(t) = \exp(\beta_1 x_{i1} +,...,+\beta_k x_{ik}) h_0(t) \tag{19}$$

Den beroende variabeln $h_i(t)$ är hazarden vid tidpunkt *t* för den *i*:te individen, och är en produkt av två funktioner. Dels är hazarden en funktion av kovariaten, $x_1, x_2, \ldots, x_k$ med respektive regressionskoefficienter $\beta_1, \beta_2, \ldots, \beta_k$. Dels är hazarden en funktion av modellens *baseline hazard function*; $h_0(t)$.

Baseline-hazarden representerar, i en modell med ett kontinuerligt kovariat, hazarden för en individ (eller observation) vars värde för variablerna i modellen är noll. I en modell med dummyvariabler eller faktorvariabel i flera nivåer representerar baseline-hazarden hazarden för de individer vars värde för faktorn eller dummyvariabeln är noll (Collett 2003).

Baslinjen kan, som nämnt ovan, lämnas ospecificerad och behöver inte skattas för att parametrarna ska kunna skattas. Detta förutsätter att antagandet om proportionallitet mellan observationerna gäller, det vill säga att även om hazardkurvorna för två individer är olika så är de parallella med varandra.

Gäller detta kan man tänka sig en jämförelse mellan hazarden för två individer, *i* och *j*, där baslinjen kommer att försvinna, och således inte behöver skattas. (Gou 2010)

$$\frac{h_i(t) = h_0 \exp(\beta_1 x_{i1} + \ldots + \beta_k x_{ik})}{h_j(t) = h_0 \exp(\beta_1 x_{j1} + \ldots + \beta_k x_{jk})} = \exp\left[\beta_1 (x_{i1} - x_{j1}) + \ldots + \beta_k (x_{ik} - x_{jk})\right] \tag{20}$$

Baslinjerna tar ut varandra, och skattningarna i modellen ger endast skillnaderna mellan individerna utan att $h_0(t)$ behöver specificeras. Man kan också tänka att *i* och *j* är olika nivåer i en faktorvariabel.

Skillnaden mellan individerna, den relativa hazarden, kan också redovisas enligt formel 21, där ψ är värdet för alla koefficienterna upphöjt av e, vid tidpunkt *t*. Den relativa risken som formeln representerar fås genom att flytta över $h_0(t)$ i den generella proportionella hazard modellen (20) till vänsterledet och därefter logaritmera på båda sidorna (Collett 2003).

$$\psi(x_i) = \exp\left(\beta' x_i\right) \tag{21}$$

Hazard ration för en variabel kan enkelt räknas fram genom att ta e upphöjt till lutningskoefficienten; $\psi = \exp(\beta)$. För en kontinuerlig variabel är tolkningen av hazard ration förändringen i hazarden då variabeln förändras med en enhet. För en faktor variabel är hazard ration skillnaden i hazarden mellan nivåerna när övriga kovariat hålls konstanta.

4.3.1 Skattning av parametrarna och baslinjen i coxmodellen.

Koefficienterna i Cox-modellen skattas, såsom koefficienterna i de parametriska modellerna, med hjälp av maximum likelihood skattning. I Cox-modellen kan parametrarna skattas med partial likelihhod. Partial likelihood har fördelar mot vanlig likelihood som marginal och conditional likelihood. Enligt Cox (1975) är partial likelihood enklare att använda än den fulla likelihoodskattningen då den endast involverar de parametrar som faktiskt är av intresse. De censurerade observationerna tas inte med i skattningen. Baslinjerna tar som tidigare nämnt ut varandra. Partial likelihood kan ses som en generalisering av marginal och conditional likelihood, och ger vid stora urval likvärdiga skattningar och teststatistikor som de senare (Cox 1975).

Partial likelihoodskattningen för parametrarna ges av följande formel enligt Cox (1975)

$$L(\beta) = \prod_{j=1}^{r} \frac{\exp(\beta' x_{(j)})}{\sum_{l \in R(t_{(j)})} \exp(\beta' x_l)} \tag{22}$$

Likelihooden för alla individer i risksetet, r, multipliceras och i nämnaren för likelihoodskattningarna summeras (endast) över de individer som fortfarande är med i risksetet.

För att närmare förklara att partial likelihoodskattningen dels kan hantera censurerade observationer och dels att summeringen i nämnaren endast är för de individer som är med i risksetet anges nedan en annan variant av formeln för likelihoodskattningen (Gou 2009).

Likelihooden för en individ, L_1, fås genom att dividera dess hazard, $h_1(1)$, med summan av alla *n* individers hazard ($h_1(1)+\ldots+h_n(n)$). När sedan alla likelihooden multipliceras ordnas de efter tid till händelse/censurering.

$$PL = \prod_{i=1}^{n} L_i = \prod_{i=1}^{n} \left[\frac{\exp(\beta x_i)}{\sum_{i=1}^{n} Y_{ij} \exp(\beta x_i)} \right]^{\delta} \tag{23}$$

I formeln är $\delta = 0$ om observationen är censurerad, så L_i för observationen blir ett, annars är $\delta = 1$. Y_{ij} är ett sätt att se till att endast observationer som är med i risksetet fortfarande kommer med i skattningarna på så sätt att $Y_{ij} = 1$ om $t_j > t_i$ och $Y_{ij} = 0$ om $t_j < t_i$.

Partial likelihood formeln förutsätter att flera händelser inte inträffar samtidigt. Eftersom tiden i studien mäts i veckor kan flera händelser inträffa vid samma tidpunkt. Det finns flera olika metoder för att hantera detta. Efron-approximationen är den metod som kommer att användas när datamaterialet analyseras i SAS.

Om modellen innehåller förklaringsvariabler kan baslinje hazarden förstås som risken för att en individ upplever händelsen i en specifik tidsenhet, minus effekten som förklaringsvariablerna har. ξ i ekvationen nedan är en funktion av variablerna och lutningskoefficienterna.

$$\hat{h}_0(t) = \frac{1 - \hat{\xi}_j}{t_{(j+1)} - t_{(j)}} \tag{24}$$

Om det inte finns några förklaringsvariabler i modellen skattas den proportionella Cox modellens baslinje enligt samma procedur som hazardkurvan i Kaplan-Meier metoden. Utan variabler följer att; $\xi_i = (n_j - d_j)/ n_j$. Baslinjen skattas enligt följande;

$$\hat{h}_0(t) = \frac{d_j}{n_j(t_{(j+1)}) - (t_{(j)})} \tag{25}$$

4.3.2 Semiparametriska test

Flera teststatistikor kommer användas för att jämföra olika Cox modellerna med varandra för att hitta den bästa variabelkombinationen. Wald-testet är ett signifikanstest som testar om β-koefficienten är skilt från noll i närvaro av de andra variablerna i modellen. Teststatistikan är chi-två fördelad. Signifikansnivå för parametrarna som Wald-testet ger kommer användas för att diskutera om variablerna är bra att ha med i modellen.

För att jämföra modeller med varandra kan -2log^L statistikan användas. Det är en metod för att jämföra modeller om den ena modellen är nested i den andra.

$$(-2\log \hat{L}(1)) - (-2\log \hat{L}(1)) \sim \chi^2_q \tag{26}$$

Om p-värdet visar på signifikant skillnad mellan modellernas $-2\log \hat{L}$ tyder det på att de extra variablerna i den större modellen bör vara inkluderade. Om $-2\log \hat{L}$ -värdet minskar då ytterligare variabel läggs till i modellen indikerar det alltså att modellen förbättras.

4.3.3 Test av proportionalitetsantagandet

För att grafiskt utvärdera proportionallitetsantagandet och få en indikation på om detta stämmer kan den logaritmerade kumulativa hazarden mot logaritmen av tiden i antal veckor undersökas. Kurvorna för de olika nivåerna i en kategorivariabel ska inte avvika från varandra utan löpa parallellt om proportionalitetsantagandet ska gälla (Collett 2003). Det kan dock vara svårt att få en riktig uppfattning om proportionaliteten med en visuell tolkning.

En annan, mer formel metod för att upptäcka avvikelser från proportionalitetsantagandet är med hjälp av Schoenfeld residualer. Varje individ har inte bara en residual, utan ett antal residualer för varje kovariat i modellen. Schoenfeldresidualerna är oberoende av tiden och kan plottas mot tiden för att upptäcka eventuell proportionalitet i modellen (Schoenfeld 1982). De är endast de observationer som inte är censurerade som får ett residualvärde. (Allison 2010)

Den *i*:te Schoenfeld residualen för den *j*:te förklarande variabeln x_j ges av.

$$r_{Pji} = \delta_i (x_{ji} - \hat{a}_{ji}) \tag{27}$$

Där x_{ji} är värdet på den *j*:te oberoende variabeln, $j = 1,2\ldots.p$, för den *i*:te individen. För de censurerade observationerna är δ lika med noll och för de icke-censurerade är δ lika med ett.

$$\hat{a}_{ji} = \frac{\sum_{I \in R(t_i)} x_{jI} \exp(\hat{\beta}' x_I)}{\sum_{I \in R(t_i)} \exp(\hat{\beta}' x_I)} \tag{28}$$

$R_{(ti)}$ är de individer utsatta för risk vid tiden t_i. Skulle den längsta överlevnadstiden vara ocensurerad är $\hat{a}_{ji}$ lika med x_{ji} och rp_{ji} får värdet noll.

