Non-life insurance mathematics

Nils F. Haavardsson, University of Oslo and DNB Skadeforsikring

Main issues so far

- Why does insurance work?
- How is risk premium defined and why is it important?
- How can claim frequency be modelled?
  - Poisson with deterministic muh
  - Poisson with stochastic muh
  - Poisson regression

Insurance works because risk can be diversified away through size

- The core idea of insurance is risk spread on many units
- Assume that policy risks $X_1, \ldots, X_J$ are stochastically independent
- Mean and variance for the portfolio total are then

$$E(\chi) = \pi_1 + \ldots + \pi_J \quad \text{and} \quad \text{var}(\chi) = \sigma_1 + \ldots + \sigma_J$$

$$\text{and} \quad \pi_j = E(X_j) \quad \text{and} \quad \sigma_j = sd(X_j). \text{ Introduce}$$

$$\bar{\pi} = \frac{1}{J}(\pi_1 + \ldots + \pi_J) \quad \text{and} \quad \bar{\sigma}^2 = \frac{1}{J}(\sigma_1 + \ldots + \sigma_J)$$

which is average expectation and variance. Then

$$E(\chi) = J\bar{\pi} = \quad \text{and} \quad \text{sd}(\chi) = \sqrt{J}\,\bar{\sigma} \text{ so that } \frac{\text{sd}(\chi)}{\text{E}(\chi)} = \frac{\bar{\sigma}/\bar{\pi}}{\sqrt{J}}$$

- The *coefficient of variation* approaches 0 as J grows large (law of large numbers)
- Insurance risk can be diversified away through size
- Insurance portfolios are still not risk-free because
    - of uncertainty in underlying models
    - risks may be dependent

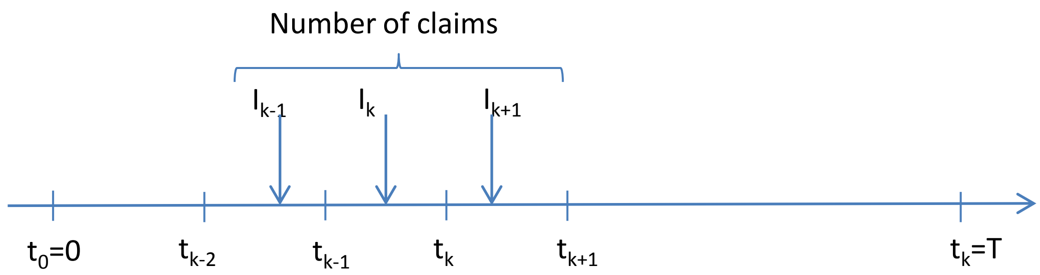
Risk premium expresses cost per policy and is important in pricing

- Risk premium is defined as P(Event)*Consequence of Event
- More formally

$$
\begin{aligned}
&\text{Risk premium} = P(\text{event}) * \text{Consequence of event} \\
&= \text{Claim frequency} * \text{Claim severity} \\
&= \frac{\text{Number of claims}}{\text{Number of risk years}} * \frac{\text{Total claim amount}}{\text{Number of claims}} \\
&= \frac{\text{Total claim amount}}{\text{Number of risk years}}
\end{aligned}
$$

- From above we see that risk premium expresses cost per policy
- Good price models rely on sound understanding of the risk premium
- We start by modelling claim frequency

The world of Poisson (Chapter 8.2)



- What is rare can be described mathematically by cutting a given time period T into K small pieces of equal length h=T/K
- On short intervals the chance of more than one incident is remote
- Assuming no more than 1 event per interval the count for the entire period is

$N=I_1+\ldots+I_K$, where I_j is either 0 or 1 for j=1,...,K

- If $p=\Pr(I_k=1)$ is equal for all k and events are independent, this is an ordinary Bernoulli series

$$\Pr(N=n)=\frac{K!}{n!(K-n)!}p^n(1-p)^{K-n}, \quad \text{for} \quad n=0,1,\ldots,K$$

- Assume that p is proportional to h and set $p=\mu h$ where μ is an intensity which applies per time unit

