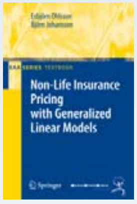
Release party:
Non-life Insurance Pricing with GLMs



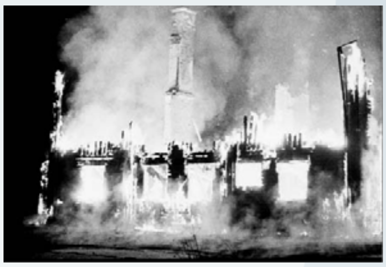
Esbjörn Ohlsson & Björn Johansson

Svenska Aktuarieföreningen 15 juni 2010

1

Brandstod enligt 1734 års lag

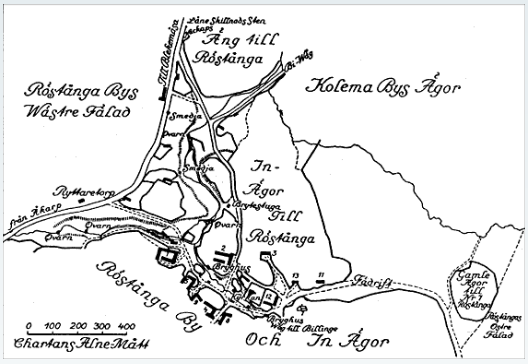
- Ersätter”för bonden nödige” hus samt säd, foder och boskap, ej annat lösöre”Af alle hemman i häradet skall brandstod gifvas efter hemmantalet” (premie i efterhand)
- Således premie i proportion till risken via variabeln ”hemmantal” (mantal) – en storleksvariabel som användes vid beskattning



2

Brandstodens prissättning

- Premie i proportion till förväntad kostnad
- Hjälpvariabel (hemmantal) används när kostnaden är okänd



Grundprincip prissättning

- Vi bortser här från administrationskostnader, vinst mm
- Skälig premie = E(kostnad)
- Försäkring utjämnar risk (osäkerhet, varians)
  – inte väntevärden
- Offentlig försäkring kan av politiska skäl fungera annorlunda
- Om oberoende risker:
  försäkringsbolagets kostnad relativt sett nära Summa(premier)