Att testa om Schoenfeld residualerna är beroende av tiden eller inte är detsamma som att testa om det finns proportionallitet i Cox modellen (Allison 2010). Skulle antagandet om proportionallitet stämma ska Schoenfeld residualerna vara oberoende av tiden. Proportionallitet testas genom en korrelationsmatris för korrelationen mellan Schoenfeld residualerna och funktioner av tiden. Om det är signifikant att det finns korrelation mellan Schoenfeld residualerna och funktioner av tiden indikerar det icke-proportionalitet.

Utifall proportionallitesantagandet inte skulle hålla för en eller flera av de förklarande variablerna i modellen finns det i huvudsak två sätt att hantera detta. Ett sätt att hantera icke-proportionalitet är att stratifiera på den variabeln som förändras med tiden. Vilket ger en proportionell hazardmodell i varje stratum för de övriga variablerna i modellen (Klein m.fl 2003). Partial likelihood funktionerna multipliceras sedan ihop till en funktion.

$$h_i = h_{i0}(t)\exp(\beta_1 x_i + ... + \beta_k x_k) \tag{29}$$

Det är en fördel om stratifieringsvariabeln är ointressant för studien då inga skattningar för koefficienten av variabeln ges (Allison 2010).

En annan strategi är att skapa en ny variabel av den variabel som bryter mot proportionalitetsantagandet och låta den nya variabel vara en lämplig transformation av tiden. Den nya interaktionsvariablen ska vara transformerad på det sättet att den fångar upp effekten av icke-proportionalliteten.

$$h(t) = h_0(t)\exp(\beta_1 x_i + \beta_2 x_i t) \tag{30}$$

Är den skattade koefficienten β_2 negativ sjunker effekten av kovariatet med tiden och är den positiv ökar effekten. β_1 ska tolkas som effekten av kovariatet när tiden är noll (Allison 2010).

4.4 Parametriska modeller

En parametrisk överlevnadsfunktion är en modell där överlevnadstiden kan antas följa en känd fördelning. De parametriska modeller som presenteras här är samtliga Accelerated Time Failure modeller, (ATF). I ATF modeller är kovariaten multiplikativa i förhållande till överlevnadstiden. Utan censurerad data skulle modellerna kunna skattas med ordinary least squares som en vanlig regression, men med censurerad data skattas modellerna med maximum likelihood metoden.

T_i är en stokastisk variabel för överlevnadstiden för den i:te individen och $x_{i1}, x_{i2}, \ldots, x_{ik}$ är värden på *k* antal kovariat för samma individ.

$$\log T_i = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik} + \lambda\varepsilon_i \tag{31}$$

ε_i är feltermen som följer en speciell distribution, och $\beta_0, \ldots, \beta_k$ samt λ är parametrarna som ska skattas, där λ är en skalparameter. Variansen för feltermen ε kan fixeras och värdet på λ står då istället för variationen av ε (Allison 2010). Den beroende variabeln och feltermen tillåts följa andra fördelningar än normalfördelningen (Guo 2010).

Den beroende variabeln i de parametriska modellerna, log T, är ett mått på överlevnadstiden och skiljer sig från Cox-modellen där den beroende variabeln är ett mått på hazarden. En hög överlevnads tid motsvaras av en låg hazard. Genom e^{β} ges den genomsnittliga överlevnadstiden i förhållande till baslinjekategorin när de övriga kovariaten hålls konstanta.

Weibull och den exponentiella modellen är både AFT-modeller och proportionella hazard modeller. Regressionskoefficienterna kan transformeras till log hazard format. För den exponentiella modellen ändras bara tecken på β-koefficienten. Log hazard ges för weibull modellen genom att ändra tecken på β-koefficienten och dividera med skalparameterskattningen.

Är formen på överlevnadsfunktionen känd blir skattningarna mer precisa och får mindre medelfel än med Cox modellen.

4.4.1 Exponentiell fördelning

Den enklaste parametriska modellen att anpassa till överlevnadsdata är den exponentiella. Hazarden i den exponentiella modellen antar att hazarden är konstant. Risken att uppleva händelsen är densamma oavsett hur lång tid som förflutit (Collett 2003).

Hazarden för den exponentiella modellen ges av

$$h(t) = \lambda \quad \text{för } 0 \le t < \infty \tag{32}$$

Med överlevnadsfunktionen

$$S(t) = e^{-\lambda t} \tag{33}$$

där parametern λ är en positiv konstant. Vid ett högt värde på λ blir hazarden hög och motsvarande överlevnadstid låg.

4.4.2 Weibull fördelning

Den vanligaste parametriska modellen för överlevnadsfunktionen är Weibull modellen. Weibull modellen har två parametrar, en skalparameter λ och en shape parameter γ som bestämmer formen på hazard funktionen.

När γ >1 ökar hazarden med tiden och när γ=1 övergår Weibull fördelningen i en exponentiell fördelning. Den exponentiella fördelningen är alltså ett specialfall av Weibull modellen. När γ<1 minskar hazarden med tiden.

Hazardfunktionen för Weibull modellen ges av

$$h(t)\lambda\gamma \cdot t^{\gamma-1} \tag{34}$$

Överlevnadsfunktionen blir då

$$S\{t\}\ \exp(-\lambda t^{\gamma}) \tag{35}$$

Shape parametern gör Weibull modellen till en flexibel modell, men hazardfunktionen är relativt enkel i sin form.

4.4.3 Log-normal fördelning

Denna fördelning skiljer sig från de andra modellerna då den antar att fördelningen fortfarande är normal. Har log *T* en normal fördelning har T en log-normal fördelning. Den log-normala distributionen har inte en monotonisk hazard funktion. Hazarden är lika med 0 när tiden är 0 och har sedan en stigande kurva som pikar och sedan sjunker mot noll när tiden går mot oändligheten.

4.4.4 Den log-logistiska fördelningen.

Den log-logistiska modellen skiljer sig från de andra AFT-modellerna som redogjorts för såtillvida att den inte är en generalisering av extended generelized gamma modellen. Till skillnad från tillexempel Weibull och exponential modellerna, men i likhet med den log-normala modellen, så har inte den log-logistiska modellen en tidsfunktion som är monoton (Collett 2003). Fördelen med det är att modellen tillåter hazardrisken att förändras över tid. Likt den log-normala modellen kan hazardfunktionen anta en ”svängd” form (Allison 2010). Om λ<1 får hazardkurvan en form liknande den log-normala, men om λ>1 liknar den en avtagande weibull.

Hazardfunktion beskrivs enligt följande

$$h(t) = \frac{e^{\theta}\kappa t^{\kappa-1}}{1+e^{\theta}t^{\kappa}} \quad (36)$$

Och överlevnadsfunktionen

$$S(t) = \frac{1}{1+e^{\theta}t^{\kappa}} \quad (37)$$

4.4.5 Gamma fördelningen

Gamma fördelningen har egenskaper som liknar Weibull modellens. Den exponentiella modellen är också ett specialfall av gamma modellen. Om shape parametern är =1 reduceras gamma modellen till en exponentiell modell. Skulle shape p. vara > 1 tar den formen av en stigande hazardkurva som planar ut vid värdet för λ. När shape < 1 råder motsatt förhållande. Hazardkurvan avtar och planar ut vid värdet för λ.

4.4.6 Generalized gamma (EGG)

Den exponentiella, Weibull och log-normal fördelningen är alla specialfall av den generaliserade gamma modellen.

Det gör den generaliserade gamma modellen till en användbar modell för att urskilja och jämföra alternativa parametriska överlevnadsfunktioner och hitta den modell som passar bäst till överlevnadstiden. EGG modellen har tre parametrar och hazardfunktionen kan anta många olika former.

- När λ=1 och γ=1 inskränks EGG modellen till en exponentiell fördelning.
- När γ=1 blir den Weibull fördelad
- När γ går mot oändligheten blir EGG modellen en log-normal modell.

4.4.7 Parametriska test Goodness of fit test – Likelihood ratio test

De olika ATF modellerna kan ge liknande skattningar av koefficienterna men gör olika antaganden om formen på hazardkurvan.

Precis som för Cox-modellerna kommer likelihood ratio-test användas för att jämföra modeller. Teststatistikan används för att utvärdera vilken av AFT modellerna som är bäst anpassat till datamaterialet. Likelihood statistikan jämför modeller som båda är specialfall av en annan modell. Är modell A ett specialfall av modell B kan modell A erhållas genom att införa restriktioner på parametrarna i modell B. (Allison 2010)

För att se om modell A är en lämplig modell multipliceras skillnaden mellan log-likelihooden för modell A och modell B med -2.

$$-2[(\log L(A)) - (\log L(B))] \quad (38)$$

Noll hypotesen i likelihood ratio testet är att restriktionen för modell A inte är skilt från de fria skattningarna av scale- och shape-parametrarna i modell B. Alternativhypotesen är att skattningarna för scale- och shape- parametrarna är skilt från noll.

Test statistikan är chi-två fördelad med antal frihetsgrader lika med antal restriktioner av parametrar. Förkastas noll hypotesen är modell B ett bättre val än modell A.

