



Stockholms  
universitet

# Prissättning av delkaskoförsäkring - en metodjämförelse

Tove Brickner

Masteruppsats 2013:7  
Försäkringsmatematik  
Oktober 2013

[www.math.su.se](http://www.math.su.se)

Matematisk statistik  
Matematiska institutionen  
Stockholms universitet  
106 91 Stockholm



Matematisk statistik  
Stockholms universitet  
Masteruppsats 2013:7  
<http://www.math.su.se/matstat>

# Prissättning av delkaskoförsäkring - en metodjämförelse

Tove Brickner\*

Oktober 2013

## Sammanfattning

I den här undersökningen kommer olika metoder för prissättningen av så kallad delkaskoförsäkring att undersökas. Delkaskoförsäkring är en fordonsförsäkring som täcker ett flertal olika typer av skador, t.ex. glasskador och stöld. Att försäkringen täcker flera olika typer av skador ger möjlighet till att beräkna tariffen på olika sätt, genom att ta hänsyn till att det är olika typer av skador eller inte. Enligt litteraturen bör prissättningen för en försäkring som täcker olika skador baseras på beräkningar för varje skadetyper för sig, vilket också stämmer överens med de antaganden som ligger till grund för tariffberäkningar. Målsättningen med undersökningen är att utreda om det alltid är bättre att genomföra tariffberäkningarna per skadetyper, och i så fall, är skillnaden så stor att det är värt det extra arbetet. Ett delmål är att finna lämpliga sätt att kunna jämföra de olika tarifferna. Resultaten tyder på att den bästa tariffen för delkaskoförsäkring är en tariff där varje skadetyper har analyserats för sig. I slutsatser diskuteras även för och nackdelar med en sådan tariff.

---

\*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.  
E-post: [tovebrickner@gmail.com](mailto:tovebrickner@gmail.com). Handledare: Andreas Norvall Lagerås.

## **Abstract**

This is a study of pricing methods of partial cover insurance, which is a motor insurance covering several different claim types, e.g. glass claims and theft. Since the insurance cover several different types of claims it is possible to calculate the tariff in several ways, whether the fact that the claims are of different types are taken into account or not. Literature in the field suggests that it is more accurate to analyse the different claim types separately, due to assumptions regarding the claims distribution. The goal of this study is to analyse if it always is better to calculate the tariff based on different claim types, and if it is, determine if the difference between the tariffs is large enough to justify the extra effort. A sub goal of the study is to find suitable methods to validate the different tariffs. The result indicates that the best tariff for the partial cover insurance is a tariff based on analyses of different claim types. Pros and cons for the tariff based on different claim types are discussed in the conclusion.

## **Förord**

Det här är ett examensarbete i försäkringsmatematik som omfattar 30 hp. Idén till examensarbetet är min egen och arbetet har till stor del genomförts under min föräldraledighet. Jag har haft två handledare, docent Andreas Nordvall Lagerås och min kollega på If, Bengt Eriksson. Jag vill tacka mina handledare för alla värdefulla tips.

# Innehåll

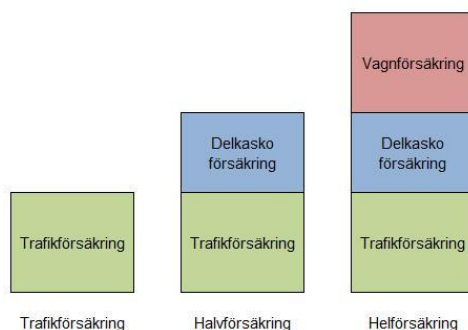
<b>1</b>	<b>Inledning</b>	<b>4</b>
1.1	Bakgrund . . . . .	4
1.2	Målsättning . . . . .	5
1.3	Metod . . . . .	5
<b>2</b>	<b>Generaliserade linjära modeller, GLM</b>	<b>6</b>
2.1	Definition GLM . . . . .	6
2.2	Begrepp . . . . .	7
2.3	Modellantaganden . . . . .	8
2.4	En multiplikativ modell . . . . .	8
2.5	Standard GLM . . . . .	9
2.5.1	Modell för skadefrekvens . . . . .	9
2.5.2	Modell för medelskadan . . . . .	9
2.6	Jungs marginalsummemetod . . . . .	10
<b>3</b>	<b>Tarifferna</b>	<b>11</b>
3.1	Data . . . . .	11
3.2	Tariffargument . . . . .	11
3.3	Tarifferna enligt Standard GLM . . . . .	12
3.3.1	Enkel S-GLM . . . . .	12
3.3.2	Komplex S-GLM . . . . .	13
3.4	Tariffen enligt Jungs metod . . . . .	14
<b>4</b>	<b>Tariffvalidering</b>	<b>16</b>
4.1	Valideringsdata . . . . .	16
4.2	Valideringsmetoder . . . . .	17
4.2.1	Goodness of Fit . . . . .	17
4.2.2	Stabilitetstest . . . . .	17
4.2.3	Kvotjämförelse . . . . .	17
4.2.4	Ginikoefficient . . . . .	19
<b>5</b>	<b>Resultat</b>	<b>22</b>
5.1	Jämförelse av komplex S-GLM och enkel S-GLM . . . . .	23
5.1.1	Validering med stabilitetstest . . . . .	24

5.1.2	Validering med kvotjämförelse . . . . .	27
5.1.3	Validering med Ginikoefficient . . . . .	27
5.2	Jämförelse av enkel S-GLM och Jung . . . . .	29
<b>6</b>	<b>Slutsatser</b>	<b>30</b>
6.1	Slutsatser av tariffjämförelsen . . . . .	30
6.2	Slutsatser av valideringsmetoder . . . . .	31
6.3	Fortsatta studier . . . . .	31
<b>A</b>	<b>Figurer, validering per premienivå</b>	<b>33</b>
<b>B</b>	<b>Figurer, validering med ginikoefficient</b>	<b>35</b>

# Kapitel 1

## Inledning

Bilförsäkring kan tecknas som en trafik-, halv- eller helförsäkring. Trafikförsäkringen täcker personskador samt skador på annans egendom och är obligatorisk enligt lag. Halv- och helförsäkring innehåller trafik- och delkaskoförsäkring men helförsäkringen innehåller också en vagnskadeförsäkring, se figur 1.1. Den här studien behandlar prissättningen av delkaskoförsäkring. De skador som täcks av delkaskoförsäkringen är stöldskador, maskinskador, brandskador, glasskador samt räddning.



Figur 1.1: Översikt av bilförsäkring

### 1.1 Bakgrund

Kunder betalar en försäkringspremie till försäkringsbolaget för att vid en eventuell skada på fordonet få ersättning. Från försäkringsbolagets sida måste den totala premieintäkten täcka den totala skadekostnaden. Olika kunder utgör olika stor risk för ett försäkringsbolag. Kunder grupperas därför efter egenskaper som kan påverka risken. För en hårt konkurrensutsatt produkt som bilförsäkring är det väsentligt för ett försäkringsbolag att kunna beräkna olika kundgruppers förväntade skadekostnad, och därmed kunna bestämma



en korrekt premienivå. Fel prissättning leder till moturval, d.v.s. att de lönsamma kunderna tycker att de har fått en för hög premie och vänder sig till en konkurrent, medan de olönsamma kunderna lockas av de låga premierna. Moturval leder till att premierna inte kommer att täcka de framtida skadekostnaderna och försäkringsbolaget blir olönsamt.

I den här undersökningen ska premien, eller den så kallade tariffen, beräknas för delkaskoförsäkringen. Att försäkringen täcker flera olika typer av skador ger möjlighet till att beräkna tariffen på olika sätt, genom att ta hänsyn till att det är olika skadetyper eller inte. Enligt tidigare studier [1, 2] bör prissättningen för en försäkring som täcker olika skador baseras på beräkningar för varje skadetyper för sig, vilket också stämmer översens med de antaganden som ligger till grund för tariffberäkningar. Dock är sådana beräkningar mer tidskrävande.

## 1.2 Målsättning

Målsättningen med undersökningen är att utreda om det alltid är bättre att genomföra tariffberäkningarna per skadetyper, och i så fall, är skillnaden så stor att det är värt det extra arbetet. Ett delmål är att finna lämpliga sätt att kunna jämföra de olika tarifferna.

## 1.3 Metod

Försäkrings- och skadedata från If Skadeförsäkring AB (If) ligger till grund för undersökningen. Tarifferna kommer att beräknas med hjälp av generaliserade linjära modeller, GLM. När tarifferna är beräknade ska några olika valideringsmetoder testas för att komma fram till vilken metod som ger den bästa tariffen, med hänsyn både till resultat och komplexitet.

## Kapitel 2

# Generaliserade linjära modeller, GLM

### 2.1 Definition GLM

Generaliserade linjära modeller (GLM) är en generalisering av den linjära regressionsmodellen till regressionsmodeller som inte har normalfördelade resonsvariabler och där en funktion av responsvariabelns medelvärde är linjär i förklarande variabler. Det är tre saker som identifierar en GLM [3]:

1. Responsvariabeln  $Y$ , med oberoende observationer  $y_1, \dots, y_N$  följer en fördelning som är en exponentiell dispersionsmodell (EDM). Sannolikhetsfunktionen eller täthetsfunktionen för en EDM är:

$$f_{y_i}(y_i; \theta_i, \phi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi/\omega_i} + c(y_i, \phi, \omega_i)\right)$$

Parametern  $\theta_i$  beror av  $i$  medan spridningsparametern  $\phi$  är densamma för alla  $i$ . Kumulantfunktionen  $b(\theta_i)$  antas vara två gånger deriverbar med inverterbar andraderivata. Givet varje val av  $b(\cdot)$  så är fördelningen bestämd av parametrarna  $\theta_i$  och  $\phi$ . Funktionen  $c(\cdot, \cdot, \cdot)$  beror inte på  $\theta_i$  och är därför inte av något stort intresse i GLM teori.