The world of Poisson

$$\Pr(N=n)=\frac{K!}{n!(K-n)!}p^n(1-p)^{K-n}$$

$$=\frac{K!}{n!(K-n)!}\left(\frac{\mu T}{K}\right)^n\left(1-\frac{\mu T}{K}\right)^{K-n}$$

$$=\frac{(\mu T)^n}{n!}\frac{K(K-1)\cdots(K-n+1)}{K^n}\left(1-\frac{\mu T}{K}\right)^K\frac{1}{\left(1-\frac{\mu T}{K}\right)^n}$$

$$\underset{K\to\infty}{\to}1 \qquad \underset{K\to\infty}{\to}e^{-\mu T} \qquad \underset{K\to\infty}{\to}1$$

$$\Rightarrow \Pr(N=n)\underset{K\to\infty}{\to}\frac{(\mu T)^n}{n!}e^{-\mu T}$$

In the limit N is Poisson distributed with parameter $\lambda=\mu T$

The world of Poisson

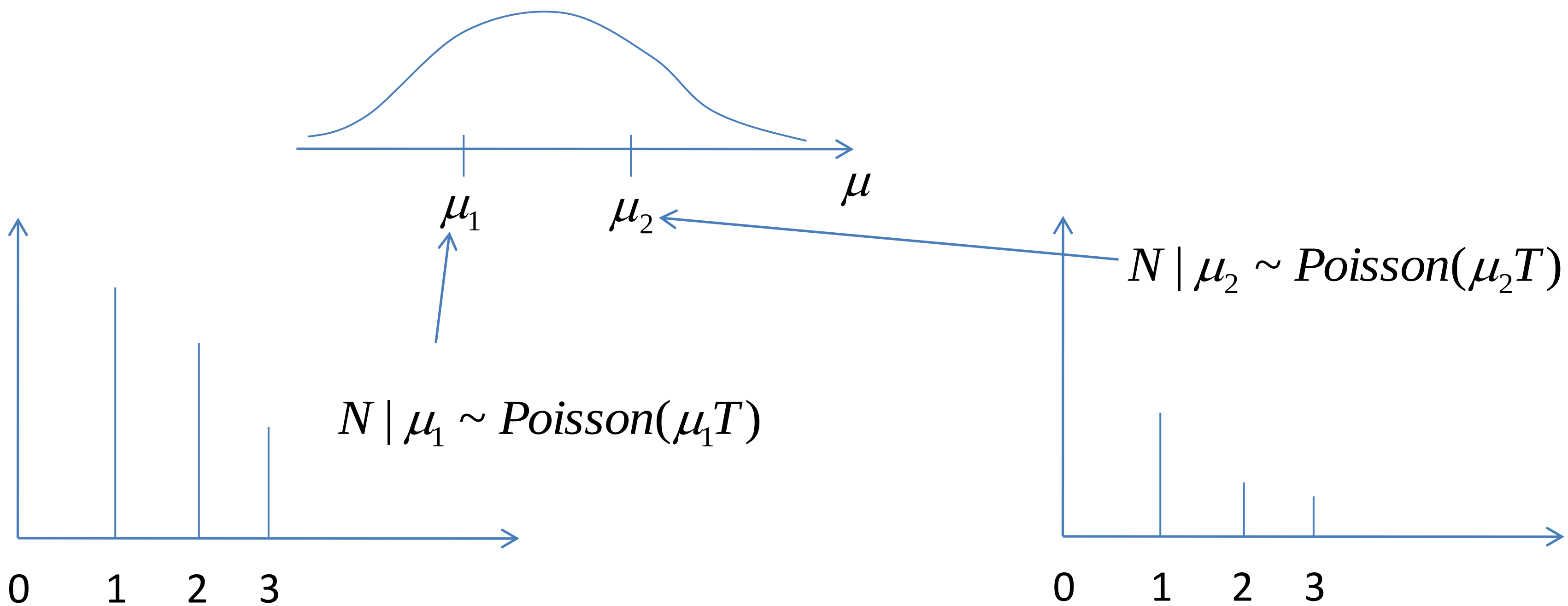
- It follows that the portfolio number of claims **N** is Poisson distributed with parameter

$$\lambda = (\mu_1 + ... + \mu_J)T = J\bar{\mu}T, \quad \text{where } \bar{\mu} = (\mu_1 + ... + \mu_J)/J$$

- When claim intensities vary over the portfolio, only their average counts

Random intensities (Chapter 8.3)