Då den exponentiella, Weibull och log-normal modellen är alla specialfall av generalized gamma modellen kan de utvärderas med ett likelihood ratio test. Passformen för den Log-logistic modellen som ju inte är ett specialfall av EGG kan inte testas med ett likelihood ratio test.

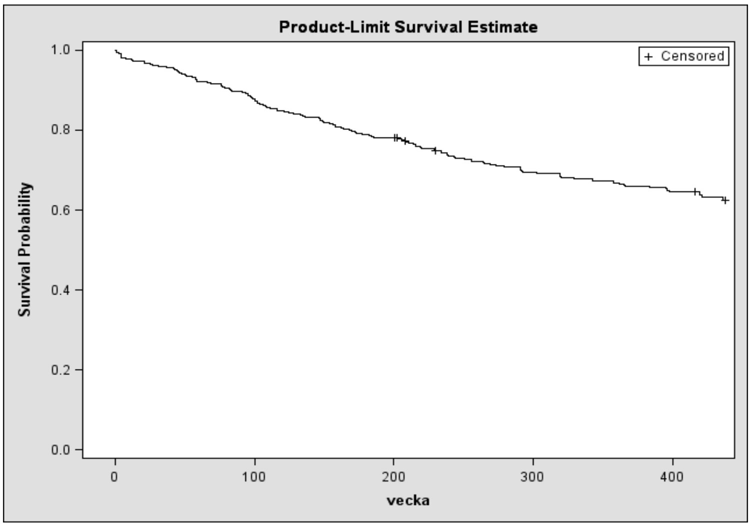
5. RESULTAT

5.1 Överlevnads- och hazardkurvor, skattade med Kaplan-Meier och Life-table.

Hazard och överlevnadskurvor kommer att redovisas både för hela datamaterialet och uppdelat på kommungrupp. Då observationer samlats in både från år 2002 och år 2006 kommer vecka 1 till och med vecka 230 representeras både av alla ledamöter som fanns med i kommunfullmäktige efter valet år 2002 och av de nyinvalda ledamöterna vid år 2006 års val. Resultaten för vecka 230 och framåt representeras bara av ledamöter som har suttit från år 2002. (Nyinvalda ledamöter från år 2006 har automatiskt censurerats efter 230 veckor)

5.1.1 Kaplan-Meier överlevnadskurvor:

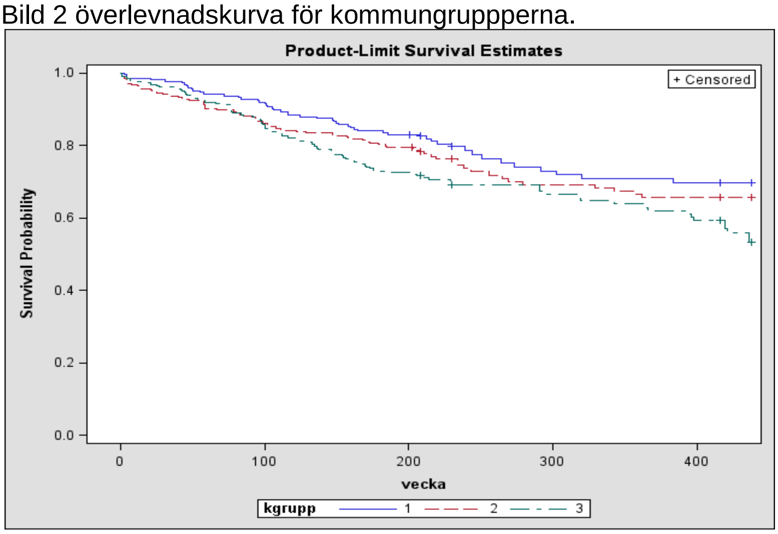
Bild 1 överlevnadskurva för hela datamaterialet



Grafen ger sannolikhet för att sitta kvar ett visst antal veckor. De lodräta strecken anger censurerade observationer. Censurering har skett vid fyra olika tidpunker; 208 veckor, 230 veckor, 416 veckor och 438 veckor (som är studiens slutpunkt).

Sannolikheten att sitta kvar som ledamot i de 438 veckorna eller längre i kommunfullmäktige är 0,62. Medelvärdet är 339 veckor, vilket motsvarar ungefär 1,5 mandatperiod.

Bild 2 överlevnadskurva för kommungruppperna.



Grafen anger precis som bild 1 sannolikheten för ledamöter att sitta kvar visst antal veckor, där de lodrätta strecken representerar tidpunkt för censurerade observationer. Kurva 1 representerar kommungrupp1, kurva 2 representerar kommungrupp 2 och kurva 3 representerar kommungrupp3.

Bild två visar en lägre överlevnadskurva för individer i kommungrupp tre gentemot de andra två. Kurvan för kommungrupp 1 är skild från de andra under hela undersökningsperioden. De andra två kurvorna skiljs åt efter knappt två år. Både log-rank och wilcoxon nedan visar på signifikanta skillnader mellan kommungrupperna. Testerna visar bara om det är signifikant att det är skillnad mellan grupperna, inte vilka av grupperna som det är skillnad emellan. Men utifrån överlevnadskurvan så verkar skillnaden mellan grupperna vara den att de ledamöter i kommuner med lägst medelinkomst har kortare tid till avgång.

5.1.2 Log-rank och wilcoxon test.

I tabell 1 anges wilcoxon och log-rank teststatistikorna för stratifierade test för variablerna kommungrupp, tidigare erfarenhet, kön, block, parti, och opposition och ålder som kategorisk variabel.

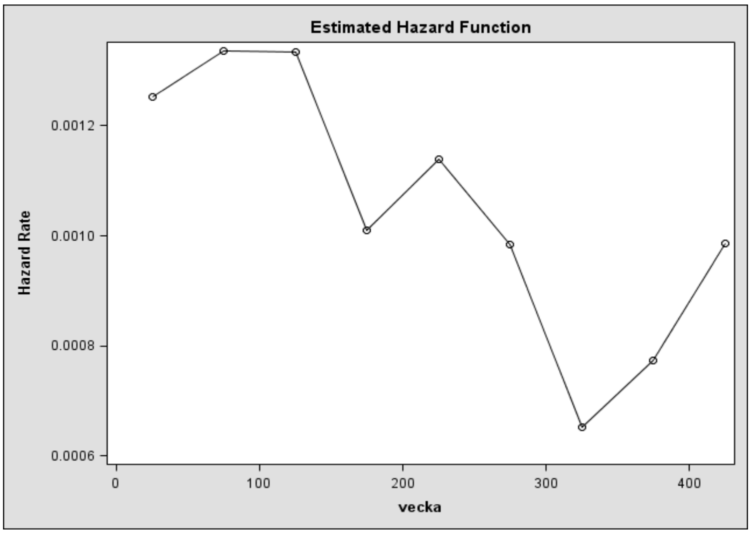
Tabell 1: Log-rank och Wilcoxontest.

Variabel	**Df**	**Log-rank-teststatistika (p-värde)**	**Wilcoxon-teststatistika (p-värde)**
Kommungrupp	2	9,2712 (**0,0097**)	8,2341 (**0,0163**)
Erf	1	27,3062 (<**0,0001**)	27,9642 (<**0,0001**)
Kön	1	0,1483 (0,7001)	0,2190 (0,6398)
Block	1	0,6123 (0,4339)	0,5656 (0,4520)
Parti	7	5,5771 (0,5899)	6,5841 (0,4729)
Opposition	1	3,5269 (0,**0604**)	5,5137 (**0,0189**)
Ålder_kat.	3	21,5610 (<0.0001)	21,8171 (<0.0001)

Båda testen ger samma signifikansnivåer för skillnad mellan grupperna. Det är signifikant för fyra av variablerna att kurvorna skiljer sig år; p-värdena för kommungrupp är likvärdiga mellan testerna 0, 0097 och 0, 0163, likaså är de p-värdena likvärdiga mellan testerna för erfarenhet, 0,0001, samt för den kategoriska åldersvariabeln, 0,0008 och 0,0004. Men för variabeln opposition, där båda testerna visar på signifikant skillnad mellan grupperna, är p-värdena relativt olika. Log-rank testet ger ett p-värde på 0. 0604 och wilcoxon testet ger ett p-värde på 0, 0189. Detta ger en indikation på att skillnaden mellan grupperna är större i början av studieperioden. Wilcoxon testet ger större vikter åt skillnader i början av perioden.

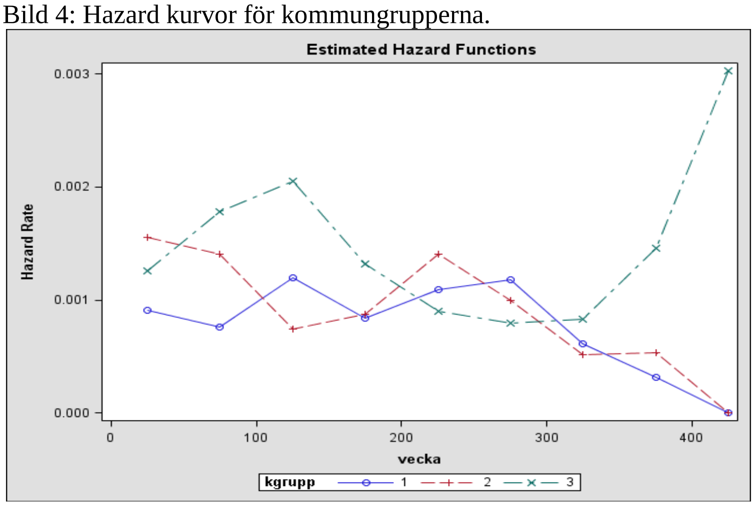
5.1.3 Hazardkurvor för hela materialet skattade med life table.