2. Den systematiska komponenten i en GLM relaterar vektorn  $(\eta_1, \dots, \eta_N)$  till de förklarande variablerna genom en linjär modell. Om  $x_{ij}$  betecknar värdet av variabel  $j$  för observation  $i$  så är

$$\eta_i = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, N.$$

3. Den tredje komponenten i en GLM är länkfunktionen. Om  $\mu_i = E[Y_i]$  så gäller att  $\eta_i = g(\mu_i)$  där länkfunktionen  $g$  är monoton och deriver-

bar. Så  $g$  länkar väntevärdet till de förklarande variablerna via formeln

$$g(\mu_i) = \sum_j \beta_j x_{ij}, \quad i = 1, \dots, N.$$

Regressionen med normalfördelade responsvariabler uppfyller detta då normalfördelningen är en EDM. Om  $Y_i \sim N(\mu_i, \sigma^2/\omega_i)$  så är täthetsfunktionen

$$f_{y_i}(y_i) = \exp\left(\frac{y_i \mu_i - \mu_i^2/2}{\sigma^2/\omega_i} + c(y_i, \sigma^2, \omega_i)\right)$$

som är en EDM med  $\theta_i = \mu_i$ ,  $\phi = \sigma^2$  och  $b(\theta_i) = \theta_i^2/2$ . Identitetslänken  $g(\mu_i) = \mu_i$  specificerar den linjära regressionen med normalfördelade väntevärden.

Detta var den generella definitionen, fortsättningsvis följer en beskrivning av hur GLM används för prissättning av försäkringar.

## 2.2 Begrepp

**Durationen** är tiden som försäkringen har gällt räknat i år. Ett försäkringsavtal som gällt i 6 månader bidrar således med 0,5 till durationen.

**Skadefrekvens** är antalet skador dividerat med durationen.

**Medelskada** är den totala skadekostnaden dividerat med antalet skador.

**Riskpremien** är den totala skadekostnaden dividerat med durationen. Riskpremien är produkten av skadefrekvens och medelskada. Riskpremien är det belopp som försäkringstagaren förväntas kosta försäkringsbolaget, exklusive omkostnader.

I tabell 2.1 finns en sammanfattning av de nyckeltal som kommer att beräknas med GLM-modeller [4].

Exponering $\omega$	Respons $X$	Nyckeltal $Y = X/\omega$
Duration	Antal skador	Skadefrekvens
Antal skador	Skadekostnad	Medelskada
Duration	Skadekostnad	Riskpremie

Tabell 2.1: Nyckeltal

**Tariffargument** är benämningen på en förklarande variabel i GLM-modellerna. Exempel på tariffargument är bilmärke och ålder. Alla försäkringskontrakt som har samma klassindelning för alla tariffargument sägs

tillhöra samma **Tariffcell**. I en tariff med endast 2 tariffargument, ålder och bilmärke tillhör alla försäkringskontrakt där föraren är 30 år och bilen är en Volvo samma tariffcell, och kommer därmed få samma premie.

## 2.3 Modellantaganden

1. **Avtalsberoende**,  $n$  är antalet försäkringskontrakt. Om  $X_i$  är respons för kontrakt  $i$ , så är  $X_1, \dots, X_n$  oberoende.
2. **Tidsberoende**,  $n$  är antalet disjunkta tidsintervall. Om  $X_i$  är respons i tidsintervall  $i$ , så är  $X_1, \dots, X_n$  oberoende.
3. **Homogenitet**, Antag 2 försäkringskontrakt i samma tariffcell med lika exponering. Om  $X_i$  är respons för kontrakt  $i$ , så har  $X_1$  och  $X_2$  samma fördelning.

Dessa antaganden ligger till grund för GLM-modellerna [4]. Dock finns det situationer då dessa modellantaganden inte uppfylls. Seriekrockar strider mot avtalsberoende och säsongeffekter såsom att risken för stenscott ökar precis efter att snön har smält strider mot tidsberoende. Homogenitet uppfylls sällan. Inom motorförsäkring finns vanligtvis ett bonussystem som ska kompensera för det. Bonusen bygger på erfarenheter av den enskilde föraren. Avvikelserna från antagandena leder i praktiken inte till några problem för analysen.

## 2.4 En multiplikativ modell

Låt  $Y_{ij}$  beteckna den förväntade riskpremien i tariffcell  $(i,j)$  och  $E[Y_{ij}] = \mu_{ij}$ . Då beskrivs den multiplikativa modellen för två tariffargument enligt

$$\mu_{ij} = \gamma_0 \gamma_{1i} \gamma_{2j}.$$

Modellen tolkas som att den förväntade riskpremien i tariffcell  $(i,j)$   $\mu_{ij}$  är lika med en baspremie  $\gamma_0$  multiplicerat med effekten av tariffargument 1 och 2,  $\gamma_{1i}$  och  $\gamma_{2j}$ .  $\gamma_0$  kan bestämmas till riskpremien i tariffcell (1,1) och då följer det att  $\gamma_{11} = \gamma_{21} = 1$ . Om  $\gamma_{21} = 1,00$  och  $\gamma_{22} = 1,08$  har alla i klass 2, en riskpremie som är 8% högre än de tillhörande klass 1 i tariffargument 2 oavsett vilken klass av tariffargument 1 de tillhör. För en modell med endast två tariffargument är det möjligt att skatta riskpremien,  $\mu_{ij}$  med  $Y_{ij} = X_{ij}/\omega_{ij}$  om varje tariffargument endast har några få klasser. Vanligtvis är det ett flertal tariffargument som inkluderas i tariffen och antalet tariffceller blir stort. I de tarifferna som kommer att beräknas för delkaskoförsäkringen kommer 7 tariffargument som innehåller mellan 2 och 6 klasser vardera inkluderas. Totalt blir det 18000 tariffceller. Att beräkna

en premie för varje tariffcell genom att använda observerat utfall är ingen lämplig metod för en tariff med 18000 tariffceller då det med största säkerhet saknas exponering i ett större antal av tariffcellerna.

## 2.5 Standard GLM

Riskpremien är produkten av skadefrekvens och medelskada. Riskpremien kan därför beräknas enligt en metod som bygger på två modeller, en för skadefrekvens och en för medelskada [4]. Skadefrekvensen antas följa en Poissonfördelning och skadebelopp en gammalfördelning. Metoden kallas här Standard GLM.

### 2.5.1 Modell för skadefrekvens

Antalet skador för varje enskilt försäkringskontrakt antas följa en Poissonfördelning. Väntevärdet för antalet skador i en tariffcell,  $E[X_i] = \omega_i \mu_i$  om  $\mu_i$  betecknar väntevärdet då  $\omega_i = 1$ .

$$P(X_i = x_i) = e^{-\omega_i \mu_i} \frac{\omega_i \mu_i^{x_i}}{x_i!}, \quad x_i = 0, 1, 2, \dots$$

Det är främst skadefrekvensen  $Y_i = X_i/\omega_i$  som är av intresse.  $Y_i$  sägs ha en relativ Poissonfördelning [4] då den inte antar heltalsvärden.

$$P(Y_i = y_i) = P(X_i = \omega_i y_i) = e^{-\omega_i \mu_i} \frac{\omega_i \mu_i^{\omega_i y_i}}{\omega_i y_i!}$$

Fördelningen är en EDM med  $\theta_i = \log(\mu_i)$ ,  $b(\theta_i) = e^{\theta_i}$ ,  $\phi = 1$  och  $c(y_i, \omega_i) = \omega_i y_i \log(\omega_i) - \log(\omega_i y_i!)$ . Länkfunktionen  $g(\mu_i) = \log(\mu_i)$  så  $\log(\mu_i) = \sum_j \beta_j x_{ij}$ . I en tariff men 2 tariffargument och  $i = 4$  tariffceller motsvaras  $\beta_j$ ,  $j = 1, 2, 3$  av  $\log(\gamma_0)$ ,  $\log(\gamma_{12})$ ,  $\log(\gamma_{22})$  och  $x_{ij}$  antar värdet 1 eller 0.

### 2.5.2 Modell för medelskadan

Skadekostnaden för en enskild skada antas följa en gammalfördelning. I fallet då  $\omega_i = 1$  är  $X_i \sim G(\alpha, \beta_i)$  med täthetsfunktionen

$$f(x_i) = \frac{\beta_i^\alpha}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta_i x_i}, \quad x_i > 0$$

Om  $X_i$  är en summa av  $\omega_i$  oberoende gammavariabler så gäller det att  $X_i \sim G(\omega_i \alpha, \beta_i)$ . Medelskadan  $Y_i = X_i/\omega_i$  har då täthetsfunktionen

$$f_{Y_i}(y_i) = \omega_i f_{X_i}(\omega_i y_i) = \frac{(\omega_i \beta_i)^{\omega_i \alpha}}{\Gamma(\omega_i \alpha)} y_i^{\omega_i \alpha - 1} e^{-\omega_i \beta_i y_i}$$

som är en  $G(\omega_i \alpha, \omega_i \beta_i)$

Detta är en EDM [4] då  $\mu_i = \alpha/\beta_i$  och  $\phi = 1/\alpha$ , vilket ger  $\theta_i = -1/\mu_i$  och  $b(\theta_i) = -\log(-\theta_i)$ .