- How μ varies over the portfolio can partially be described by observables such as age or sex of the individual (treated in Chapter 8.4)
- There are however factors that have impact on the risk which the company can't know much about
  - Driver ability, personal risk averseness,
- This randomeness can be managed by making μ a stochastic variable



Random intensities (Chapter 8.3)

- The models are conditional ones of the form

$$N \mid \mu \sim Poisson(\mu T) \quad \text{and} \quad \mathbf{N} \mid \mu \sim Poisson(J\mu T)$$

Policy level Portfolio level

- Let $\xi = E(\mu)$ and $\sigma = \text{sd}(\mu)$ and recall that $E(N \mid \mu) = \text{var}(N \mid \mu) = \mu T$

  which by double rules in Section 6.3 imply

$$E(N) = E(\mu T) = \xi T \quad \text{and} \quad \text{var}(N) = E(\mu T) + \text{var}(\mu T) = \xi T + \sigma^2 T^2$$

- Now E(N)<var(N) and N is no longer Poisson distributed

The Poisson regression model (Section 8.4)

- The idea is to attribute variation in μ to variations in a set of observable variables $x_1,\ldots,x_v$. Poisson regressjon makes use of relationships of the form

$$\log(\mu) = b_0 + b_1 x_1 + \ldots + b_v x_v \qquad (1.12)$$

- Why $\log(\mu)$ and not μ itself?
- The expected number of claims is non-negative, where as the predictor on the right of (1.12) can be anything on the real line
- It makes more sense to transform μ so that the left and right side of (1.12) are more in line with each other.

- Historical data are of the following form

  - n_1 T_1 $x_{11}\ldots x_{1x}$
  - n_2 T_2 $x_{21}\ldots x_{2x}$

  - n_n T_n $x_{n1}\ldots x_{nv}$

  Claims exposure covariates

- The coefficients $b_0,\ldots,b_v$ are usually determined by likelihood estimation

The model (Section 8.4)

- In likelihood estimation it is assumed that n_j is Poisson distributed $\lambda_j = \mu_j T_j$ where μ_j is tied to covariates $x_{j1},\ldots,x_{jv}$ as in (1.12). The density function of n_j is then

$$f(n_j) = \frac{(\mu_j T_j)^{n_j}}{n_j!} \exp(-\mu_j T_j)$$

or

$$\log(f(n_j)) = n_j \log(\mu_j) + n_j \log(T_j) - \log(n_j!) - \mu_j T_j$$

- $\log(f(n_j))$ above is to be added over all j for the likehood function $L(b_0,\ldots,b_v)$.
- Skip the middle terms $n_j T_j$ and $\log(n_j!)$ since they are constants in this context.
- Then the likelihood criterion becomes

$$L(b_0,\ldots,b_v) = \sum_{j=1}^{n} \{n_j \log(\mu_j) - \mu_j T_j\} \text{ where } \log(\mu_j) = b_0 + b_1 x_{j1} + \ldots + b_j x_{jv} \quad (1.13)$$

- Numerical software is used to optimize (1.13).
- McCullagh and Nelder (1989) proved that $L(b_0,\ldots,b_v)$ is a convex surface with a single maximum
- Therefore optimization is straight forward.