Bild 3: Hazardkurva för hela datamaterialet.



Hazardkurvan stiger initialt för att sedan stadigt sjunka i etapper med uppgångar runt fyra och åtta år. Risken för avhopp ökar alltså I samband med valen. De första två åren har högre risk för avhopp jämfört med resten av undersökningsperioden.

Bild 4: Hazard kurvor för kommungrupperna.



Hazard kurvorna visar inte på stora skillnader mellan grupperna. Kommungrupp tre har ett avvikande mönster mot de andra grupperna. Dels stiger hazarden för grupp 3 kraftigt efter 330 veckor, och vid 120 veckor. Detta indikerar att av ledamöter som avgick efter valet 2010 och fram till studietidens slut tillhör många grupp 3. Kommungrupp 2 har en högre hazard vid studiens början och efter cirka 208 veckor. Risken för avhopp i kommungrupp 2 är störst strax efter valen. Risken för avhopp ligger relativt konstant i kommungrupp 1, men sjunker efter ca 300 veckor.

5.2 Cox proportionella hazard modell.

5.2.1 Proportionalitetstest.

Cox proportionella hazard modell bygger på att antagandet om proportionalitet ska gälla. För att testa proportionaliteten testas korrelationen av Schoenfeld residualerna för alla variabler med tre vanliga funktioner av tiden. Det första är tiden i antal veckor den andra är logaritmering av tiden i antal veckor, och den tredje tiden i antal veckor i kvadrat.

Tabell 2: Schoenfeld residualerna

Schoenfeld residual	Vecka p-värde	LogVecka p-värde	$Vecka^2$ p-värde
Ålder_kont:	0.4367	0.7255	0.4577
Erfarenhet	0.3733	0.8487	0.1809
Kommunstratum	0.2958	0.5743	0.1283
Opposition	0.0206	0.0012	0.1495
Parti	0.7254	0.8258	0.7244
Block	0.5735	0.2649	0.5763
Kön	0.7252	0.9803	0.5150
Ålder_kat	0,7006	0,8852	0,6985

Då ålder_kat och ålder_kont ej kan inkluderas i samma modell har en korrelationsmatris med resp. variabel skattat. Skattningarna för de övriga variablerna blev snarlika oavsett modell, varför endast Ålder_kat visas från korrelationsmatrisen med den variabeln.

Vi testar nollhypotesen, att det inte finns någon beroende mellan tiden och Schoenfeld residualen, det vill säga ingen korrelation mellan Schoenfeld residualerna och funktionerna av tiden, mot alternativhypotesen att det finns beroende. I tabell 2 visas p-värdena för korrelatonsmatrisen. Ett signifikant p-värde påvisar beroende, och problem med proportionalitetsantagandet.

För variablerna; ålder som kategorisk variabel, ålder som kontinuerliga variabel erfarenhet, kommunstratum, parti, block och kön kan vi inte förkasta nollhypotesen om proportionalitet.

Den enda variabel som signifikant visar på beroende mellan felterm och tid är opposition. Dock kan inte nollhypotesen om proportionalitet förkastas på samtliga funktioner av tiden då det inte är signifikant för alla funktioner av tid att proportionalitetsantagandet inte gäller. På tidsavstånden veckor respektive, log(vecka) är det signifikant på 5 procents- och 1 procents nivå att proportionalitetsantagandet kan förkastas, men det går inte att förkasta nollhypotesen då vi testar beroende mellan felterm och tid i kvadrat.

I appendix E redovisas en graf för den log kumulativa hazarden mot logaritmen av tiden i antal veckor för variabeln opposition. Grafen ger en visuell bild av proportionaliteten mellan grupperna. Tolkningen av huruvida kurvorna är proportionella eller inte är subjektiv, och därför svårtolkad.

Som nämnts i metoddelen finns det flera sätt att hantera problemet med ickeproportionella variabler. Stratifiering av en variabel leder till att inga skattningar ges för variabeln. Då opposition tillför något till att förklarar avhopp bland kommunfullmäktige är det intressant att ha den kvar modellen. Istället kommer en modell skattas där variabeln opposition transformeras till ett tidsberoende kovariat.

5.2.2 Jämförelse mellan Cox-modeller.

Nästa steg i processen för att hitta den modellen med de bästa förklaringsvariablerna, kommer att göras genom att skatta Cox proportionella hazard modell.

För att anpassa en modell med ett bra variabelset kan man använda sig av olika tillvägagångssätt (Collet 2003). En metod är att testa varje variabel för sig. För att sedan bygga ihop modellen genom att lägga till en variabel i taget, utifrån dess signifikansnivå, och behålla varje variabel som tillför modellen något. En annan metod är att börja med en modell med alla variabler, och sedan plocka bort de variabler som är inte är signifikanta och heller inte tillför modellen något. Vi kommer att använda oss av den senare metoden.

Två cox-modeller kommer initialt att anpassas. Det är två modeller, båda innehållande kommungrupp, tidigare erfarenhet, opposition, block och parti, samt ålder som kategorisk variabel för modell(1) och ålder som kontinuerlig variabel för modell(2).

Modell(1) :

$h_i(t) = \exp\{\beta_1 ålder\_kont_i + \beta_2 k\_grupp_i + \beta_3 erf_i + \beta_4 opp_i + \beta_5 kön_i + \beta_6 block_i + \beta_7 parti_i\}h_0(t)$

Modell(2) :

$h_i(t) = \exp\{ \beta_1 ålder\_kat_i + \beta_2 k\_grupp_i + \beta_3 erf_i + \beta_4 opp_i + \beta_5 kön_i + \beta_6 block_i + \beta_7 parti_i\}h_0(t)$

Variabler kommer sedan att plockas bort ifrån modellerna till dess att den bästa modellen eller modellerna har anpassats med hänsyn till -2logL statistikan och signifikansnivå för parameterskattningarna, d.v.s. wald-statistikan.

Tabell 3: Maximum likelihoodskattningar för modell (1).

Parameter	**df**	**(β)**	**SE(β)**	**χ^2**	**P-värde**
ålder_kont	1	-0.01547	0.00534	8.4050	0.0037
Erf	1	-0.66585	0.15479	18.5041	<.0001
Opp	1	0.29327	0.14527	4.0756	0.0435
kgrupp1	1	-0.62414	0.18131	11.8500	0.0006
kgrupp2	1	-0.34778	0.15997	4.7263	0.0297
Kön	1	0.04408	0.13859	(0.1011	0.7505)
Block	1	0.16977	0.40807	(0.1731	0.6774)
parti 1	1	0.06082	0.32222	(0.0356	0.8503)
parti2	1	-0.24189	0.35916	(0.4536	0.5006)
parti 3	1	-0.60321	0.57931	(1.0842	0.2978)
parti 4	1	-0.13154	0.40937	(0.1032	0.7480)
parti 5	1	-0.33759	0.31767	(1.1294	0.2879)
parti 6	1	-0.18805	0.39556	(0.2260	0.6345)
parti 7	0	0			

I modellen gäller följande baslinjer; kommungrupp tre är baslinjen för variabeln kommungrupp, ingen tidigare erfarenhet är baslinjen för variabeln erfarenhet, ledamot ej i opposition för opposition, kvinnor för variabeln kön, det borgliga blocket för variabeln block, parti8 slutligen är baslinjen för variabeln parti. Inga skattningar ges för parti 7, detta har förmodligen att göra med kollinäritetsproblem mellan variabeln block och parti.

Variablerna ålder, tidigare erfarenhet, huruvida ledamoten satt i opposition samt kommungrupp har alla parameterskattningar som är signifikanta. Skattningarna för ålder, opposition och kommungrupp2 med p-värden på 0,0037 0,0435 och 0,0297 är signifikanta på 5% nivå. Skattningarna för kommungrupp1 med p-värde på 0,0006 och för tidigare erfarenhet med ett p-värde på <0,0001 är signifikanta på 1% signifikansnivå. Ingen av de andra variablerna har skattningar som är signifikant skilda från noll.

Tabell 4: Maximum likelihoodskattningar för modell (2).

Parameter	**Df**	**(β)**	**SE(β)**	χ^2	**P-värde**
ålder_kat 1	1	0,90962	0,38312	5,6371	0,0176
ålder_kat 2	1	0,80278	0,30044	7,1398	0,0075
ålder_kat 3	1	0,42958	0,28523	2,2683	0,1320
Erf	1	-0,65688	0,15422	18,1414	<,0001
Opp	1	-0,57003	0,18143	9,8715	0,0017
kgrupp1	1	-0,29936	0,15992	3,5040	0,0612
kgrupp2	1	0,29599	0,14581	4,1207	0,0424
Kön	1	0,04266	0,13913	(0,0940	0,7591)
Block	1	0,10800	0,40821	(0,0700	0,7913)
parti 1	1	0,01661	0,32247	(0,0027	0,9589)
parti2	1	-0,27052	0,35905	(0,5677	0,4512)
parti 3	1	-0,69376	0,57881	(1,4367	0,2307)
parti 4	1	-0,19938	0,40884	(0,2378	0,6258)
parti 5	1	-0,31336	0,31756	(0,9737	0,3238)
parti 6	1	-0,17921	0,39654	(0,2043	0,6513)
parti 7	0	0			

Samma baslinjer gäller som för modell 1. För den kategoriska åldersvariabeln är baslinjen ålderskategori 4, ledamöter som är 65 år eller mer.