## 2.6 Jungs marginalsummemetod

Svensken Jan Jung [5] introducerade marginalsummemetoden för tariffberäkning. Modellen bygger på ett antagande om att det inte finns tillräckligt med data i varje tariffcell, men att det gör det i marginalerna, d.v.s. per klassindelning för tariffargumenten. I Jungs metod är skattningar alltid väntevärdesriktiga. I modellen med endast 2 tariffargument erhålls följande ekvationer:

$$\sum_j E[X_{ij}] = \sum_j x_{ij}; \quad i = 1, \dots, m_1$$

$$\sum_i E[X_{ij}] = \sum_i x_{ij}; \quad j = 1, \dots, m_2$$

som med den multiplikativa modell blir

$$\sum_j \omega_{ij} \gamma_0 \gamma_1 \gamma_2^j = \sum_j \omega_{ij} y_{ij}; \quad i = 1, \dots, m_1$$

$$\sum_i \omega_{ij} \gamma_0 \gamma_1 \gamma_2^j = \sum_i \omega_{ij} y_{ij}; \quad j = 1, \dots, m_2$$

Ekvationerna löses numeriskt. Jungs marginalsummemetod tillämpas direkt för riskpremien.

# Kapitel 3

## Tarifferna

### 3.1 Data

Till grund för tariffberäkningarna ligger försäkringsinformation och skadehistorik från åren 2007 till 2011. Försäkringsinformationen innehåller variabler såsom gällande försäkringsperiod, information om försäkringstagaren, t.ex. ålder och adress, samt information om fordonet, t.ex. klassning, årsmo-  
dell och märke. Skadehistoriken innehåller all information om de inträffade skadorna såsom datum, typ av skada och utbetalda belopp. Försäkringar som inte har fullständig information eller inte gällt i mer än ett par dagar har tagits bort från beräkningsunderlaget. Skadekostnaden har inflationsju-  
sterats med 2% årligen från år 2007. Totalt består data av cirka 1 200 000 försäkringsår och 130 000 skador.

### 3.2 Tariffargument

I tabell 3.1 visas en sammanfattning av de tariffargument som kommer att ingå i tariffanalyserna. Totalt ger detta  $2 \cdot 5 \cdot 5 \cdot 6 \cdot 2 \cdot 6 \cdot 5 = 18000$  möjliga tariffceller. Körsträckeklasserna 1–5 motsvarar en av försäkringstagaren uppskattad körsträcka där klass 1 är den kortaste och klass 5 den längsta körsträckan. Område är en geografisk indelning av landet där område 1 är minst riskfylld och område 5 innehåller de delar av landet som är mest riskfyllda inom försäkringsgrenen. Exempelvis ingår Stockholmsregionen i område 5. Bilklassning är en gruppering av fordonen där klass 1 motsvarar lägre förväntad skadekostnad än övriga klasser. Egenskaper som inkluderas i bilklassningen är exempelvis priset på reservdelar och hur kraftig motor bilen har.

Nr	Tariffargument	Klasser	Antal klasser
1	Import	Ja / Nej	2
2	Körsträcka	1–5	5
3	Område	1–5	5
4	Bilklassning	1–6	6
5	Garage	Ja / Nej	2
6	Förarålder	0–29 år, 30–39 år, 40–49 år, 50–59 år, 60–69 år, 70–99 år	6
7	Bilålder	0–2 år, 3–6 år, 7–10 år, 11–15 år, 16– år	5

Tabell 3.1: Översikt av tariffargumenten

### 3.3 Tarifferna enligt Standard GLM

Relationstalen för skadefrekvens,  $f_i$ , och medelskada,  $m_i$ , kommer att skattas enligt Standard GLM (S-GLM). Skattningarna av relationstalen och dess konfidensintervall beräknas med hjälp av funktionen PROC GENMOD i programvaran SAS [6].

#### 3.3.1 Enkel S-GLM

Enkel S-GLM betecknar den tariff som baseras på skattningarna utifrån all data. Premien beräknas enligt:

$$Premie = B \cdot \prod_{i=1}^k f_i \cdot \prod_{i=1}^k m_i$$

där  $B$  betecknar baspremien och  $f_i$  och  $m_i$  betecknar relationstalen för frekvens respektive medelskada för de  $k = 7$  tariffargumenten.  $B$  motsvarande  $\gamma_0$  och  $f_1 \cdot m_1$  motsvarar  $\gamma_{1i}$  i kapitel 2.4. Alla sju tariffargument är signifikanta och kommer att inkluderas i tariffen. Bilklassningen har stor betydelse både för skadefrekvens och medelskada. Förarålder och område påverkar skadefrekvensen i större utsträckning än medelskadan. Körstäckeklass har motsatta effekter på skadefrekvens och medelskada. Skadefrekvensen ökar med körsträckan, medan medelskadan minskar. I tabell 3.2 visas relationstalen  $f_i$  samt ett konfidensintervall för två av sju tariffargument.



Tariffargument	Klass	Undre gräns	Relationstal	Övre gräns
Område	1	0,587	0,592	0,606
Område	2	0,690	0,703	0,716
Område	3	0,844	0,861	0,878
Område	4	0,895	0,912	0,929
Område	5	1,000	1,000	1,000
Bilklassning	1	0,342	0,369	0,398
Bilklassning	2	0,494	0,509	0,526
Bilklassning	3	0,601	0,619	0,638
Bilklassning	4	0,752	0,775	0,798
Bilklassning	5	0,871	0,899	0,929
Bilklassning	6	1,000	1,000	1,000

Tabell 3.2: Relationstal för skadefrekvens

### 3.3.2 Komplex S-GLM

Komplex S-GLM betecknar tariffen som baseras på olika skattningar av relationstalen beroende på skadetyper. Tariffen beräknas som en summa av sex deltariffer, en för varje skadetyper, och premien beräknas således enligt:

$$Premie = \sum_{j=1}^6 \left( B_j \cdot \prod_{i=1}^{k_j} f_{ij} \cdot \prod_{i=1}^{k_j} m_{ij} \right)$$

$B_j$  betecknar baspremien för skadetyper  $j$ ,  $f_{ij}$  och  $m_{ij}$  är relationstalen för skadefrekvens och medelskada för tariffargument  $i$  inom skadetyper  $j$ . Alla tariffargumenten är inte signifikanta för alla deltariffer. För varje deltariff undersöks vilka tariffargument som ska inkluderas. Detta blir tidskrävande då det är totalt 12 stycken analyser, en för skadefrekvens och en för medelskada, och detta för alla sex skadetyper, som ska bedömmas. Att exkludera tariffargument som inte är signifikanta borde påverka stabiliteten i tariffen positivt. I tabell 3.3 presenteras vilka tariffargument som ingår i respektive deltariff.

Tariff	Import	Körsträcka	Område	Bilklassning	Garage	Förarålder	Bilålder
Glasbyte	F M	F	F M	F M	F M	F	F M
Stöld	F M		F M	F M	F M	F M	F M
Maskin	F M		F	F M	F	F	F M
Brand	F M		F	F M	F	F M	F M
Räddning	F	F	M	F M		F	F
Glasreparation	F	F	F	F		F	F

Tabell 3.3: Tariffargument i tariffen S-GLM komplex

Enligt tabell 3.2 har bilklassningen stor betydelse för att prediktera skadefrekvensen i delkaskoförsäkringen. Bilklassningen har även betydelse för skadefrekvensen i varje deltariff, men i olika stor utsträckning. Exempelvis så är det större skillnad i stöldfrekvens beroende på bilklassningen jämfört med glasbytesfrekvens, se tabell 3.4 och 3.5. Det är endast skadefrekvensen

för glas och räddning som beror av körsträckan. Att medelskadan i tariffen enligt enkel S-GLM sjunker med körsträcka förklaras av att det inträffar fler glas och räddningsskador, som jämförelsevis är billiga skador, i de längre körsträckeklasserna. Område har betydelse för alla skadefrekvenser utom räddning. I tabell 3.6 redovisas relationstalen för räddningsfrekvenserna. Konfidensintervallen för samtliga områden inkluderar 1, d.v.s. att alla områden kan använda samma relationstal. Konfidensintervallen för respektive deltariff blir bredare än för tariffen enligt enkel S-GLM. De mest osäkra skattningarna har brandtariffen då endast en dryg procent av alla skador är brandskador.

Tariffargument	Klass	Undre gräns	Relationstal	Övre gräns
Bilklassning	1	0,158	0,195	0,240
Bilklassning	2	0,343	0,376	0,413
Bilklassning	3	0,415	0,453	0,495
Bilklassning	4	0,628	0,684	0,745
Bilklassning	5	0,766	0,841	0,922
Bilklassning	6	1,000	1,000	1,000

Tabell 3.4: Relationstal för skadefrekvens stöld

Tariffargument	Klass	Undre gräns	Relationstal	Övre gräns
Bilklassning	1	0,366	0,411	0,461
Bilklassning	2	0,570	0,598	0,627
Bilklassning	3	0,696	0,728	0,762
Bilklassning	4	0,825	0,862	0,901
Bilklassning	5	0,899	0,944	0,991
Bilklassning	6	1,000	1,000	1,000

Tabell 3.5: Relationstal för skadefrekvens glasbyte

Tariffargument	Klass	Undre gräns	Relationstal	Övre gräns
Område	1	0,922	0,987	1,057
Område	2	0,979	1,040	1,106
Område	3	0,904	0,966	1,031
Område	4	0,886	0,944	1,006
Område	5	1,000	1,000	1,000

Tabell 3.6: Relationstal för skadefrekvens räddning

### 3.4 Tariffen enligt Jungs metod

Den här studien är i första hand en jämförelse mellan enkel S-GLM och komplex S-GLM. Men under vissa omständigheter kan det vara intressant att även jämföra med en tariff enligt Jungs metod och därför beräknas även en sådan tariff. Relationstalen kommer att skattas med SAS PROC GENMOD

och premien beräknas enligt:

$$Premie = B \cdot \prod_{i=1}^k r_i$$

där  $r_i$  betecknar relationstalen för riskpremien för tariffargument  $i$ . Alla sju tariffargument kommer att ingå i tariffen enligt Jungs metod.