Marknaden avreglerades på 90-talet
Fri konkurrens – så här funkar det

Skälig premie 1500 kr

Skälig premie 2500 kr

2000 kr

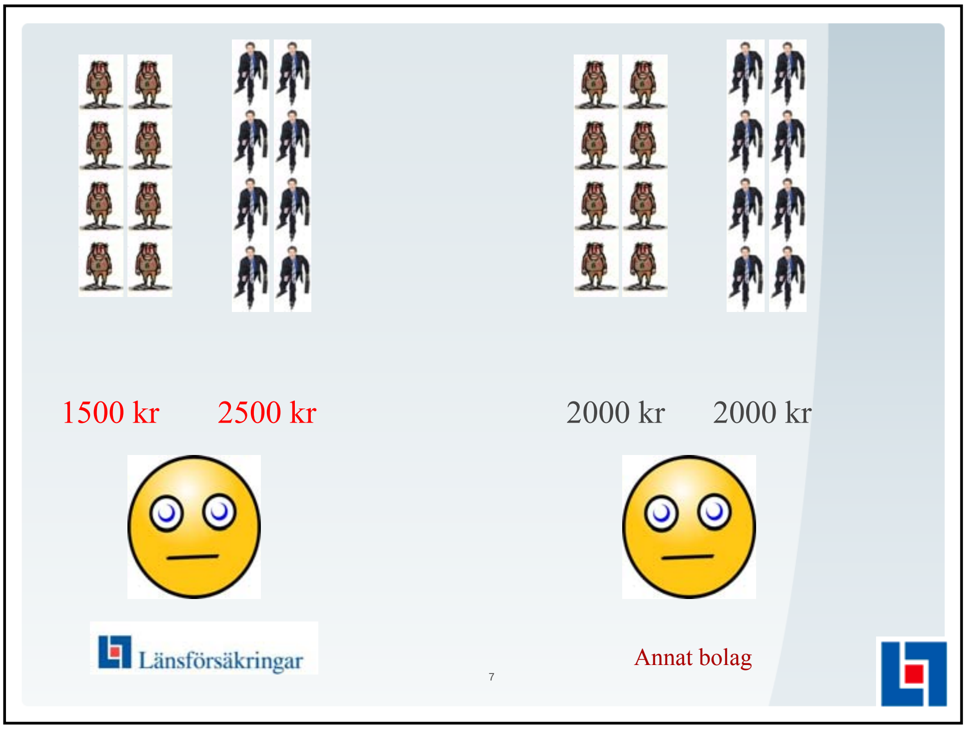
2000 kr

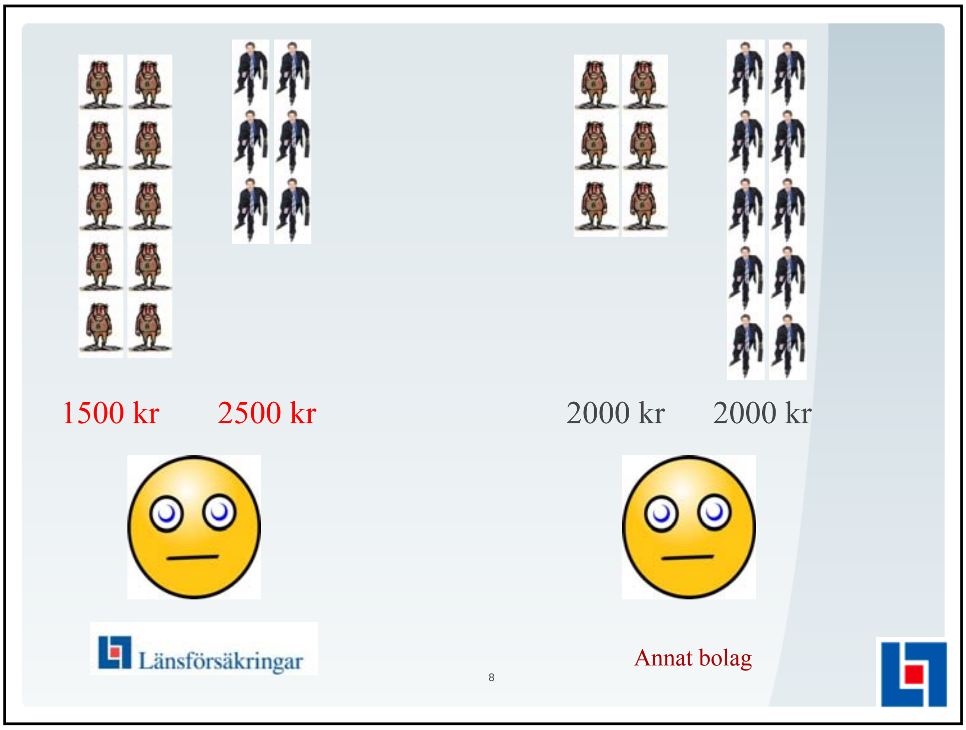
Länsförsäkringar

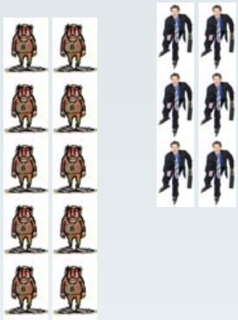
2000 kr

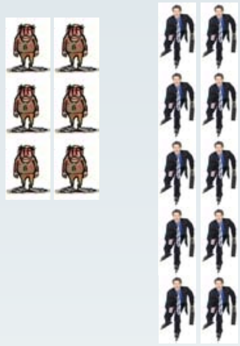
2000 kr

Annat bolag


1500 kr
2500 kr
2000 kr
2000 kr
Länsförsäkringar
7
Annat bolag


1500 kr
2500 kr
2000 kr
2000 kr
Länsförsäkringar
8
Annat bolag





1500 kr 2500 kr

2000 kr 2000 kr

Länsförsäkringar

Annat bolag

9

Hur gör man?

- Skadekostnad per försäkring = riskpremie
- Riskpremie = Skadefrekvens × Medelskada

$$R = S \times M$$

- I princip: skatta den enskilda försäkringens R
- I praktiken: gruppera försäkringarna efter några variabler $x_1, x_2, x_3, \ldots$ (jämför hemmantal)

10

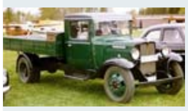
Tänkbara variabler, fallande acceptans

- Objektens skillnader
  (husstorlek, bilmodell)
- Kundens beteende
  (brandvarnare, larm, sprinklers)
- Andras beteende
  (farligt bostadsområde, dålig bilsäkerhet)
- Indirekta beteendevariabler som kunden
  inte kan styra (kön, ålder)

Förmodad acceptans
fallande skada

11

Exempel: observerad skadefrekvens lätt lastbil, promille (verkliga data)



Skade-frekvens	*Nytt fordon*	*Gammal*	***Totalt***
Låg kör-sträcka	33	25	**26**
Hög kör-sträcka	67	49	**61**
Totalt	**51**	**28**	

- 2 ggr så hög risk hög/låg körsträcka oberoende av ålder
- 1.3 ggr så hög risk nytt/gammalt oberoende av körsträcka

12

Multiplikativ tariff

- Exemplet: tillräckligt med data för att skatta R_{ij} direkt
- I praktiken många variabler och utjämning behövs
- Multiplikativ modell $R_{ij} = a_i b_j$
- Färre parametrar ger säkrare skattning
- Additiv modell $R_{ij} = a_i + b_j$ passar sällan

13

Sverige 50- och 60-tal

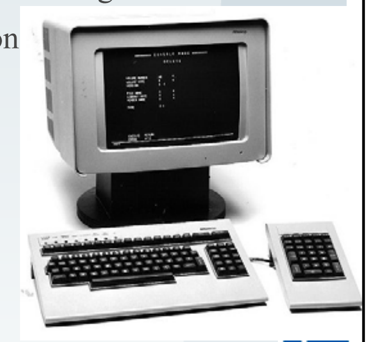
- Möte med aktuarieföreningen 1954:
  Bertil Almér, Fylgia, visade att multiplikativ modell passade väl för motorförsäkring
- 1 feb 1966: ny gemensam motortariff baserad på ADB-program för multiplikativ modell av Gunnar Andreasson
  - 1970 modifierat tillsammans med M-L Wennander