Main steps

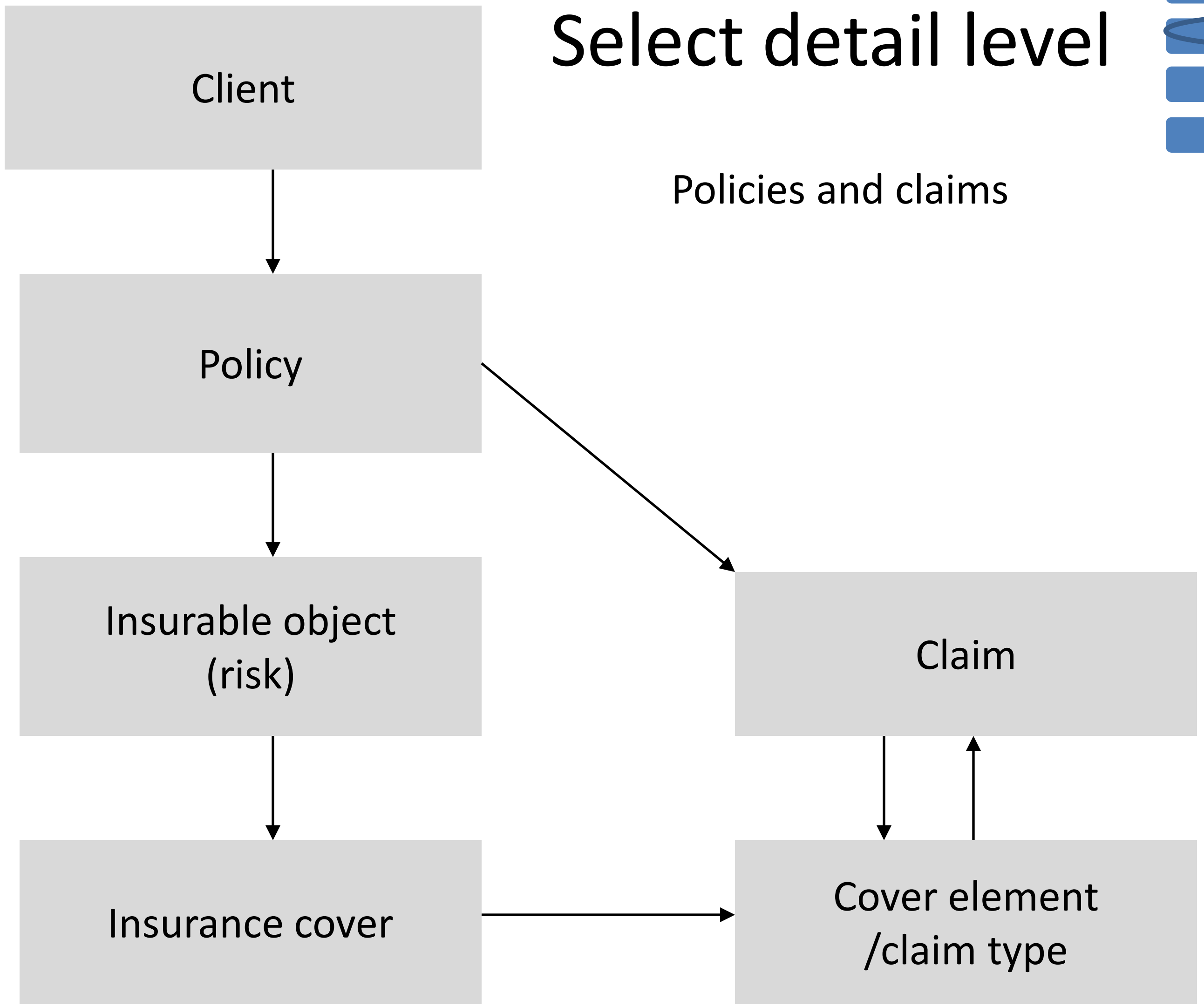
- How to build a model
- How to evaluate a model
- How to use a model

How to build a regression model

- Select detail level
- Variable selection
- Grouping of variables
- Remove variables that are strongly correlated
- Model important interactions