## Kapitel 4

# Tariffvalidering

### 4.1 Valideringsdata

För att validera tarifferna bör ett dataset som inte varit underlag för tariffanalyserna användas. Modellen kan vara väl anpassad till data men det som är relevant är att modellen passar till framtida data, d.v.s. okänt data. Valideringsdata kan vara ett slumpat dataset från den totala datamängden eller så kan data från den senaste tidsperioden användas som valideringsdata.

I den här studien kommer båda varianterna att testas. För att validera tarifferna på slumpmässigt data har den totala datamängden slumpats i fem delar. Fyra delar används som underlag till tarifferna, så kallat träningsdata, och en del som valideringsdata. Det upprepas fem gånger enligt en typ av korsvalidering [7] där varje del används till validering exakt en gång. Figur 4.1 illustrerar hur den totala datamängden har delats upp i träningsdata och valideringsdata för de fem grupperna. I en annan validering kommer

Grupp1	V	T	T	T	T
Grupp2	T	V	T	T	T
Grupp3	T	T	V	T	T
Grupp4	T	T	T	V	T
Grupp5	T	T	T	T	V

Figur 4.1: Valideringsdata, V och träningsdata, T för de 5 grupperna.

det senaste året, här år 2011, att exkluderas från tariffmodelleringen och användas som valideringsdata.

## 4.2 Valideringsmetoder

### 4.2.1 Goodness of Fit

Pearsons  $\chi^2$  [4] är ett välkänt goodness-of-fit mått inom regressionsteori. Detta mått jämför observerade data med skattningarna från den skattade modellen. Pearsons  $\chi^2$  (oskalad) definieras enligt

$$\chi^2 = \sum_i \frac{(y_i - \mu_i)^2}{\sigma_i^2}$$

där  $\mu_i$  respektive  $\sigma_i$  är väntevärde och standardavvikelse enligt modellantagandet. I modeller för skadefrekvens har en Poissonfördelning antagits och därmed är väntevärdet och varians lika och i modeller för medelskadan har en gammafördelning antagits och då är variansen proportionell mot kvadraten av väntevärdet. Pearsons  $\chi^2$  är approximativt  $\chi^2(k)$  fördelad där antalet frihetsgrader,  $k$ , är antalet tariffceller minus antalet skattade parametrar. De klassiska goodness-of-fit-måtten fokuserar vanligtvis på analyser med en svarsvariabel. Det innebär att det inte går att jämföra tariffmodeller enligt standard GLM. För Jungs metod är Pearsons  $\chi^2$  odefinierad.

### 4.2.2 Stabilitetstest

Eftersom en tariff ska prediktera framtida skadekostnader är det viktigt att parametrarna som inkluderas är tidskonsistenta. För att testa om relationstalen förhåller sig på samma sätt över tid kan analyserna göras på delmängder av data som är indelat efter tidsperioder, exempelvis riskår.

En annan form av stabilitetstest är att testa hur premierna förändras beroende på vilket data som ligger till grund för tariffberäkningarna. Eftersom data slumpats i fem delar, där fyra delar används som tariffunderlag kommer fem tariffer att beräknas. Då är det intressant att beräkna den genomsnittliga premieskillnaden per försäkringskontrakt beroende på vilket data som används som tariffunderlag. Hur valet av träningsdata och valideringsdata påverkar tarifferna har inte påträffats i någon artikel. Exempelvis kan den genomsnittliga premieskillnaden per försäkringskontrakt beroende på om träningsdata från grupp 1 eller grupp 2 används som tariffunderlag beräknas.

### 4.2.3 Kvotjämförelse

Inom försäkring är det stora variationer i data och målet med en modell är inte att prediktera de observerade skadorna, utan att modellera noggranna skattningar av de förväntade värdena på skadorna. Goodness-of-fit kan vara svårt att tolka när det gäller en tariff med tusentals tariffceller. För att

förstå innebörden av en alternativ tariff kan en valideringsmetod, som i den här undersökningen benämns kvotjämförelse, vara mer lämpligt. Kvotjämförelsen har använts på If och en liknande jämförelser beskrivs i [8].

Första steget för en kvotjämförelse är att beräkna kvoten mellan två tariffer. Den ena tariffen skulle kunna vara en gällande tariff, och den andra en alternativ tariff. Kvoten, eller relationen, beräknas för varje kontrakt enligt

$$R_i = \frac{P_i^A}{P_i^G},$$

där  $P_i^G$  betecknar den gällande premien för försäkringskontrakt  $i$  och  $P_i^A$  den alternativa premien för försäkringskontrakt  $i$ . I nästa steg grupperas kvoterna enligt större eller mindre än 1. Avslutningsvis studeras den normerade skadekostnadsprocenten både för den gällande och den alternativa tariffen, se exempel i tabell 4.1. Den normerade skadekostnadsprocenten definieras som skadekostnaden dividerat med premieintäkten givet att total premieintäkt är lika med total skadekostnad.

	Alternativ tariff	Gällande tariff
Kvot mindre än 1	102%	90%
Kvot större än 1	98%	110%

Tabell 4.1: Kvotjämförelse för 2 tariffer

De kontrakt som får en sänkt premie om den gällande tariffen ersätts med alternativet ingår i gruppen kvot mindre än 1, och de kontrakt som får en högre premie ingår i gruppen kvot större än 1. Då kan tabell 4.1 tolkas som att en grupp med sänkning respektive höjningsbehov har identifierats och skadekostnadsprocenten i gruppen med sänkingsbehov går från 90% till 102% och gruppen med höjningsbehov från 110% till 98%. Den alternativa tariffen har värden närmre 100% och är därför en bättre tariff. Resultatet i tabell 4.1 beräknas enligt tabell 4.2 där  $S_i$  betecknar skadekostnad för försäkringskontrakt  $i$ .

	Alternativ tariff	Gällande tariff
Kvot mindre än 1	$\frac{\sum_{i:R_i < 1} S_i}{\sum_{i:R_i < 1} P_i^A}$	$\frac{\sum_{i:R_i < 1} S_i}{\sum_{i:R_i < 1} P_i^G}$
Kvot större än 1	$\frac{\sum_{i:R_i > 1} S_i}{\sum_{i:R_i > 1} P_i^A}$	$\frac{\sum_{i:R_i > 1} S_i}{\sum_{i:R_i > 1} P_i^G}$

Tabell 4.2: Beräkningar för kvotjämförelse

Det finns ytterligare en tolkning av kvotjämförelsen. I en perfekt konkurrenssituation där kunder alltid väljer att teckna försäkring i det försäkringsbolag som ger den billigaste premien innebär det att om ett konkurrerande försäkringsbolag prissätter enligt den alternativa tariffen så kommer konkurrenten att vara mer lönsam, se figur 4.3.

	Alternativ tariff	Gällande tariff
Normerad Skk%	102%	110%

Tabell 4.3: Tariffjämförelse vid perfekt konkurrens

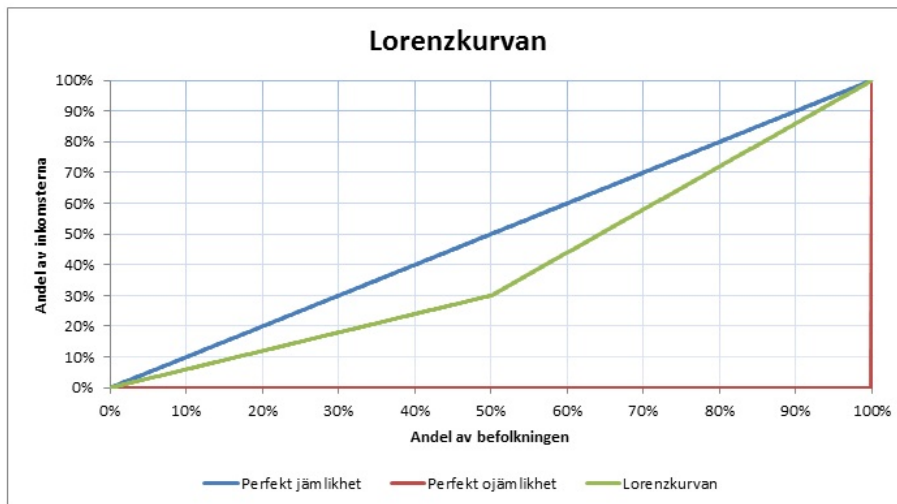
Resultaten av kvotjämförelserna kommer presenteras enligt tabell 4.3 som ger ett konkret mått på vilken tariff som är den bästa.

#### 4.2.4 Ginikoefficient

Ginikoefficienten [9] är ett ekonomiskt mått på ojämlikhet. Ett användningsområde är att mäta inkomstfördelningen hos en befolkning. För att beräkna Ginikoefficienten behövs den så kallade Lorenzkurvan. Lorenzkurvans [10] y-värde givet ett x-värde visar hur stor andel av de totala inkomsterna som intjänas av en viss andel av den fattigaste befolkningen. I figur 4.2 står den 50% fattigaste delen av befolkningen för 30% av de totala inkomsterna. Ginikoefficienten definieras som två gånger arean mellan perfekt jämlikhet och Lorenzkurvan. Ginikoefficienten antar värden mellan 0% och 100% där 0% är total jämlikhet och 100% är total ojämlikhet.