Precis som i modell (1) är variablerna kommungrupp, tidigare erfarenhet, huruvida ledamoten satt i opposition samt åldersvariabeln de variabler för vilka skattningarna är signifikanta. Skattningarna för variablerna opposition och tidigare erfarenhet är relativt lika som de I modell (1). Vad gäller kommungrupp har signifikansnivån för skillnad mot kommungrupp 3, baslinjen, ändrats till 10% med ett p-värde på 0,0612 för kommungrupp2 och till strax över 1% för kommungrupp1 med ett p-värde på 0,0018.

För den kategoriska åldersvariabeln gäller att endast nivå två, 25-34 åringarna är signifikant skild från baslinjen, pensionärer, med ett p-värde på 0,0780.

Två nya modeller anpassas, modell (3) och modell (4), som är utvecklingar av modell (1) och modell (2), där variablerna block, kön och parti plockats bort då inga av deras parameterskattningar är signifikanta.

Modell(3) : $h_i(t) = \exp\{ \beta_1 \text{ålder\_kont}_i + \beta_2 \text{k\_grupp}_i + \beta_3 \text{erf}_i + \beta_4 \text{opp}_i \} h_0(t)$

Modell(4) : $h_i(t) = \exp\{ \beta_1 \text{ålder\_kat}_i + \beta_2 \text{k\_grupp}_i + \beta_3 \text{erf}_i + \beta_4 \text{opp}_i \} h_0(t)$

Modell(5) – modell(8) är generaliseringar av modell(3) där modell(5) är utan variabeln: k_grupp, modell(6) är utan variabeln: erf, modell(7) är utan variabeln: opp, och slutligen modell(8) är utan variabeln ålder_kont.

Liknande procedur gäller för modell(9) till modell(12). Modell(9) är en generalisering av modell(4) där modell(9) är en modell utan variabeln: k_grupp, modell(10) är utan variabeln: erf, modell(11) är utan variabeln: opp, och modell(12) är utan variabeln ålder_kat.

Med hjälp av -2log^L statistikan testas om förklaringsnivån på modellerna har minskas signifikant då de ickesignifikanta variablerna tas bort. Har det inte gjort det antas den mindre modellen vara bättre. Med samma procedur kontrolleras även om förklaringsgraden minskar då någon av de signifikanta variablerna tas bort. Teststatistikan fås genom att ta skillnaden för värdet för -2logL för den större modellen från den mindre. Om modell 1 är nested i modell 2 blir teststatistikan -2logL(1)- -2logL(2). Antalet frihetsgrader är det samma som skillnaden i variabler mellan modellerna. Vad som testas är alltså om den större modellen är bättre en den mindre, utifrån om dess variabler tillför något vad gäller förklaringsgrad.

H_0: -2logL(för den mindre modellen) - -2logL(för den större modellen) = 0
H_1: -2logL(för den mindre modellen) - -2logL(för den större modellen) > 0

Tabell 5: -2logL test

Större modellen vs den mindre	**-2logL – (-2logL)**	**df**	$\chi^2_{(q)}$	**p-värde**
Modell (1) vs Modell (3)	(-2638,574) – (-2633,973)	8	4,601	0,7995
Modell (2) vs Modell (4)	(-2646,148) – (-2641,424)	8	4,724	0,7881
Modell (3) vs Modell (5)	(-2651,053) – (-2638,574)	2	12,479	<0,002
Modell (3) vs Modell (6)	(-2658,930) – (-2638,574)	1	20,356	< 0,001
Modell (3) vs Modell (7)	(-2642,425) – (-2638,574)	1	3,851	
Modell (3) vs Modell (8)	(-2658,092) – (-2638,574)	1	19,518	< 0,001
Modell (4) vs Modell (9)	(-2656,372) – (-2646,148)	2	10,224	
Modell (4) vs Modell (10)	(-2665,946) – (-2646,148)	1	19,798	< 0,001
Modell (4) vs Modell (11)	(-2650,068) – (-2646,148)	1	3,920	
Modell (4) vs Modell (12)	(-2658,092) – (-2646,148)	3	11,944	0,004

Variablerna block, kön och parti förbättrar inte förklaringsgraden för någon av modellerna. Det är inte signifikant att modell(1) har en högre förklaringsgrad än modell(3), likaså är det inte heller signifikant att modell(2) har en bättre förklaringsgrad än modell(4). Dessa tre variabler bör således inte vara med i modellen.

Det har därefter testas huruvida de fyra andra variablerna bidrar till förklaringsgraden i respektive modell. Tidigare erfarenhet, kommungrupp och opposition, bidrar signifikant till att öka förklaringsgraden när de inkluderas både till modellen med kategorisk åldersvariabel (modell(3)) och till modellen med kontinuerlig åldersvariabel modell(4). Det är signifikant på 5 procent eller mindre att alla variablerna tillför något till modellerna. Det som skiljer sig åt mellan modellerna är att det är signifikant med 0,1 procent att variabeln kommungrupp bidrar till att förklara överlevnadstiden i modell(3) och på 1 procent i modell (4).

Åldersvariabeln bidrar signifikant till att öka förklaringsgraden både om den inkluderas som kontinuerlig variabel, och om den inkluderas som kategorisk variabel. Signifikansnivån är bättre om den inkluderas som kontinuerlig.

Det kan dock finnas en fördel med att inkludera åldersvariabeln som en kategorisk variabel istället för kontinuerlig variabel. Om det finns skillnader mellan åldersgrupper som inte är linjära så kommer det visa sig då modell (4) skattas.

Tabell (6): Parameterskattningar för modell (3) och modell (4)

	Modell(3)			**Modell(4)**		
Variabel	β (se(β))	χ^2 (p-värde)	Hazard ratio (e^{β})	β (se(β))	χ^2 (p-värde)	Hazard ratio (e^{β})
Ålder_kont	-0.01542 (0,00530)	8.4574 (0.0036)	0.985			
Ålder_kat1				0.93923 (0.37793)	6.1763 (0.0129)	2.558
Ålder_kat2				0.80448 (0.29812)	7.2822 (0.0070)	2.236
Ålder_kat3				0.43463 (0.28425)	2.3379 (0.1263)	1.544
Erf	-0.67930 (0.15372)	19.5287 (<.0001)	0.507	-0.66748 (0.15311)	19.0044 (<.0001)	0.513
Opp	0.26868 (0.13673)	3.8613 (0.0494)	1.308	0.27188 (0.13713)	3.9307 (0.0474)	1.312
Kgrupp 1	-0.59786 (0.17511)	11.6569 (0.0006)	0.550	-0.54476 (0.17525)	9.6624 (0.0019)	0.580
Kgrupp 2	-0.33301 (0.15793)	4.4461 (0.0350)	0.717	-0.28564 (0.15799)	3.2688 (0.0706)	0.752

Samma baslinjer gäller som för modell 1 respektive modell 2.

För den kontinuerliga åldersvariabeln gäller att ett års ökning i ålder minskar risken för avgång med 1,5 procent. Att skillnad existerar är i högsta grad signifikant, med ett p-värde på 0,0036, skillnaden är dock marginell. Men då förändring i drygt en procent kan tyckas lite bör det tas med i beräkning att det gäller också bara för förändring i ledamots ålder med ett år.

För den kategoriska åldersvariabeln jämförs ålderskategorierna 18-24 år, 25-39 år och 40-64 år med baslinjen 65 år och äldre. Risken för avgång är 156 procent högre för dem i ålderkategori 1, 18 tom 24 år, jämfört med dem som är 65 och äldre. Detta är signifikant med ett p-värde på 0,0129. Risken för avgång är 124 procent större för dem i ålderkategori 2, 25 tom 39 år, jämfört med basgruppen. Detta är signifikant med ett p-värde på 0,007. Risken är högre även för ålderskategori 3 jämfört med referensgruppen. Ledamöter i åldersgruppen 40 till och med 65 år har en 54 procents högre hazard. Parameterskattningarna för kategori 3 är dock inte signifikanta.

Skattningarna för den kategoriska variabeln bidrar till att förklara hur förhållandet mellan ålder och avhoppsrisk ser ut. Åldersgruppen 18-24 åringar har en högst skillnad i hazard för avhopp gentemot pensionärerna, åldersgruppen 25-39 något mindre skillnad i avhoppsrisk gentemot basgruppen, och åldersgruppen 40-65 åringarna har minst skillnad. Detta tyder på att skillnad i risken för avhopp utifrån ledamotens ålder verkar vara linjärt.