Källor: Jung (1968), Ajne (1982)

14

Jan Jung, ASTIN Colloquium 1966 Arnhem + ASTIN Bulletin 1968

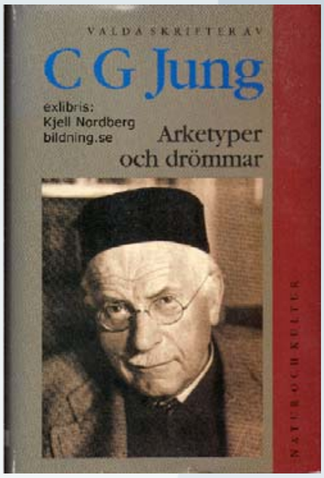
- Metod för iterativ skattning av parametrar i multiplikativ modell
- Baseras på ML-skattning av Poissonfördelning
- Metoden föreslås även för icke-Poisson ("heuristic approach")
- Björn Ajne (1982), *ASTIN Bulletin* Simuleringsstudie för att mäta standardavvikelsen i skattningarna
  - Tackar E. Elvers och P. Lindstroem för valuable discussions



15

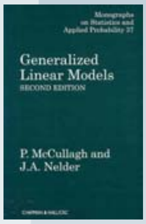
Internationellt

- Bailey & Simon (1960) studerar multiplikativ modell, *ASTIN Bulletin*
- Bailey (1963) först med "Jungs metod" (the method of marginal totals), *Proceedings of CAS*



16

Generaliserade linjära modeller GLM



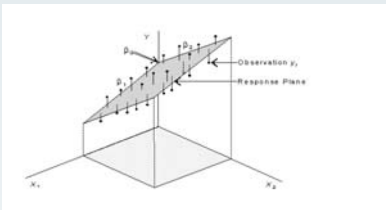
- 1972 introducerar Nelder & Wedderburn GLM
- 1977 McCullagh-Nelder klassisk bok om GLM
  - ett försäkringsexempel, men ej multiplikativ modell
- 1984 Coutts PhD thesis City University om motortariff med GLM
- 1992 Brockman & Wright: praktiskt introduktion till GLM inom motor-försäkring, med GLIM-kod
- 1998 Brockman m fl på EMB lanserar Emblem
- 2010 GLM används inom sakförsäkring över hela (?) världen

GLM börjar användas på försäkringsbolagen

17

GLM vs linjär regression (1)

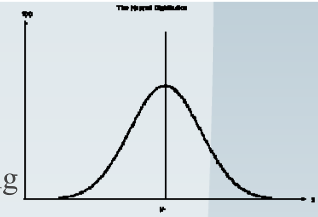
- Tariffanalys är som sagt ett regressionsproblem: hur beror R på x_1, x_2,…



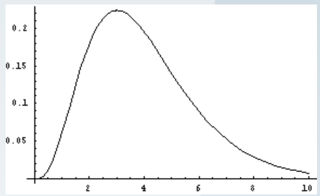
- Men dess additiva/linjära struktur för väntevärdet av R passar ej försäkring
- GLM generaliserar: en funktion av väntevärdet ska vara linjär
  - Speciellt ingår multiplikativ modell

18

GLM vs linjär regression (2)



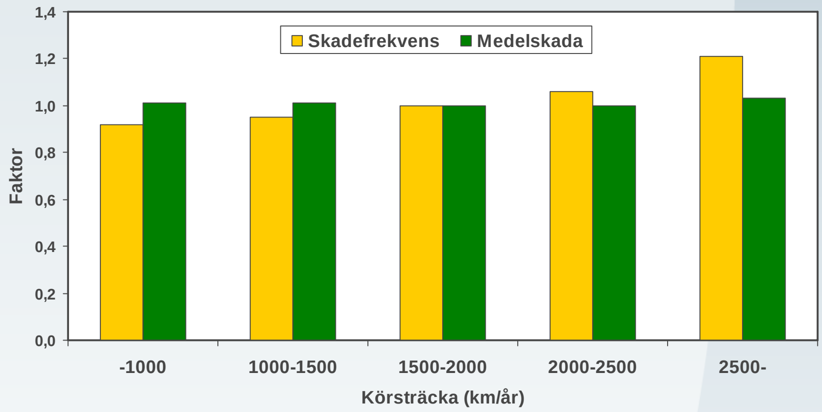
- Linjär regression antar normalfördelning
- Försäkringsdata:
  - Poissonfördelat antal skador
  - Positiv kontinuerlig fördelning för skadebeloppen, t ex Gamma



- GLM tillåter stor familj fördelningar, inkl Poisson och Gamma
- I praktiken bara relationen väntevärde-varians som är viktig, inte exakt fördelning

19

Separat analys av skadefrekvens och medelskada fick sitt genombrott med introduktionen av GLM



20

Vad har GLM mer tillfört?