En Ginikoefficient kan även användas för att mäta olikheter mellan premier och skador. Försäkringskontrakten motsvarar befolkningen och skadorna motsvarar inkomsterna. Om  $P_i^G$  betecknar den gällande premien för försäkringskontrakt  $i$  och  $P_i^A$  betecknar den alternativa premien för försäkringskontrakt  $i$  så definieras relationen  $R_i$  enligt

$$R_i = \frac{P_i^A}{P_i^G}.$$



Figur 4.2: Lorenzkurvan

För att beräkna Ginikoefficienten rangordnas försäkringskontrakten enligt stigande relationstal så att  $R_1 \leq R_2 \leq \dots \leq R_n$ . Den empiriska ordnade Lorenzkurvan [11] är då given av  $(a_0 = 0, b_0 = 0), (a_1, b_1), \dots, (a_n = 1, b_n = 1)$  för  $n$  försäkringskontrakt. Om  $S_i$  betecknar skadekostnad för kontrakt  $i$  så är

$$(a_i, b_i) = \left( \frac{\sum_{k=1}^i P_k^G}{\sum_{k=1}^n P_k^G}, \frac{\sum_{k=1}^i S_k}{\sum_{k=1}^n S_k} \right).$$

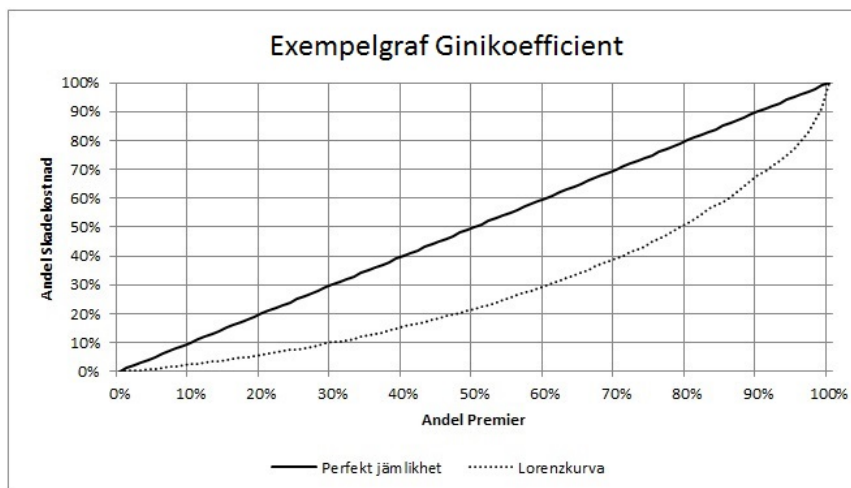
Den empiriska Ginikoefficienten beräknas enligt

$$\begin{aligned} \widehat{Gini} &= 2 \sum_{j=0}^{n-1} (a_{j+1} - a_j) \left( \frac{a_{j+1} + a_j}{2} - \frac{b_{j+1} + b_j}{2} \right) \\ &= 1 - \sum_{j=0}^{n-1} (a_{j+1} - a_j)(b_{j+1} + b_j), \end{aligned}$$

där  $a_j$  är den empiriska fördelningsfunktionen för de gällande premierna,  $\hat{F}_p(R_j)$  och  $b_j$  är den empiriska fördelningsfunktionen för skadorna,  $\hat{F}_s(R_j)$

I exemplet i figur 4.3 har en konstant gällande premie jämförts med en av de beräknade tarifferna som alternativ premie. Figur 4.3 visar att 30% av de försäkringskontrakt som har lägst alternativ premie endast står för 10% av skadorna. En hög Ginikoefficient säger att den alternativa premien har stor spridning jämfört med den gällande premien. Ginikoefficienten i det här exemplet är 44% vilket tyder på att en stor del av variationen går förlorad om inte alternativet väljs. I den här studien kommer Ginikoefficienten vara en jämförelse mellan komplex S-GLM och enkel S-GLM.





Figur 4.3: Ginikoefficienten för försäkring

## Kapitel 5

# Resultat

I första hand kommer jämförelser mellan komplex S-GLM och enkel S-GLM att presenteras, men även en jämförelse mellan enkel S-GLM och Jung kommer att visas. Resultat från de olika tariffvalideringarna kommer att presenteras dels för de fem slumpmässiga grupperna, där 80% av data används som tariffunderlag och 20% används som valideringsmaterial, samt dels för tariffen där åren 2007-2010 använts som tariffunderlag och år 2011 som valideringsmaterial.

Ett beräkningsmoment som ingår i tariffen enligt komplex S-GLM är att fördela den totala premien mellan de olika skadetyperna. Om inte alla skadetyper har samma skadefrekvensutveckling och medelskadautveckling över tid blir fördelningen inte korrekt för kommande år. De olika skadetyperna som ingår i delkaskoförsäkringen utvecklas olika över tid, t.ex. ökar andelen maskinskadador medan andelen stöldskador minskar, se tabell 5.1. För den här jämförelsen betyder det att fördelningen mellan skadetyper kommer att vara olika för validering på slumpmässigt data, som inte är tidsberoende, och validering på senaste data, år 2011.

Skadetyper	2007	2008	2009	2010	2011	Medel
Glasbyte	31,2%	29,2%	33,8%	33,4%	35,3%	32,5%
Stöld	39,8%	37,4%	28,5%	25,1%	23,6%	31,1%
Maskin	13,3%	13,9%	19,9%	24,5%	24,3%	19,0%
Brand	8,3%	12,1%	10,1%	9,2%	8,9%	9,8%
Räddning	3,2%	3,2%	3,6%	3,9%	3,6%	3,5%
Glasreparation	3,4%	3,7%	3,6%	3,6%	4,1%	3,7%

Tabell 5.1: Skadekostnadsandel per skadetyper

## 5.1 Jämförelse av komplex S-GLM och enkel S-GLM

Syftet med försäkring är att risken ska delas mellan försäkringstagare, så att följa upp tariffen genom att på kontraktsnivå undersöka premieintäkt mot skadekostnad är ingen bra idé. För att undersöka tariffens kvalitet måste skadekostnad och premieintäkt jämföras på en aggregerad nivå. Pearsons  $\chi^2$  mäter avvikelsen mellan observerad och predikerad skadefrekvens och medelskada på tariffcellsnivå. Då tarifferna innehåller ett flertal tariffargument blir antalet tariffceller stort och många tariffceller har lite eller ingen exponering. Pearsons  $\chi^2$  kan därför bli ett svårtolkat mått. Om Pearsons  $\chi^2$  hade varit det bästa måttet borde skattningarna av relationstalen anpassas så att Pearsons  $\chi^2$  minimeras.

I tabell 5.2 och 5.3 presenteras skadekostnadsprocenten för två av de sju tariffargumenten för en av de fem grupperna. Genom att jämföra den normalade skadekostnadsprocenten för klasserna inom ett och två tariffargument är det svårt att säga något om huruvida den ena tariffen är bättre än den andra. Det finns en antydning till att komplex S-GLM predikterar skadekostnaden något bättre än enkel S-GLM. Exempelvis ligger skadekostnadsprocenten närmre 100% för komplex S-GLM där exponeringen är som störst, bilklassning klass 3 och ålder 50-69 år. Å andra sidan predikterar enkel S-GLM skadekostnaden bättre för bilklassning klass 6.

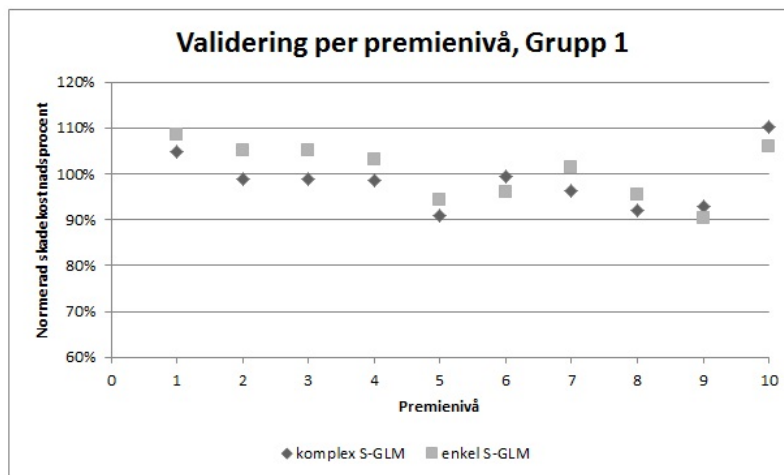
Klasser	00-29 år	30-39 år	40-49 år	50-59 år	60-69 år	70-99 år	Totalt
1	191,0%	64,6%	98,3%	93,2%	134,6%	63,7%	100,3%
2	85,3%	78,5%	89,6%	91,8%	92,2%	86,8%	88,3%
3	62,6%	96,9%	98,4%	<b>130,8%</b>	<b>110,5%</b>	99,7%	105,2%
4	96,1%	102,5%	82,2%	97,8%	96,9%	90,1%	93,6%
5	138,6%	124,1%	96,5%	88,5%	86,5%	121,7%	101,5%
6	193,9%	194,3%	82,8%	75,1%	101,1%	62,8%	<b>113,9%</b>
Totalt	106,1%	117,3%	89,2%	100,4%	99,3%	93,4%	100,0%