Modell (3) och modell (4) uppvisar i övrigt inte mycket skillnader vad gäller skattningarna för de andra variablerna. Det är huvudsakligen skattningarna för kommungrupp som skiljer åt mellan modellerna. Risken för avhopp för kommungrupp 1 gentemot kommungrupp 3 är 45 % lägre, och för kommungrupp 2 är den 28 % lägre I modell (3), men I modell (4) 42 % respektive 25 % lägre. Vad som verkligen skiljer åt mellan modellerna är signifikansnivåerna för dessa skattningar. Skattningarna för kommungrupp 1 är signifikanta på 1 % i båda modellerna, p-värdet har förändrats från 0,0006 i modell (3) till 0,0019 I modell (4). Signifikansnivån för kommungrupp 2 har förändrats mer. Med ett p-värde i modell (3) på 0,035 och på 0,0752 i modell (4) förändras signifikansnivån från 5 % till 10 %.

Hazarden samt parameterskattningarna och deras signifikansknivåer är likvärdiga för variablerna erfarenhet och opposition. Erfarenhet har stor inverkan på risken för avhopp i båda modellerna och förändras inte av hur åldersvariabeln är uppbyggd. Risken för avhopp är runt 50 % lägre för ledamöter med tidigare erfarenhet av kommunfullmäktigearbete jämfört med dem utan tidigare erfarenhet i båda modellerna. Det är signifikant på mindre än 0,0001 procent. För variabeln opposition gäller att risken för avhopp är 30 % (modell(3)), eller 31 % (modell (4)) högre för ledamöter som sitter i opposition. Detta är signifikant på 5 %.

Det som huvudsakligen skiljer modellerna åt är således åldersvariabeln. Eftersom sambandet mellan ålder och risk för avhopp i förtid från kommunfullmäktige verkar vara linjär tillför den kontinuerliga åldersvariabeln någon mer information än den kategoriska variabeln. Modell (3) är att föredra framför modell (4).

5.2.3 Cox-modell med tidsberoende kovariat.

För att hantera problemet med eventuell icke-proportionallitet i variabeln opposition skapas en variabel som är en funktion av tiden. En ny modell skattas där den nya variabeln oppvecka är en interaktionsterm för opposition gånger antal veckor.

Modell(13) :

$$h_i(t) = \exp\{\beta_1 \text{ ålder\_kont}_i + \beta_2 \text{k\_grupp}_i + \beta_3 \text{erf}_i + \beta_4 \text{opp}_i + \beta_5 \text{opp}_i * \text{vecka} \} h_0(t)$$

Tabell 7: Modell med opposition som tidsberoende kovariat.

	Modell		
Variabel	B (se(β))	χ^2 (p-värde)	Hazard ratio
Ålder_kont	-0,0155 (0,0053)	8,5548 (0,0034)	0,985
Erf	-0,6670 (0,1537)	18,8357 (<.0001)	0,513
Opp	0,7071 (0,2342)	9,1180 (0,0025)	2,028
Oppvecka	-0,0032 (0,0014)	5,2630 (0,0218)	0,997
Kgrupp 1	-0,6019 (0,1750)	11,8251 (0,0006)	0,548
Kgrupp 2	-0,3222 (0,1580)	4,1597 (0,0414)	0,725

Eftersom interaktionstermen är signifikant tyder detta på att variabeln opposition inte är proportionell vilket också stämmer överens med proportionallitetstestet med schoenfeldresidualerna.

Genom ett enkelt wald-test testas om det är signifikant att gruppernas hazard skiljer sig åt vid en viss tidpunkt. I tabell 8 visas resultaten för huruvida grupperna signifikant skiljer sig åt vid ett antal tidpunkter.

Tabell 8: Wald-test för interaktionsvariabeln

Tidpunkt	χ^2	df	(p-värde)
52 veckor	8.9220	1	0.0028
104 veckor	6.6158	1	0.0101
148 veckor	2.7989	1	0.0943
156 veckor	2.1661	1	0.1411
208 veckor	0.0629	1	0.8019
312 veckor	1.0456	1	0.3065
416 veckor	2.2425	1	0.1343
430 veckor	2.3627	1	0.1243

Det kan bara dras slutsatser utifrån interaktionsvariabeln ungefär fram till och med vecka 148. Från och med vecka 156 visar interaktionsvariabeln inte på någon signifikant skillnad för risken för avhopp mellan att sitta i opposition och att inte göra det.

Koefficienten för interaktionstermen är negativ. Effekten på hazarden av att sitta i opposition minskar därför med tiden. Den ursprungliga variabeln för opposition kan tolkas som variabelns effekt vid studiens början, vid tiden noll. Risken för avhopp från kommunfullmäktige är dubbelt så stor för ledamöter som sitter i opposition jämfört med dem som inte gör det. Efter ett år har hazarden minskat till, exp(0.7071-0.0032*52) = 1,71. Risken för avhopp för en ledamot under opposition är då 71 % högre. Efter två år in i studien är risken för avhopp 45% större för ledamöter som sitter i opposition, och efter knappt tre år är den 26 % högre.

Skattningarna för de övriga variablerna, kommungrupp, erfarenhet och ålder, har inte påverkats av att variabeln opp*vecka inkluderats, de är i stort sätt samma i modell (13) som modell (3).

5.3 Parametrisk modellanpassning

I avsnittet kommer olika parametriska modeller att anpassas till överlevnadsdatat för kommunfullmäktige. Det undersöks om överlevnadstiden, variabeln T, följer någon känd fördelning.

5.3.1 Likelihood ratio test

Datamaterialet anpassas till de parametriska fördelningarna Extended generalizied gamma, Weibull, Lognormal, Ordinary Gamma, Exponential, och log-logistisk. Som nämnts i metoddelen är alla utom log-logistisk specialfall av EGG. Anpassningen för de parametriska fördelningar som ingår i EGG-familjen kan således testas med ett likelihood ratio test. Med likelihood ratio testet jämförs de olika parametriska fördelningarna med EGG.

Eftersom modell (3) har visat sig vara den bättre modellen, så kommer fokus ligga på att anpassa en modell med ålder som kontinuerlig variabel till de parametriska fördelningarna. Endast de signifikanta variablerna är med i de modeller som jämförs nedan. För parameterskattningar med alla variabler (modell (1)) se appendix. I likhet med cox modellen är ålder_kont, tidigare erfarenhet, huruvida ledamoten avgick i opposition eller inte, och kommungrupp signifikanta variabler.

H_0 : restriktionerna för parametrarna i den specifika modellen stämmer.
H_1 : de fria skattningarna för parametrarna är skilt från restriktionerna.

Tabell 9: Likelihoodratiotest.

Jämförelse	**H_0**	**Loglikelihood för modell under H_0**	**Loglikelihood för modell under H_1**	**Chi-square**	**Df**	**P-värde**
Weibull vs EGG	$\gamma = 1$	-673,176	-673,172	0,008	(1)	>0,95
Lognormal vs EGG	$\gamma = 0$	-681,425	-673,172	16,50	(1)	<0,005
Gamma vs EGG	$\lambda = 1$ givet $\gamma > 0$	-673,248	-673,172	0,15	(1)	>0,5
Exponential vs EGG	$\lambda = 1$ givet $\gamma = 1$	-674,476	-673,172	2,61	(2)	>0,25
Exponential vs Weibull	$\lambda = 1$ givet $\gamma = 1$	-674,476	-673,176	2,60	(1)	>0,1

Den Log normal modellen är dåligt anpassad till datamaterialet och kan förkastas 0,5 % signifikansnivå. För de övriga modellerna kan nollhypotesen om ett visst värde för scale, och eller shape parametern inte förkastas. Det kan inte förkastas att datat följer en ordinary Gamma, Weibull eller exponentiell fördelning.

Den log-logistiska modellen kan inte heller uteslutas. Modellen kan inte jämföras med de andra modellerna eftersom den inte är ett specialfall av EGG. Den anpassade log-logistiska modellen har ett Log Likelihood värde på -673,791. Vilket bara är en aning högre än EGG modellen.

5.3.2 Jämförelse mellan parametriska modeller

De fördelningar som passar datamaterialet anpassas med de signifikanta variablerna, modell (3).

Tabell 10: Parameterskattningar för Weibull, Gamma, exponentiell och Log-logistic

Variabel	**EGG β SE(β)**	**Weibull β SE(β)**	**Weibull e^{β}**	**Exp β SE(β)**	**Gamma β SE(β)**	**L- Logistic β SE(β)**
Intercept	5.8417*** (0,3478)	5.8393*** (0,3485)		5.8098*** (0,3514)	5.8512*** (0,3387)	5.5416*** (0,3718)
Ålder_kont	0.0171*** (0,0060)	0.0172*** (0,0059)	1.017	0.0156*** (0,0053)	0.0166*** (0,0057)	0.0189*** (0,0064)
Kgrupp 1	0.6582*** (0,1959)	0.6591*** (0,1963)	1,93	0.6062*** (0,1750)	0.6503*** (0,1908)	0.6520*** (0,2060)
Kgrupp 2	0.3716** (0,1749)	0.3710** (0,1756)	1,45	0.3397** (0,1579)	0.3713** (0,1703)	0.3416* (0,1889)
Erf	0.7535*** (0,1679)	0.7545*** (0,1681)	2,13	0.7152*** (0,1505)	0.7460*** (0,1636)	0.7619*** (0,1759)
Opp	-0.2913* (0,1552)	-0.2948* (0,1516)	0,74	-0.2723** (0,1367)	-0.2764* (0,1464)	-0.3537** (0,1628)
Scale	1.0848	1.1042		1.0000	1.0000	0.9893
Shape	1.0281	1.0000		1.0000	1.1465	.