- Tillgång till allmän, vedertagen statistisk teori och metodik
  - Konfidensintervall
  - Residualer
  - Test
- Möjlighet att använda standardprogram (SAS, GLIM, R)
- Med statistisk modell i botten finns god terräng för metodutveckling

God terräng

Exempel: Bilmärkesklassning

- Data: Försäkrings- och skadehistorik för personbil
- Variabler:
  - Bilmodell (ca 3800 olika)
  - Hjälpvariabler per bilmodell (bilmärke, vikt, vikt/effekt,…)
  - Alla övriga tariffvariabler: Ägarålder, körsträcka, geografisk zon…
- Problem: Skatta premiefaktorer för bilmodell

Analyser som bilmärkesklassningen kräver avancerade statistiska modeller

- Tre grundläggande typer av variabler:
  - Kategoriska, med få klasser (t ex körsträckeklass)
  - Multiklass, kategoriska med många klasser (t ex bilmodell)
  - Kontinuerliga (t ex fordonsvikt)
- Samspelande variabler ger upphov till ytterligare typer:
  - Kategorisk / kontinuerlig (t ex kön / ägarålder)
  - Kontinuerlig / kontinuerlig (t ex fordonsålder / fordonsvikt)
- Behovet att modellera alla typer av variabler på ett bra sätt har lett till utvidgningar av GLM: GLMM, HGLM, GAM…

23

Analys av nyckeltal

X_i	Antal skador	Skadekostnad	Skadekostnad
w_i	Antal försäkringsår	Antal skador	Antal försäkringsår
$Y_i = X_i / w_i$	Skadefrekvens	Medelskada	Riskpremie

24

Standard-GLM inom sakförsäkring

- Modell för väntevärdet $\mu_i = E[Y_i]$:
  $\eta_i = g(\mu_i)$
  $\eta_i = u_1(x_{1i}) + u_2(x_{2i}) + u_3(x_{3i},x_{4i}) + \ldots$
  $u_k(x_{ki}) = \beta_{k1} f_{k1}(x_{ki}) + \beta_{k2} f_{k2}(x_{ki}) + \ldots$
- Multiplikativ modell: $g(\mu) = \log \mu \rightarrow g^{-1}(\eta) = \exp\{\eta\}$

  $\mu_i = \exp\{u_1(x_{1i})\} \exp\{u_2(x_{2i})\} \exp\{u_3(x_{3i},x_{4i})\} + \ldots$
- Exempel:
  x_{1i} = körsträcka (låg/hög) x_{2i} = fordonsålder (ny/gammal)

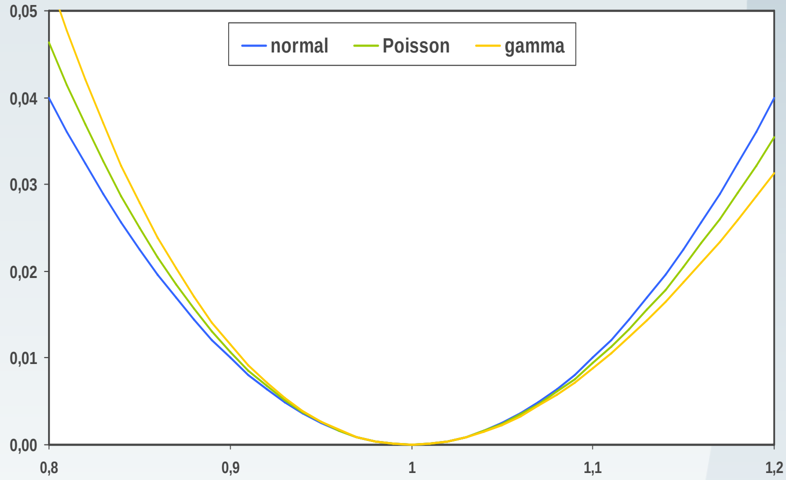
  Låg körsträcka, gammalt fordon:
  $\eta_i = \beta_{11} + \beta_{22}$
  $\mu_i = \exp\{\beta_{11} + \beta_{22}\} = \exp\{\beta_{11}\} \exp\{\beta_{22}\} = \gamma_{11} \gamma_{22}$

Deviansen mäter anpassningen till data

- Variansantagande:
  $V(Y_i) = \varphi\, v(\mu_i) / w_i$
- Deviansen $D = \Sigma_i\, w_i\, d(y_i,\mu_i)$

  $v(\mu) = 1 \rightarrow d(y,\mu) = (y - \mu)^2$ ("normal")
  $v(\mu) = \mu \rightarrow d(y,\mu) = 2\,(y \log y - y \log \mu - y + \mu)$ ("Poisson")
  $v(\mu) = \mu^2 \rightarrow d(y,\mu) = 2\,(y/\mu - 1 - \log (y/\mu))$ ("gamma")

Deviansen d(1,μ)



Standard-GLM inom sakförsäkring

- Variansantaganden:

  Skadefrekvens: $v(\mu) = \mu$

  Medelskada: $v(\mu) = \mu^2$

- Finn β-parametrar som minimerar deviansen:

  $D = \Sigma_i \; w_i \; d(y_i,\mu_i)$

  $\mu_i = g^{-1}(\eta_i) = \exp\{\eta_i\}$

  $\eta_i = u_1(x_{1i}) + u_2(x_{2i}) + u_3(x_{3i},x_{4i}) + \ldots$

  $u_k(x_{ki}) = \beta_{k1} \; f_{k1}(x_{ki}) + \beta_{k2} \; f_{k2}(x_{ki}) + \ldots$