Tabell 5.2: Skadekostnadsprocent per ålder och bilklassning för enkel S-GLM

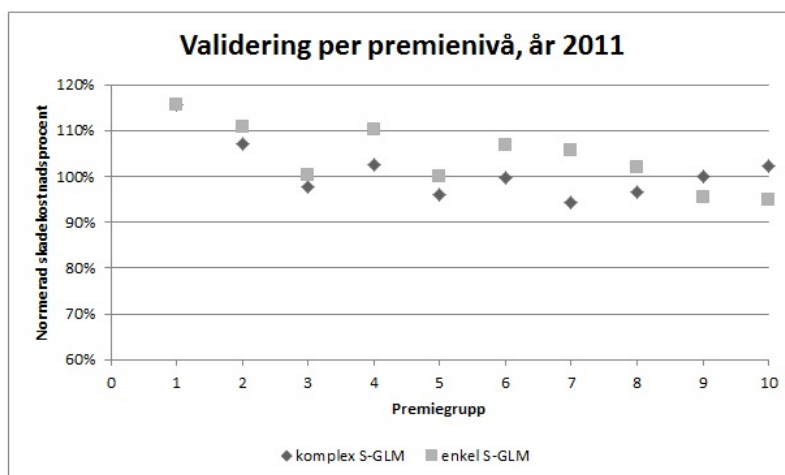
Klasser	00-29 år	30-39 år	40-49 år	50-59 år	60-69 år	70-99 år	Totalt
1	160,3%	62,7%	103,5%	88,3%	136,4%	65,4%	99,8%
2	81,9%	78,6%	94,1%	83,6%	86,1%	82,8%	85,3%
3	62,1%	98,4%	103,9%	<b>119,2%</b>	<b>102,4%</b>	95,1%	102,4%
4	86,4%	101,5%	86,3%	89,9%	91,7%	85,2%	91,0%
5	119,7%	124,9%	105,6%	85,5%	90,2%	122,3%	103,6%
6	159,4%	210,3%	102,7%	84,2%	126,1%	75,0%	<b>130,7%</b>
Totalt	96,7%	119,0%	96,9%	95,2%	97,0%	90,4%	100,0%

Tabell 5.3: Skadekostnadsprocent per ålder och bilklassning för komplex S-GLM

En annan indelning av data som ger en uppfattning av tariffernas anpassning visas i figur 5.1. På x-axeln visas premienivån sorterade enligt respektive tariff sådan att 10% av exponeringen med lägst premie tillhör premienivå 1 och de 10% som har de högsta premierna tillhör premienivå 10. Gemensamt för båda tariffen i grupp 1 är att både de högsta men framförallt de lägsta premierna är för låga. Det kan tyda på att modellantaganden som multiplikativitet inte är uppfyllda. För grupp 1 finns en liten tendens att tariffen S-GLM komplex har en normerad skadekostnadsprocent som håller sig närmast 100%. Det samma gäller även för år 2011, se figur 5.2. Motsvarande figurer för grupperna 2, 3, 4 och 5 redovisas i bilaga A.



Figur 5.1: Skadekostnadsprocent per premienivå i grupp 1



Figur 5.2: Skadekostnadsprocent per premienivå i grupp år 2011

### 5.1.1 Validering med stabilitetstest

Det är viktigt att kontrollera att de skattade relationstalen är stabila över tid. Nedan följer två exempel, tabell 5.4 och 5.5, på att relationstal som ligger till grund för beräkningarna av den komplexa tariffen är stabila över tid. Om relationstalen för ett tariffargument inte är stabila över tid är de inte lämpliga för att prediktera den framtida skadekostnaden.

Körsträcka	2007	2008	2009	2010	2011
1	0,85	0,85	0,85	0,84	0,82
2	1,05	1,05	1,03	1,05	1,07
3	1,24	1,24	1,27	1,28	1,28
4	1,56	1,57	1,48	1,51	1,63
5	1,83	1,85	2,08	2,09	2,19

Tabell 5.4: Relationstalen per körsträckeklass för antal glasskador

Garage	2007	2008	2009	2010	2011
Nej	1,19	1,29	1,24	1,26	1,23
Ja	0,81	0,71	0,79	0,78	0,82

Tabell 5.5: Relationstalen för garage för antal stölskador

Ett annat mått för stabiliteten är premieomfördelningen. Premieomfördelningen definieras här som den absoluta relativa premieskillnaden per försäkringskontrakt mellan två tariffer. Tabell 5.6 visar premieomfördelningen mellan grupperna för tarifferna enkel S-GLM och komplex S-GLM. Exempelvis skiljer det i genomsnitt 5,9% på premien mellan grupp 1 och grupp

2 enligt enkel S-GLM medan det endast skiljer 2,8% i samma jämförelse för komplex S-GLM. Alla grupper har alltid 75% gemensamt träningsdata med övriga grupper. Stabilitetstesten tyder på att tariffen enligt komplex S-GLM är stabilare, vilket skulle kunna förklaras av att tariffargument som bara bidrar med brus har exkluderats i varje deltariff.

Grupp	1	2	3	4	5
1	0,0%	5,9%	5,7%	6,3%	5,9%
2	5,9%	0,0%	4,9%	5,4%	5,1%
3	5,7%	4,9%	0,0%	3,4%	4,1%
4	6,3%	5,4%	3,4%	0,0%	4,2%
5	5,9%	5,1%	4,1%	4,2%	0,0%

(a) Omfördelning enkel S-GLM

Grupp	1	2	3	4	5
1	0,0%	2,8%	3,1%	3,9%	2,8%
2	2,8%	0,0%	2,9%	4,0%	3,2%
3	3,1%	2,9%	0,0%	4,0%	2,9%
4	3,9%	4,0%	4,0%	0,0%	3,9%
5	2,8%	3,2%	2,9%	3,9%	0,0%

(b) Omfördelning komplex S-GLM

Tabell 5.6: Omfördelning mellan grupper för respektive tariff

I genomsnitt är premieomfördelningen beroende på grupp 5,1% för enkel S-GLM och 3,4% för komplex S-GLM. Komplex S-GLM visas alltså vara stabilare än enkel S-GLM trots att enkel S-GLM innehåller många färre parametrar.

Omfördelningen kan även beräknas mellan de två tarifferna. I tabell 5.7 redovisas den genomsnittliga omfördelningen mellan de två tarifferna. I de flesta fall så får en försäkring drygt 10% skillnad i premie beroende på vilken metod som används vid tariffberäkningarna. Störst skillnad mellan enkel S-GLM och komplex S-GLM är det för år 2011.

	Premieomfördelning
Grupp 1	10,7%
Grupp 2	10,6%
Grupp 3	11,5%
Grupp 4	11,1%
Grupp 5	9,9%
År 2011	12,9%

Tabell 5.7: Genomsnitt av omfördelning mellan tarifferna enkel S-GLM och komplex S-GLM

### 5.1.2 Validering med kvotjämförelse

Det är den alternativa premien enligt kapitel 4.2.3 som representeras av komplex S-GLM och den gällande premien av enkel S-GLM. Tabell 5.8 visar kvotjämförelsen för grupp 1. De försäkringskontrakt som får en lägre premie om tariffen bestäms enligt komplex S-GLM jämfört enkel S-GLM går från en skadekostnadsprocent på 94,8% för enkel S-GLM till en skadekostnadsprocent på 105,5% för komplex S-GLM och det som får en högre premie givet komplex S-GLM jämfört enkel S-GLM går från 107,0% till 94,1%. Tabellen kan tolkas som att en grupp med premiehöjnings- respektive sänkingsbehov har identifierats och att justeringen sedan blivit för stor. Detta är ett fenomen som inträffar då tarifferna är ungefär lika. En liten fördel visas dock för komplex S-GLM då den normerade skadkostnadsprocenten är närmre 100% för komplex S-GLM.

	komplex S-GLM	enkel S-GLM
Kvot mindre än 1	<b>105,5%</b>	94,8%
Kvot större än 1	94,1%	<b>107,0%</b>

Tabell 5.8: Kvotjämförelse i grupp 1

Givet att ett försäkringsbolag prissätter enligt komplex S-GLM och ett annat enligt enkel S-GLM skulle, allt annat lika, försäkringsbolaget med tariffen enligt komplex S-GLM ha en skadekostnadsprocent på 105,5% och försäkringsbolaget med tariffen enligt enkel S-GLM 107,0%, då alla försäkringstagare väljer det billigaste bolaget. Motsvarande jämförelse för alla valideringsunderlag presenteras i tabell 5.9. Komplex S-GLM visas vara den bästa tariffen i alla fall, förutom i grupp 5. Resultaten varierar beroende på vilket data som slumpmässigt väljs ut som valideringsmaterial, detta visar på vikten av att inte enbart testa tariffen på ett valideringsmaterial.

	komplex S-GLM	enkel S-GLM
Grupp 1	<b>105,5%</b>	<b>107,0%</b>
Grupp 2	102,4%	111,1%
Grupp 3	105,5%	108,2%
Grupp 4	103,2%	110,4%
Grupp 5	108,3%	104,0%
År 2011	105,4%	110,0%

Tabell 5.9: Kvotjämförelse, komplex S-GLM jämfört enkel S-GLM

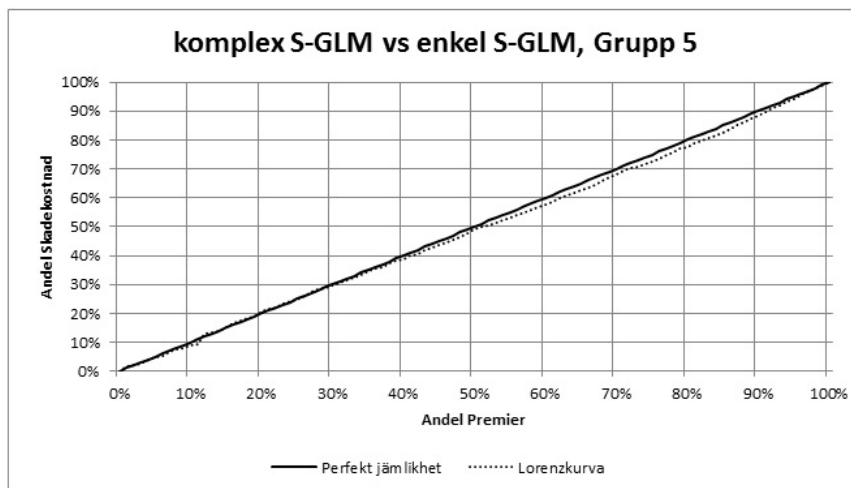
### 5.1.3 Validering med Ginikoefficient