Signifikansnivåerna för parameterskattningarna anges enligt; *= 10% **= 5% ***=1%.

Skattningarna för weibull modellernas scale och shape parametrar ligger något närmare skattningarna i EGG modellen än vad scale och shape parameterararna i de andra modellerna gör. Weibull modellen, med ett scale-parametervärde över ett, indikerar en avtagande hazard. (se figur 1) Med ett scale-parametervärde på mindre än ett liknar den log-logistiska modellen den log-normala. Den stiger initialt men sjunker sedan mot noll. Även gamma modellen med en shape parameter större än ett indikerar en stigande hazard som sedan planar ut mot värdet på scale parametern.

Parameterskattingarna för de olika variablerna är förhållandevis lika för alla modellerna och så är även signifikansnivåerna för skattningarna. Den log-logistiska modellen har genomgående lite större standardavvikelser än de övriga modellerna och har högre parameterskattningar för erfarenhet och opposition. Den exponentiella modellen har något lägre standardavvikelser och har också något lägre parameterskattingar. Närmast EGG ligger parameterskattningarna för Weibull modellen. Även ordinary-gamma modellens skattningar ligger relativt nära EGG modellen.

Genom att ta e^{β} ges den genomsnittliga överlevnadstiden, vilket gör koefficienterna lättare att tolka. I tabell 9 ges e^{β} för Weibull modellen. Ett års ökning i ålder ökar tiden till avhopp med 1,7 %. För individerna i kommungrupp 1 och 2 är den förväntade överlevnadstiden 93 % respektive 43 % högre än för gruppen med lägst medelinkomst. En kommunledamot med tidigare erfarenhet har en genomsnittlig överlevnad på 113 % jämfört med en ledamot utan tidigare erfarenhet. För de ledamöter som sitter i opposition minskar överlevnadstiden med 26 %.

6. SLUTDISKUSSION OCH VIDARE FORSKNING

Uppsatsens syfte var att hitta förklaringsvariabler till kommunfullmäktigeledamöters avgångar mellan mandatperioderna. Uppsatsens syfte har också varit att anpassa en lämplig modell för materialet. Flera olika variabler och modeller, parametriska och icke-parametriska, har prövats.

Ålder, tidigare erfarenhet, opposition och kommungrupp är de variablerna som har visat sig ha tydlig påverkan på tid till avgång. Ålder har provats både som en kategorisk och som kontinuerlig variabel. Den kontinuerliga åldersvariabeln visade sig fungera bäst då den fångar upp mest av ålderns effekt på tid till avgång eftersom det finns ett linjärt samband. Variablerna parti, block och kön bidrar inte till att förklara tid till avhopp.

6.1 Modellval

Flera modeller har visat sig fungera bra för att skatta kommunfullmäktigeledamöternas överlevnadstid i kommunfullmäktige. Cox proportionella hazard modell samt de parametriska AFT -modellerna Weibull, exponentiella, ordinary gamma modellen och den log-logistiska modellen ger alla relativt bra och likvärdiga estimeringar av variablerna.

Av de parametriska modellerna är den exponentiella och Weibull modellen förhållandevis enkla modeller, vilket kan vara en anledning att välja en av dessa framför ordinary-gamma och log-logistisk.

De har visat sig att skattningarna för scale och shape parametrarna för Weibull modellen ligger något närmare EGG skattningarna än skattningarna för den exponentiella modellen gör. Parameterskattingarna i Weibull modellen ligger också något närmare EGG modellens än vad den exponentiella gör. Weibull modellen följer alltså den mer flexibla EGG modellen bättre och kan enligt detta vara modellen att föredra. Det som talar för den exponentiella modellen är att den dels har lägre medelfel och dels att den enligt log-likelihood testet inte går att förkasta till förmån för Weibull.

Fördelen med att använda en parametrisk metod framför cox modellen är att i parametriska modeller behöver inte proportionalitetsantagandet gälla. Då det finns indikationer på att oppositionsvariabeln eventuellt äventyrar proportionalitetsantagandet kan det finnas fördelar med att använda en parametrisk modell. Problemet med icke-proportionallitet i Cox-modellen kan också lösas med att inkludera en interaktionsterm av/med tiden. En sådan modell, modell (13), skattades. Nackdelen med denna modell är att den är mer svårtolkad.

Undersökningen mynnar ut i tre möjliga modeller.

Weibull-fördelning.

$\text{Log } T_i = 5{,}84 + 0{,}017 \text{ ålder\_kont}_i + 0{,}658 \text{ k\_grupp1}_i + 0{,}372 \text{ k\_grupp2}_i + 0{,}754 \text{ erf}_i - 0{,}291 \text{ opp}_i$

Exponentiell-fördelning.

$\text{Log } T_i = 5{,}810 + 0{,}016 \text{ ålder\_kont}_i + 0{,}606 \text{ k\_grupp1}_i + 0{,}340 \text{ k\_grupp2}_i + 0{,}715 \text{ erf}_i - 0{,}273 \text{ opp}_i$

Cox-model med opposition som tidsberoende kovariat.

$h_i(t) = \exp\{-0{,}016 \text{ ålder\_kont}_i - 0{,}602 \text{ k\_grupp1}_i - 0{,}322 \text{ k\_grupp2}_i - 0{,}667 \text{ erf}_i + 0{,}707 \text{ opp}_i -$

$0{,}003 \text{ opp}_i\text{*vecka } \}h_0(t)$

6.2 Variabler

Enligt den här studien har ledamöter från kommuner i Stockholms län med högre medelinkomst lägre risk för avhopp än ledamöter från kommuner med lägre medelinkomst.

Det som har störst effekt för om ledamoten hoppar av sitt uppdrag i förtid är om han eller hon har någon tidigare erfarenhet. I till exempel artikeln ”En av tio har slutat” (svenska dagbladet) framställs teorin att många (unga) väljer att hoppa av kommunfullmäktige innan uppdraget är slutfört av anledningen att de inte hade uppdragets innebörd klart för sig. I artikeln anges att skäl till avhopp till exempel kan vara just bristen på erfarenhet av hur uppdraget fungerar. Resultaten i den här studien stödjer den teorin.

Det har också visat sig att ju äldre ledamoten är, desto mindre är risken att han eller hon väljer att lämna sitt uppdrag i förtid. Detta överrensstämmer med resultaten i valmyndighetens rapport. Det är intressant att notera att detta resultat går tvärtemot resultaten från Ahlbäck Öberg m.fl. undersökning av avhopp från riksdagen. Enligt den undersökningen har riksdagsledamöter med högre ålder en större tendens att hoppa av från riksdagen innan mandatperioden är slut. (Ahlbäck Öberg m.fl. 2007).

6.3 Vidare forskning

Erfarenhet från minst en tidigare mandatperiod påverkar alltså mycket om man sitter mandatperioden ut och är bland annat ett mått på kunskap om uppdragets natur. Å ena sidan ger vår undersökning stöd för den tesen, men å andra sidan finns det risk att variabeln inte mäter just denna kunskap om uppdragets natur. Erfarenhet från andra nämnder i kommunen har inte räknats med som tidigare erfarenhet. Har man suttit i en annan nämnd eller haft liknande uppdrag finns förmodligen kunskap om hur politiska uppdrag fungerar.
Vidare så mäter variabeln inte heller kvantiteten av erfarenhet, ledamoten kan ha suttit en tidigare mandatperiod, eller, som fallet verkar vara för många ledamöter betydligt fler perioder.

Det enda vi kan konstatera är alltså att, om en ledamot satt i kommunfullmäktige mandatperioden innan så är sannolikheten att han eller hon inte hoppar av sitt uppdrag i förtid betydligt större. En intressant vidareutveckling av undersökningen skulle vara att modellera olika typer av erfarenheter av längre eller kortare politiskt arbete.

Effekten av kommungrupp kan komma från andra egenskaper än just medelinkomst. Till exempel följer andelen valdeltagande i kommunerna till viss del samma mönster som medelinkomst. En annan egenskap som den här den här kommuniondelningen fångar upp är om kommunens styre är borgligt eller socialdemokratiskt styrt. En intressant vidareutveckling av studien skulle vara att jämföra kommuner utifrån fler egenskaper än medelinkomst.

En mycket intressant vidareutveckling som den här uppsatsen inte modellerar är anledningen till avhopp. En modell med konkurrerande risker skulle till exempel kunna undersöka om anledningen till avhopp är frivillig, eller om den ligger utanför ledamotens kontroll.

8 Källförteckning

Tryckta källor:

Ahlbäck Öberg, S. Hermansson, J. Wängnerud, L. (2007). *Exit riksdagen*, första upplagan. Malmö: Liber.

Allison, P. D. (2010). *Survival Analysis Using SAS® : A practical Guide*, Andra upplagan. Gary, NC: SAS institute Inc.

Collett, D. (2003). *Modelling Survival Data in Medical Research*, Andra upplagan. London: Chapman & Hall.

Cox, D. R. (1975). Partial Likelihood. *Biometrica Trust*, Vol. 62, nr. 2, 269-276.

Guo, S. (2010). *Survival Analysis*, New York: Oxford University Press,cop.

Klein, J. P. Melvin L. Moeschberger (2003). *Techniques for Censored and Truncated Data*, Andra upplagan, Springer science + Business Media , LLC.