Multiklassvariabler: GLM med stokastiska effekter

- Kategorisk: $u_k(x_{ki}) = \beta_{k1}\, f_{k1}(x_{ki}) + \beta_{k2}\, f_{k2}(x_{ki}) + \ldots$
- Multiklass: $u_k(x_{ki}) = Z_{k1}\, f_{k1}(x_{ki}) + Z_{k2}\, f_{k2}(x_{ki}) + \ldots$,
  där $Z_{k1}, Z_{k2}, \ldots$ är stokastiska variabler
- Ohlsson-Johansson: $Z_{12}, Z_{k2}, \ldots$ skattas m h a kredibilitetsteori
- Inspirationskällor:
  - Svenska aktuarieidéer, t ex Campbell (1986)
  - Kredibilitetsteori (Bühlmann-Straub, Jewells hierarkiska modell)
  - HGLM (Lee & Nelder, 1996)

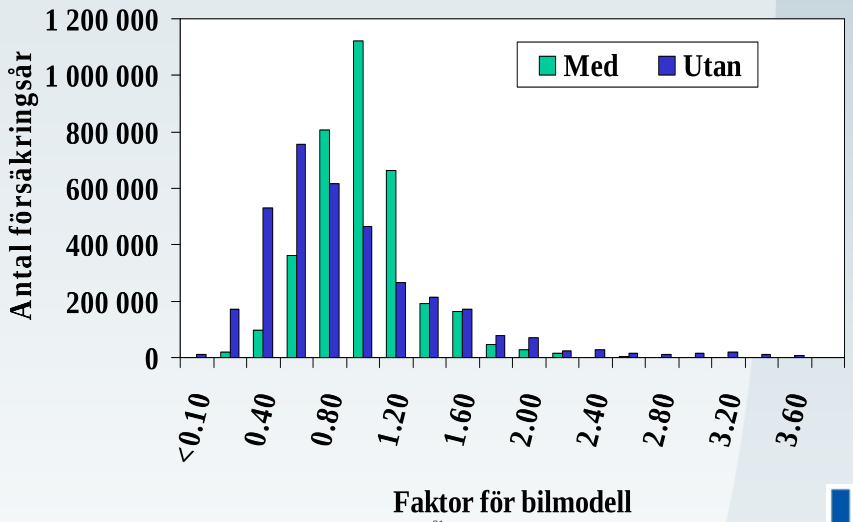
29

Bilmärkesklassningen: Kombinera tekniska data med historik för enskilda modeller

- Bilmodellfaktor = (faktor för vikt) × (faktor för vikt/effekt) × …(faktorer för andra tekniska variabler) × (faktor för bilmärke) × (faktor för bilmodell)
- Stort dataunderlag: Skadehistoriken väger tyngst
- Litet dataunderlag: Tekniska data väger tyngst

30

De tekniska variablerna minskar behovet av individuell skadehistorik



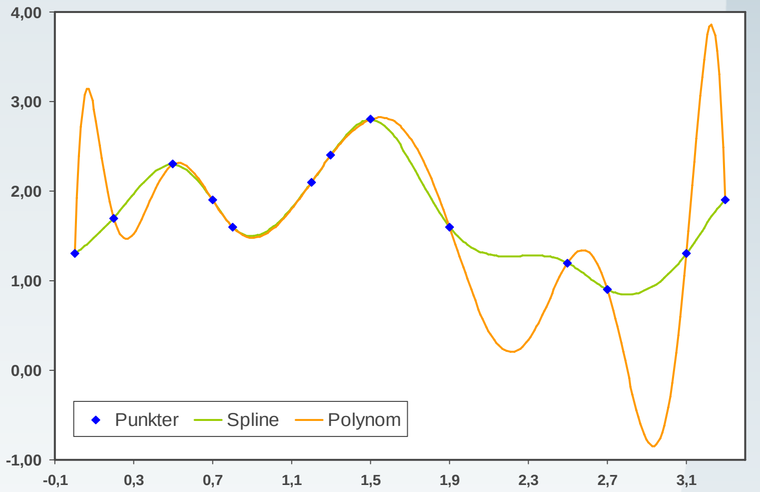
31

Generaliserade additiva modeller (GAM) används för att modellera kontinuerliga variabler

- Modell för väntevärdet:
  $\eta_i = g(\mu_i)$
  $\eta_i = u_1(x_{1i}) + u_2(x_{2i}) + u_3(x_{3i}) + \ldots$
- Antag att x_{1i} är kontinuerlig
  $u_1(x_{1i}) = h(x_{1i})$
  $u_2(x_{2i}) = \beta_{21}\, f_{21}(x_{2i}) + \beta_{22}\, f_{22}(x_{2i}) + \ldots$
  …
  h är en godtycklig två gånger deriverbar funktion
- Minimera den penaliserade deviansen :
  $\Delta = \Sigma_i\, w_i\, d(y_i,\mu_i) + \lambda \int (h''(x))^2\, dx$