Precis som för kvotjämförelsen är den alternativa premien enligt kapitel 4.2.4 som representeras av komplex S-GLM och den gällande premien av enkel S-

GLM. På x-axeln är det den ackumulerade premien för enkel S-GLM som visas, sorterad efter relationerna  $R_i$  i stigande ordning. På y-axeln visas motsvarande ackumulerande skadekostnad. Relationerna  $R_i$  är här alternativet komplex S-GLM dividerat med enkel S-GLM. I tabell 5.10 presenteras ginikoefficienterna för samtliga valideringsunderlag och figur 5.3 och 5.4 visar Lorenzkurvan för minsta respektive största ginikoefficienten. Enligt figur 5.3 finns det ingen anledning att ersätta den gällande tariffen, enkel S-GLM, med alternativet, komplex S-GLM, då fördelningarna är lika. Enligt figur 5.4 finns en märkbar skillnad mellan fördelningarna och alternativet komplex S-GLM bör väljas. Motsvarande figurer för grupp 1, 2, 3 och 4 redovisas i bilaga B.

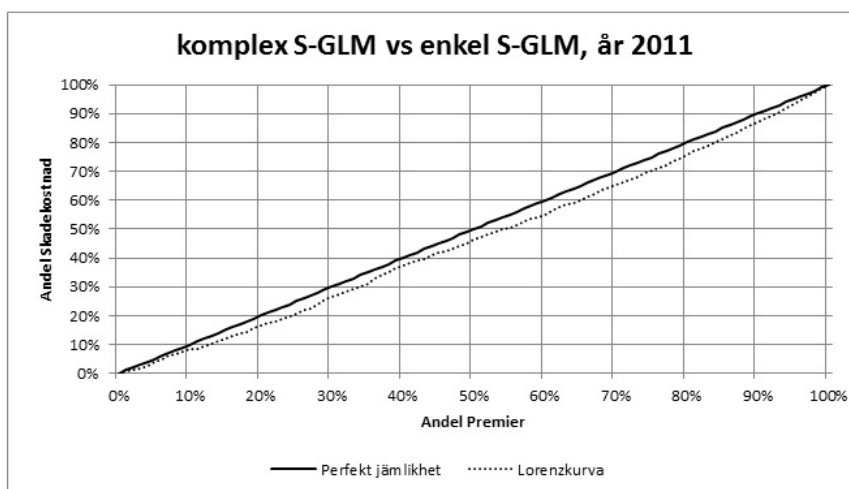
	Ginikoefficient
Grupp 1	2,5%
Grupp 2	5,1%
Grupp 3	5,2%
Grupp 4	6,0%
Grupp 5	2,1%
År 2011	6,8%

Tabell 5.10: Ginikoefficienter



Figur 5.3: Lorenzkurvan grupp 5





Figur 5.4: Lorenzkurvan år 2011

## 5.2 Jämförelse av enkel S-GLM och Jung

Enligt jämförelsen av komplex S-GLM och enkel S-GLM tyder resultaten på att komplex S-GLM ger en bättre tariff. Om tariffberäkningarna av någon anledning ändå inte skulle genomföras enligt komplex S-GLM, exempelvis beroende på tidsbrist eller att en multiplikativ tariff eftersträvas, är det intressant att undersöka om det finns någon anledning att beräkna skadefrekvens och medelskada separat. Fördelningsantaganden för skadefrekvens och medelskada uppfylls inte p.g.a. de olika skadetyperna.

Enligt [12] är enkel S-GLM bättre än Jung, men i den här undersökningen råder andra förutsättningar för jämförelsen mellan enkel S-GLM och Jung. I tabell 5.11 presenterar kvotjämförelsen mellan enkel S-GLM och Jung. Jung visar sig vara minst lika bra som enkel S-GLM. Alltså, om inte uppdelningen per skadetyg görs, behöver inte heller skadefrekvens och medelskada modelleras separat.

	enkel S-GLM	Jung
Grupp 1	106,7%	100,0%
Grupp 2	102,9%	103,2%
Grupp 3	102,8%	104,5%
Grupp 4	106,2%	100,6%
Grupp 5	110,2%	94,7%
År 2011	105,9%	100,5%

Tabell 5.11: Kvotjämförelse, enkel S-GLM jämfört Jung

# Kapitel 6

## Slutsatser

### 6.1 Slutsatser av tariffjämförelsen

Alla tariffjämförelser, oavsett valideringsmetod, tyder på att den bästa tariffen för delkaskoförsäkringen är en tariff enligt komplex S-GLM, d.v.s. en tariff som är beräknad utifrån modeller av skadefrekvens och medelskada per skadetyper. Stabilitetstest visar att resultaten av tariffjämförelserna varierar kraftigt beroende på vilken del av data som använts som träningsdata respektive valideringsdata, detta trots att det i den här studien användes ett tariffunderlag som motsvarade 5 års statistik av personbilsförsäkringar hos ett av landets största försäkringsbolag.

#### Fördelar med en tariff enligt komplex S-GLM

1. Ger enligt resultaten en bättre tariff.
2. Förståelse för vad som driver risken, både för skadefrekvens och medelskada per skadetyper.
3. Möjlighet att kalibrera tariffen på skadetypsnivå, t.ex. om det annonseras höjda reparationskostnader för glasskador är det enkelt att justera baspremien för glasskador.

#### Nackdelar med en tariff enligt komplex S-GLM

1. Tidskrävande, både för tariffkonstruktionen och underhåll.
2. Många och komplexa beräkningar t.ex. för att avgöra signifikans av tariffargument och för att bestämma alla baspremier.
3. Premien blir svår att förklara för kunden då tariffen inte är multiplikativ. Exempelvis kan effekterna av en flytt innebära olika premieförändringar för kundens två bilar, i värsta fall kan en bli billigare och en dyrare.

4. Det förekommer att tariffargument är givna på förhand. Exempelvis görs den branchgemensamma bilklassningen för hela delkasko. Om tariffen delas upp per skadetyper blir det svårt att normera för detta.

För- och nackdelarna måste vägas mot varandra för att avgöra om det är motiverat att beräkna en tariff utifrån modeller av skadefrekvens och medelskada per skadetyper. Den här undersökningen visar att det är motiverat resultatmässigt men skulle tariffunderlaget vara mindre, t.ex. motsvarande försäkring för MC istället för personbil, är det högst tveksamt om det är motiverat att beräkna en tariff per skadetyper. Omständigheter som kan förminska betydelsen av att beräkna en tariff enligt komplex S-GLM är att delkaskoförsäkringen nästan alltid säljs tillsammans med åtminstone en trafikförsäkring.

## 6.2 Slutsatser av valideringsmetoder

Korsvalidering och validering med sista årets data har gjorts eftersom båda är viktiga. Korsvalidering är viktig för att förstå stabiliteten i skattningarna och för andra valideringsmetoder. Validering av senaste data indikerar hur väl tariffen fungerar i praktiken då det kan ligga oidentifierade trender i skadekostnaden.

Kvotjämförelsen är en bra valideringsmetod för att den är enkel att beräkna och för att det blir enkelt att jämföra två tariffer. Ginikoefficienten med motsvarande graf för Lorenzkurvan är mer tidskrävande än kvotjämförelsen. Det finns klara samband mellan Lorenzkurvan och kvotjämförelsen i och med sorteringen enligt relationen  $R_i$ . Lorenzkurvan ger en generell bild av fördelningen mellan skador och premier men kvotjämförelsen är tydligare för att kunna jämföra tarifferna. En kvotjämförelse, kompletterat med ett stabilitetstest, ger en bra förutsättning för att kunna jämföra tariffer.

## 6.3 Fortsatta studier

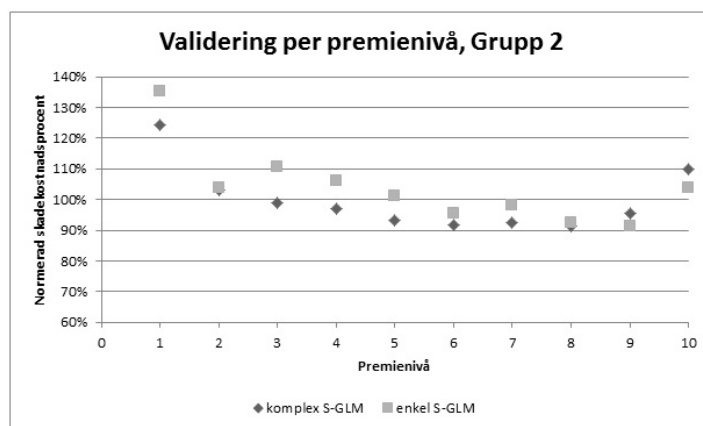
Slutsatserna om hur stor skillnad det är mellan tariffmetoderna är specifika för de rådande förutsättningarna som gäller i undersökningen. Det skulle vara intressant att undersöka hur antalet skadetyper, storleken av tariffunderlaget och antalet tariffceller påverkar resultatet.