Kleinbaum, D. G. Mitchel Klein (2005). *Survival Analysis, A Self-Learning Text*, Andra upplagan.

Schoenfeld, D. A. (1982). Partial residuals for the proportional hazard regression model. *Biometrika*, vol. 69, nr. 1, 239-241.

Elektroniska källor:

SCB: Medelinkomst. (2011-03-28)
http://www.ssd.scb.se/databaser/makro/start.asp?SSD=Sveriges+statistiska+databaser&xu=C9233001&yp=tansss&lang=1&SSD=Sveriges+statistiska+databaser

Svenska Dagbladet. (2011-04-15)
http://www.svd.se/nyheter/stockholm/en-av-tio-har-slutat_1795733.svd

Valmyndigheten. (2011-04-29)
http://www.val.se/om_oss/rapporter/avgangna_redovisning2011_tillUS.pdf

9. Appendix

Appendix A – Variabler

Kommungrupp – Indelning av kommuner utifrån tre stratum baserad på medelinkomst för både män och kvinnor 2007.

K_grupp1= Högst medelinkomst, Danderyd, Lidingö och Nacka
K_grupp2= Mellan medelinkomst, Järfälla, Solna och Sigtuna
K_grupp3=Lägst medelinkomst, Botkyrka, Södertälje och Haninge

Parti

parti1= Moderaterna
parti 2= Folkpartiet
parti 3= Centerpartiet
parti 4= Kristdemokraterna
parti 5= Socialdemokraterna
parti 6= Vänsterpartiet
parti 7= Miljöpartiet
parti 8= Övriga partier

Block

block1= Högerblocket (M,Fp,Kd,C,Övriga)
block 2= Vänsterblocket (S,V,Mp)

Ålder kontinuerlig – Indelat efter ålder vid inträde i studien

Ålder kategorisk

Ålder_kat1 = 18 – 24
Ålder_kat2 = 25 – 39
Ålder_kat3 = 40 – 64
Ålder_kat4 = 65 och äldre

Tidigare erfarenhet – Ledamöternas tidigare erfarenhet som ledamöter i kommunfullmäktige innan inträdet i studien

Erf = 1 om Ledamoten har tidigare erfarenhet
Erf= 0 om ej tidigare erfarenhet

Kön

Kön = 1 om ledamoten är man
Kön = 0 om ledamoten är kvinna

Utträde under opposition – Huruvida ledamoten har avgått under mandatperiod då denne var i opposition eller ej

Opp = 1 om ledamoten avgått under opposition
Opp = 0 om ledamoten inte suttit i opposition när han/hon avgått

Appendix B
Stratumindelning efter sammanräknad förvärvsinkomst 2007

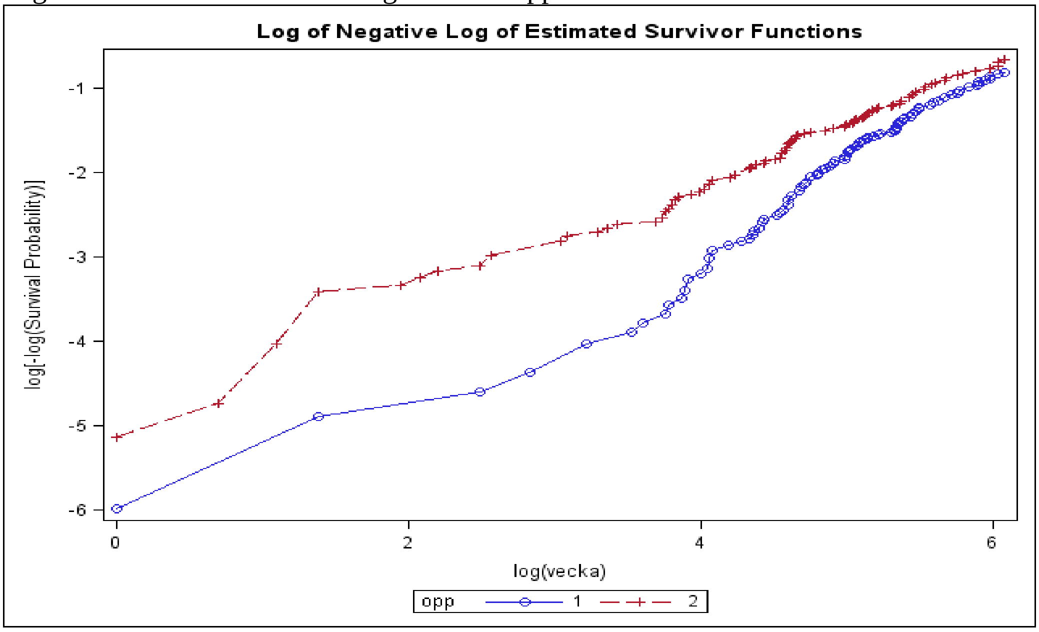
Danderyd	409
Lidingö	342
Täby	324
Nacka	299
Sollentuna	293
Vaxholm	291
Ekerö	282
Värmdö	265
Österåker	264
Nykvarn	263
Tyresö	262
Stockholm	259
Vallentuna	259
Salem	257
Solna	252
Järfälla	251
Upplands Väsby	242
Sundbyberg	239
Huddinge	238
Sigtuna	234
Upplands-Bro	233
Haninge	225
Nynäshamn	221
Norrtälje	212
Södertälje	210
Botkyrka	202

Appendix C - Majoriteter i kommunfullmäktige

	Mandatperiod		
Kommun	2002-2006	2006-2010	2010-2014
Danderyd	M+KD	M+KD	M+KD+C
Lidingö	M+FP	M+FP	M+FP+KD
Nacka	M+FP+KD	M+FP+KD	M+FP+KD+C
Solna	M+FP+KD	M+FP+KD	M+FP+KD+C
Järfälla	S+V+MP	M+FP+KD+C	M+FP+KD+C
Sigtuna	S+V+MP	S+MP+C	S+MP+C
Haninge	S+FP	M+FP+KD+C+MP	M+FP+KD+C+MP
Södertälje	S+V+MP	S+V+MP	S+V+MP
Botkyrka	S+V+MP	S+V+MP	S+V+MP

Appendix D – Visuellt proportionalitetstest.

Log kumulativa hazarden mot log tiden för opposition.



Den övre kurvan visar ledamöter som sitter i opposition och den undre kurvan visar de som inte sitter i opposition.

Appendix E - Parametriska modeller med samtliga variabler.

Variabel	EGG β SE(β)	Weibull β SE(β)	Exp β SE(β)	Gamma β SE(β)	L- Logistic β SE(β)
Intercept	6,0226*** (0,8567)	6,0221*** (0,8651)	5,9645*** (0,7869)	6,0324*** (0,8381)	5,7112*** (0,9473)
Ålder_kont	0,0170*** (0,0060)	0,0172*** (0,0059)	0,0157*** (0,0053)	0,0166*** (0,0057)	0,0188*** (0,0064)
Kgrupp1	0,6829*** (0,2007)	0,6842*** (0,2018)	0,3546*** (0,1813)	0,6780*** (0,1969)	0.6781 *** (0.2115)
Kgrupp2	0,3859** (0,1755)	0,3847** (0,1769)	0,3546** (0,1599)	0,3856** (0,1721)	0,3553* (0,1906)
Kgrupp3	.	.	.	.	.
Erf	0,7319*** (0,1677)	0,7335*** (0,1684)	0,6973*** (0,1516)	0,7269*** (0,1643)	0,7369*** (0,1769)
Opp	-0,3123* (0,1640)	-0,3193** (0,1604)	-0,2955** (0,1454)	-0,3002* (0,1554)	-0,3833** (0,1711)
Kön	-0,0494 (0,1511)	-0,0480 (0,1522)	-0,0425 (0,1385)	-0,0517 (0,1481)	-0,0176 (0,1637)
Block	-0,1763 (0,4463)	-0,1760 (0,4480)	-0,1634 (0,4078)	-0,1759 (0,4343)	-0,1818 (0,4883)
Parti1	-0,0600 (0,3505)	-0,0623 (0,3539)	-0,0518 (0,3223)	-0,0558 (0,3426)	-0,0827 (0,3866)
Parti2	0,2748 (0,3911)	0,2765 (0,3943)	0,2605 (0,3590)	0,2712 (0,3836)	0,2789 (0,4233)
Parti3	0,6666 (0,6368)	0,6627 (0,6371)	0,6112 (0,5792)	0,6685 (0,6326)	0,6398 (0,6450)
Parti4	0,1563 (0,4456)	0,1563 (0,4493)	0,1473 (0,4092)	0,1559 (0,4375)	0,1407 (0,4819)
Parti5	0,3640 (0,3449)	0,3638 (0,3481)	0,3536 (0,3169)	0,3639 (0,3383)	0,3695 (0,3750)
Parti6	0,1981 (0,4311)	0,1945 (0,4344)	0,1814 (0,3955)	0,2023 (0,4225)	0,1481 (0,4708)
Parti7	.	.	.	.	.
Parti8	.	.	.	.	.
Scale	1,0629	1,0982	1,00	1,00	0,9835
Shape	1,0518	1,00	1,00	1,1409	.

Signifikansnivåerna för parameterskattningarna anges enligt; *= 10% **= 5% ***=1%.