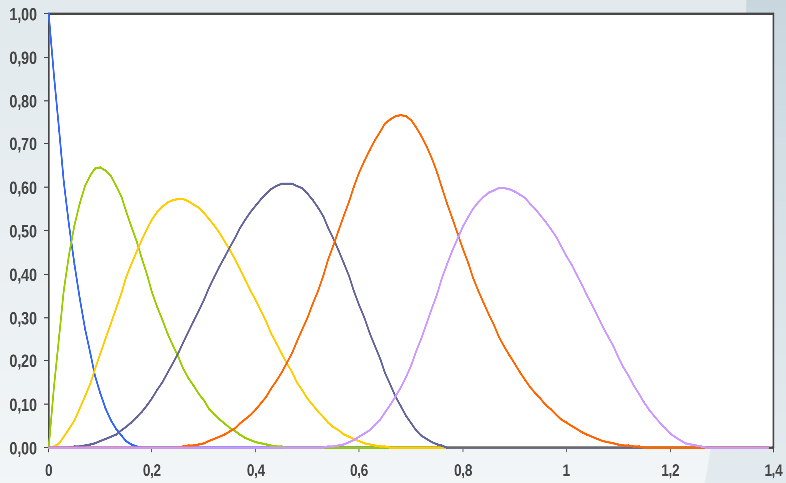
32

Bland alla interpolerande två gånger deriverbara funktioner h, minimeras $\int (h''(x))^2\,dx$ av en kubisk spline



33

Varje kubisk spline är en linjärkombination av kubiska B-splines



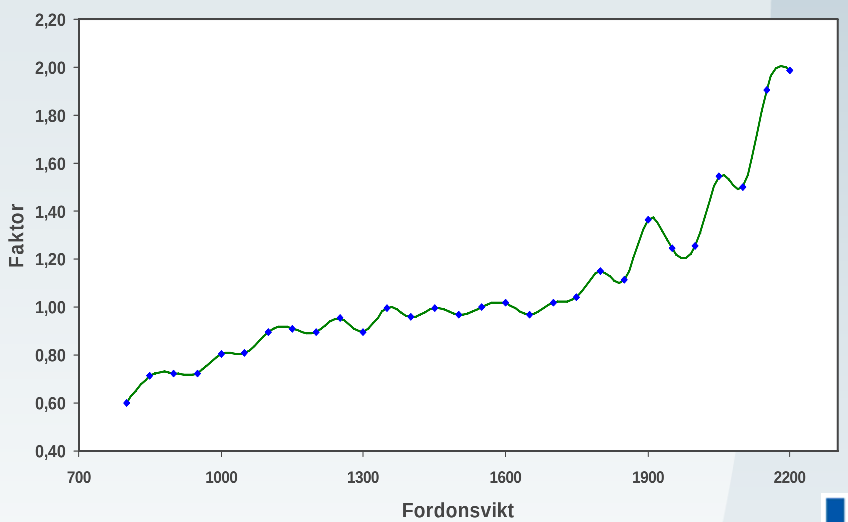
34

→ Modellen går att parametrisera

- Funktionerna $u_k(x)$ är linjära:
  $u_1(x_{1i}) = h(x_{1i}) = \beta_{11} B_{11}(x_{1i}) + \beta_{12} B_{12}(x_{1i}) + \ldots$
  $u_3(x_{3i}) = \beta_{31} f_{31}(x_{3i}) + \beta_{32} f_{32}(x_{3i}) + \ldots$
  $\ldots$
- $\int (h''(x))^2\, dx = \Sigma_m \Sigma_n\, \beta_{km} \beta_{kn} \int B_{km}''(x)\, B_{kn}''(x)\, dx$
- Finn β-parametrar som minimerar den penaliserade deviansen :
  $\Delta = \Sigma_i\, w_i\, d(y_i,\mu_i) + \lambda \int (h''(x))^2\, dx$
- Det finns ett flertal metoder för att skatta ”smoothing-parametern” λ