# Litteraturförteckning

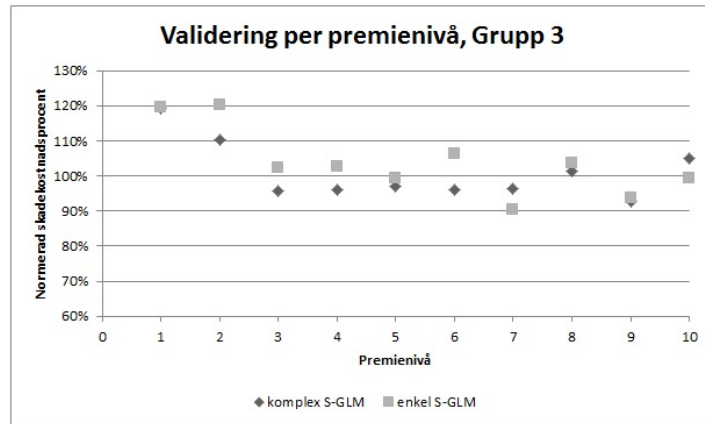
- [1] M. J. Brockman, "Statistical Motor Rating: Making effective use of your data," *JIA*, vol. 119, 1992.
- [2] E. W. Frees, G. Meyers, and A. D. Cummings, "Predictive Modeling of Multi-Peril Homeowners Insurance," *Variance*, vol. 6, 2011.
- [3] A. Agresti, *Categorical Data Analysis*. Wiley, 2002.
- [4] E. Ohlsson and B. Johansson, *Non-Life Insurance Pricing with Generalized Linear Models*. Springer, 2010.
- [5] J. Jung, "On Automobile Insurance Rating," *Astin Bullentin*, vol. 5, 1968.
- [6] <http://support.sas.com>, September 2013.
- [7] <http://sv.wikipedia.org/wiki/Korsvalidering>, Juli 2013.
- [8] G. Meyers and D. Cummings, "Goodness of Fit vs. Goodness of Lift," *www.casact.org*, 2010.
- [9] <http://sv.wikipedia.org/wiki/Ginikoefficient>, Juli 2013.
- [10] <http://sv.wikipedia.org/wiki/Lorenzkurvan>, Juli 2013.
- [11] E. W. Frees, G. Meyers, and A. D. Cummings, "Summarizing Insurance Scores Using a Gini Index," *www.casact.org*, 2010.
- [12] E. Larsson, "En jämförande studie av GLM, Jungs metod och Tweedie-modell för premiesättning av multiplikativ tariff.," *Examensarbete*, 2004.

## Bilaga A

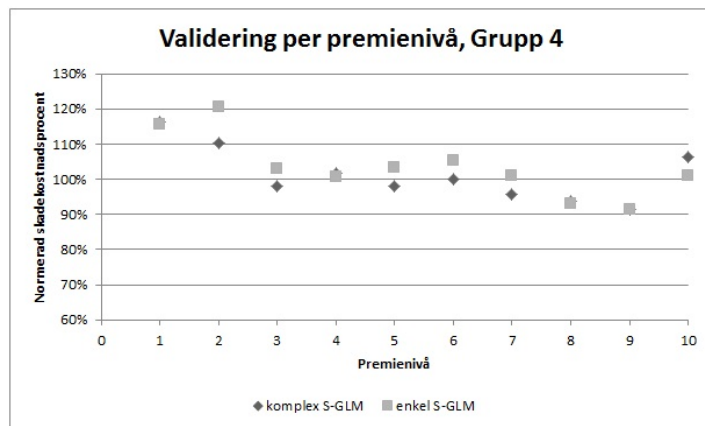
# Figurer, validering per premienivå



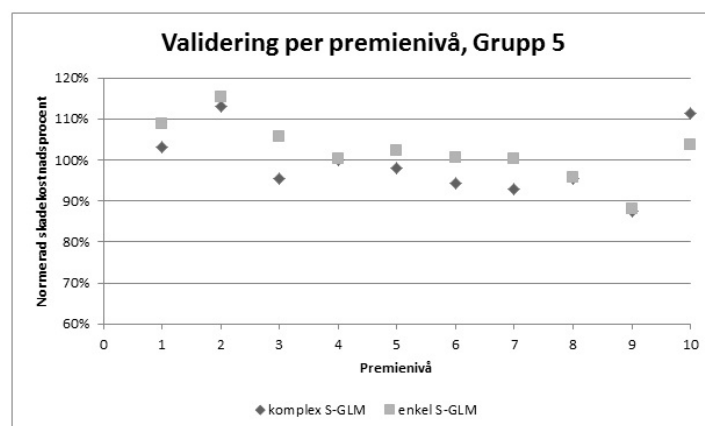
Figur A.1: Skadekostnadsprocent per premienivå för grupp 2



Figur A.2: Skadekostnadsprocent per premienivå för grupp 3



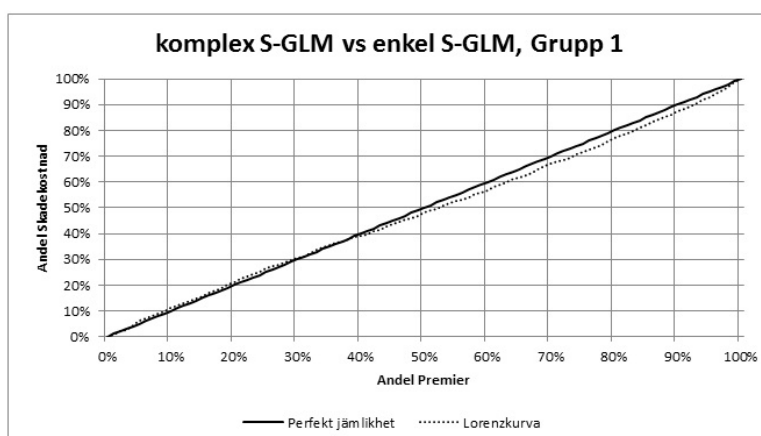
Figur A.3: Skadekostnadsprocent per premienivå för grupp 4



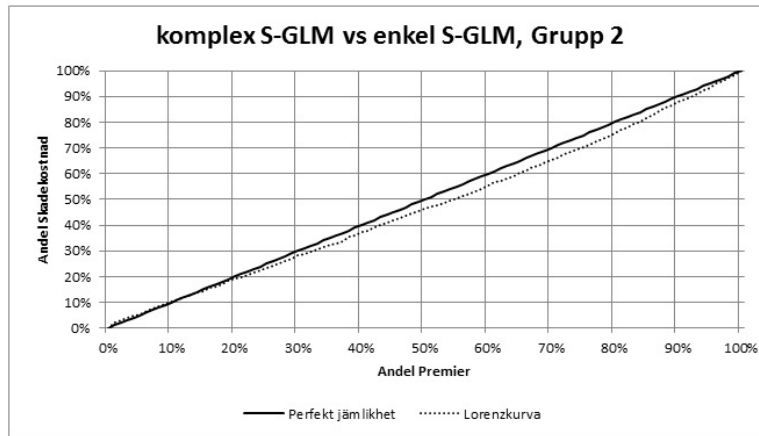
Figur A.4: Skadekostnadsprocent per premienivå för grupp 5

## Bilaga B

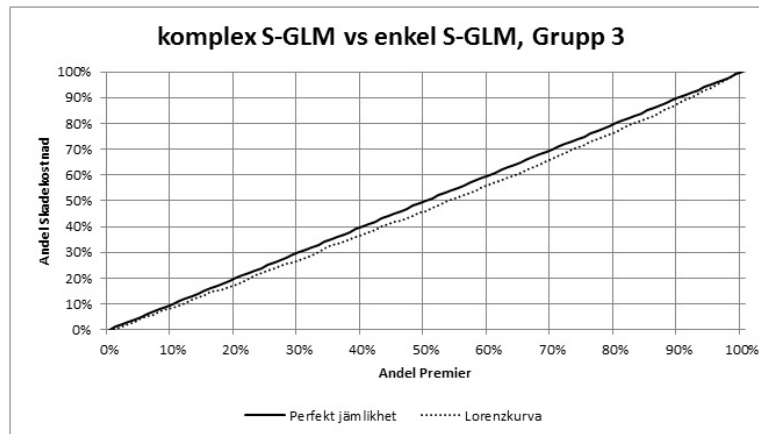
# Figurer, validering med ginikoefficient



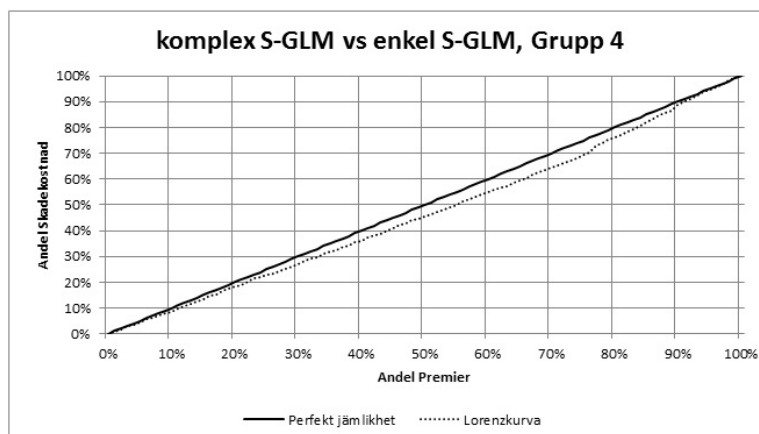
Figur B.1: Lorenzkurvan grupp 1



Figur B.2: Lorenzkurvan grupp 2



Figur B.3: Lorenzkurvan grupp 3



Figur B.4: Lorenzkurvan grupp 4