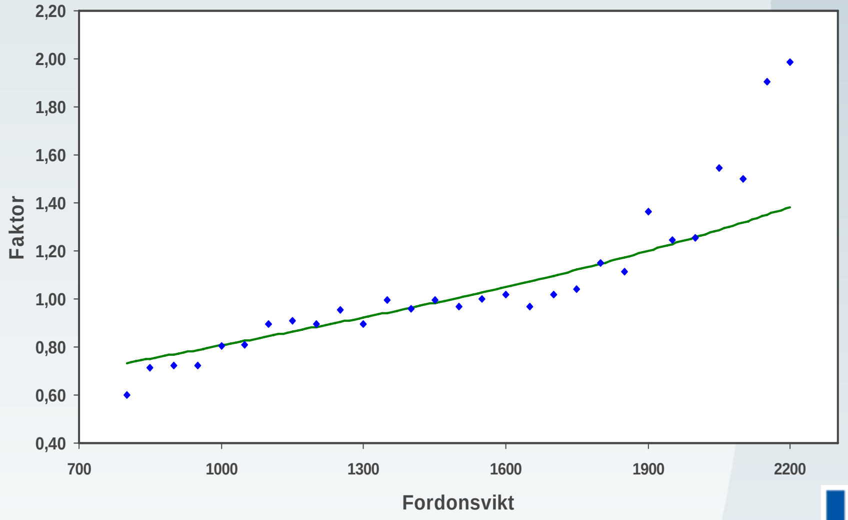
35

Lågt värde på smoothing-parametern



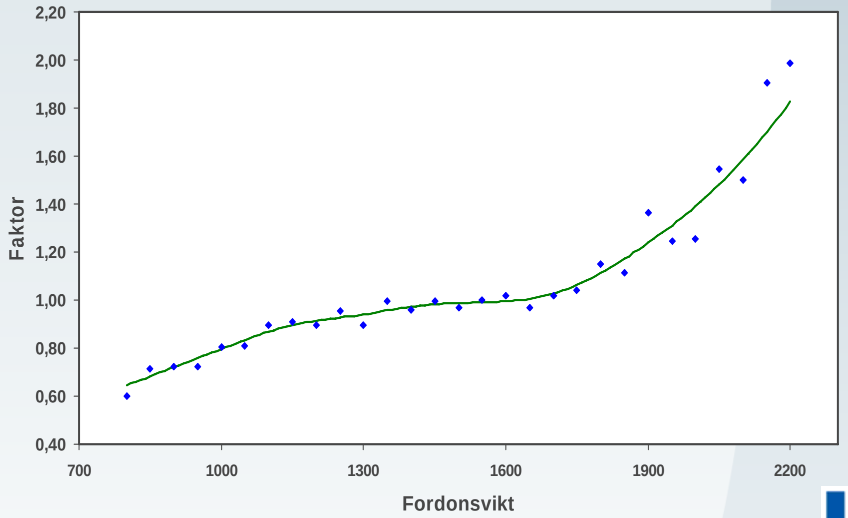
36

Högt värde på smoothing-parametern



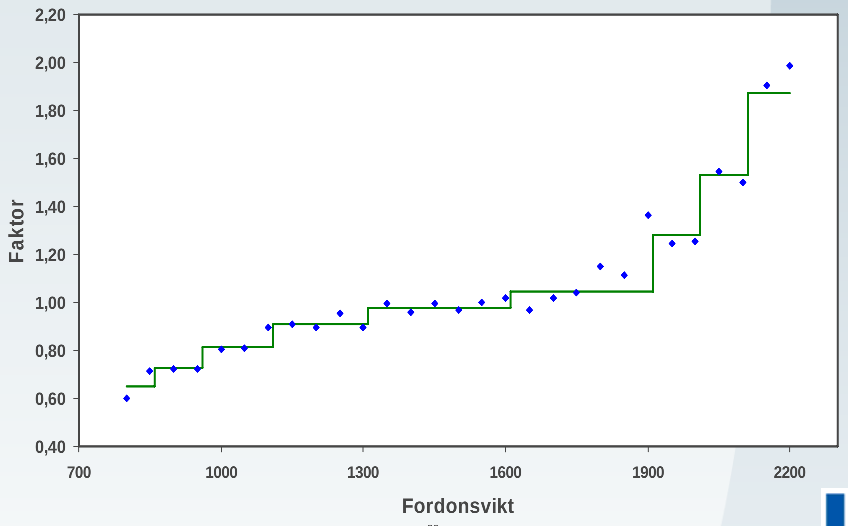
37

Skattad smoothing-parameter



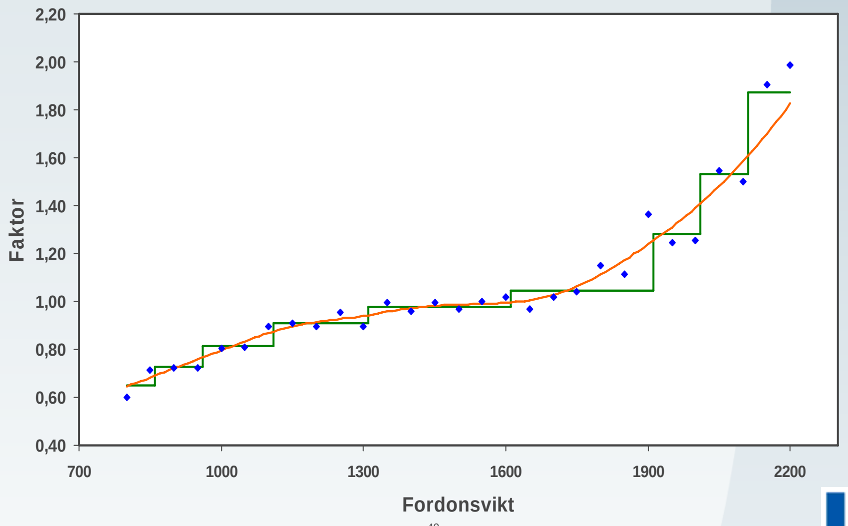
38

Traditionell metod



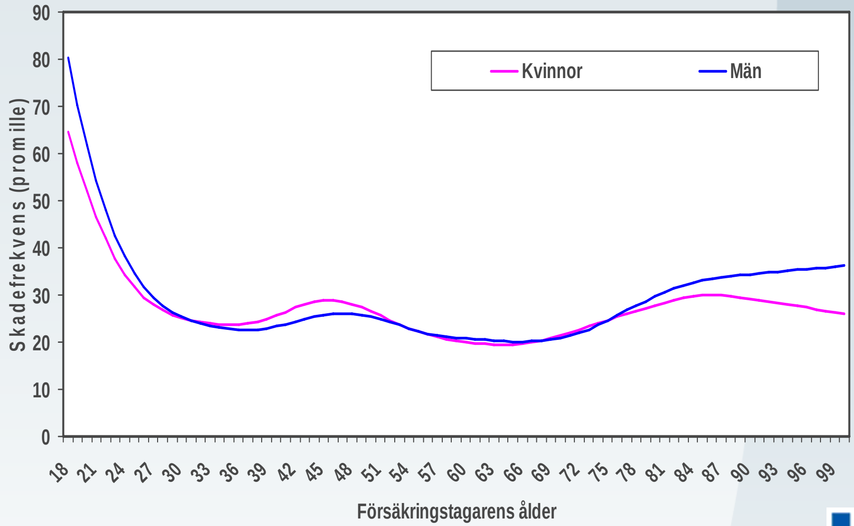
39

En kontinuerlig kurva ger bättre riskanpassning



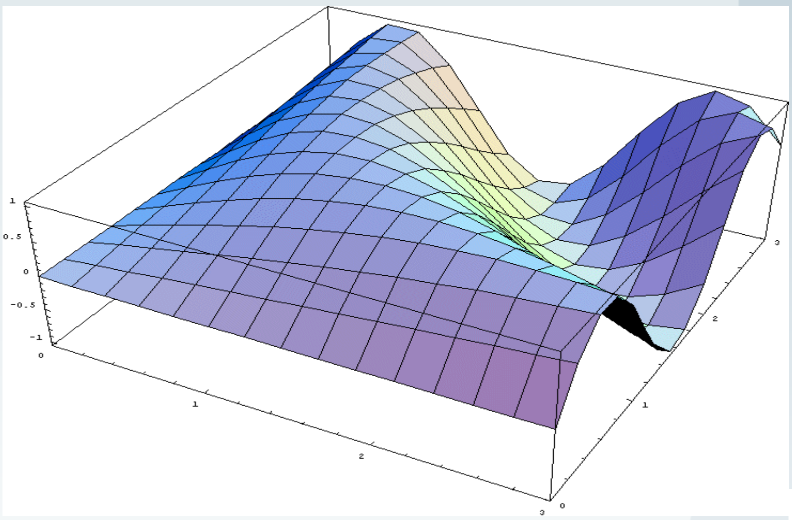
40

Koppling mellan kategoriska och kontinuerliga variabler

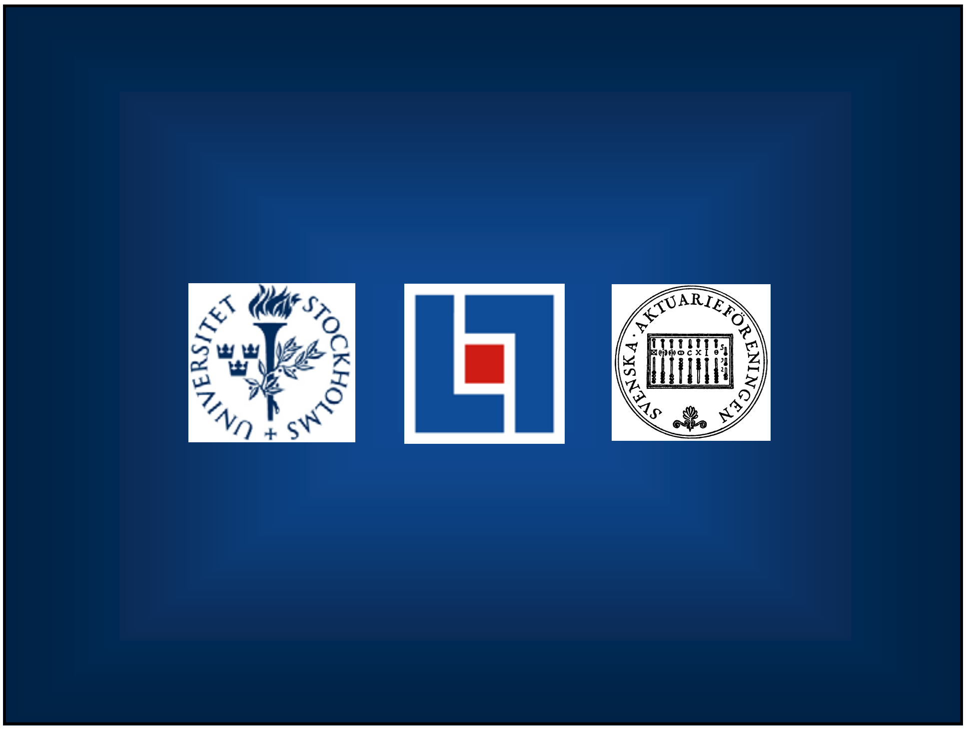


41

Koppling mellan två kontinuerliga variabler – Thin plate splines



42


STOCKHOLMS UNIVERSITET
SVENSKA AKTUARIEFÖRENINGEN