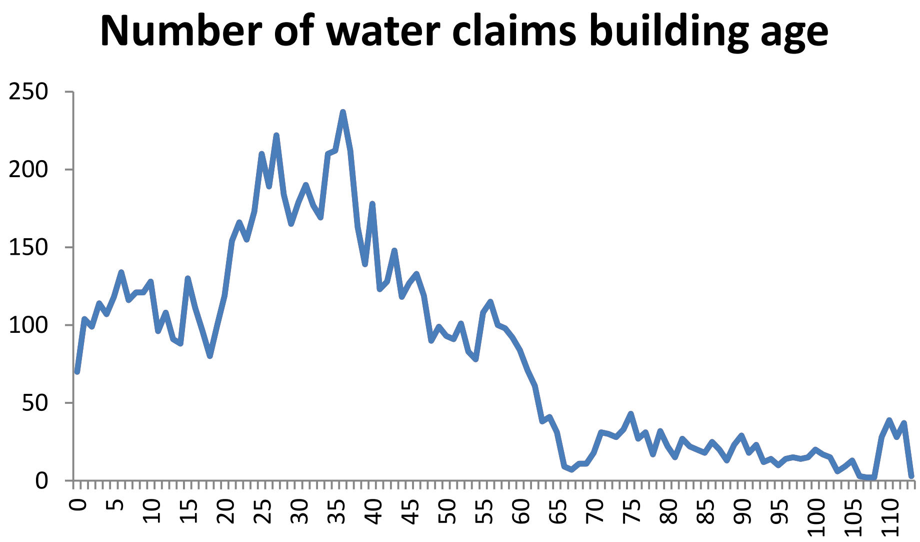
Select detail level

Policies and claims



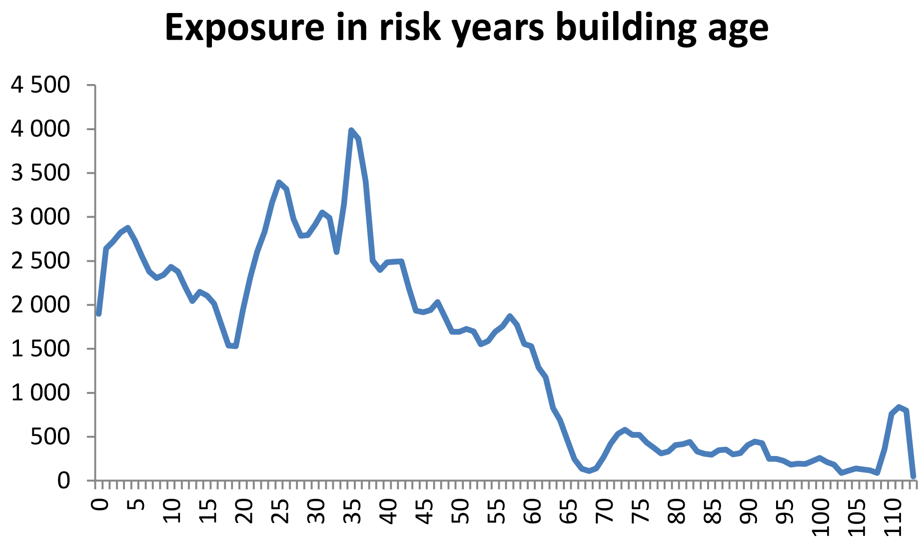
Step 1 in model building: Select detail level

Grouping of variables

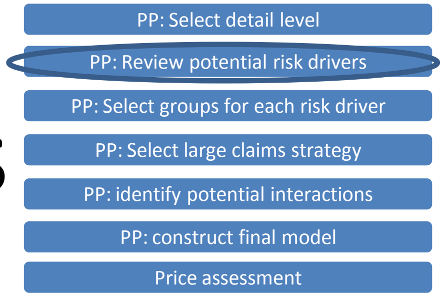


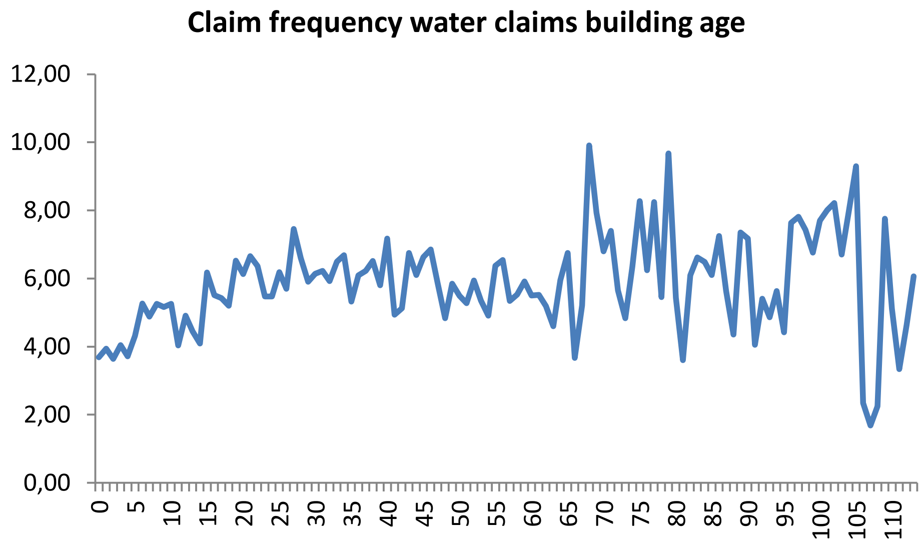
Grouping of variables

PP: Select detail level
PP: Review potential risk drivers
PP: Select groups for each risk driver
PP: Select large claims strategy
PP: identify potential interactions
PP: construct final model
Price assessment



Grouping of variables



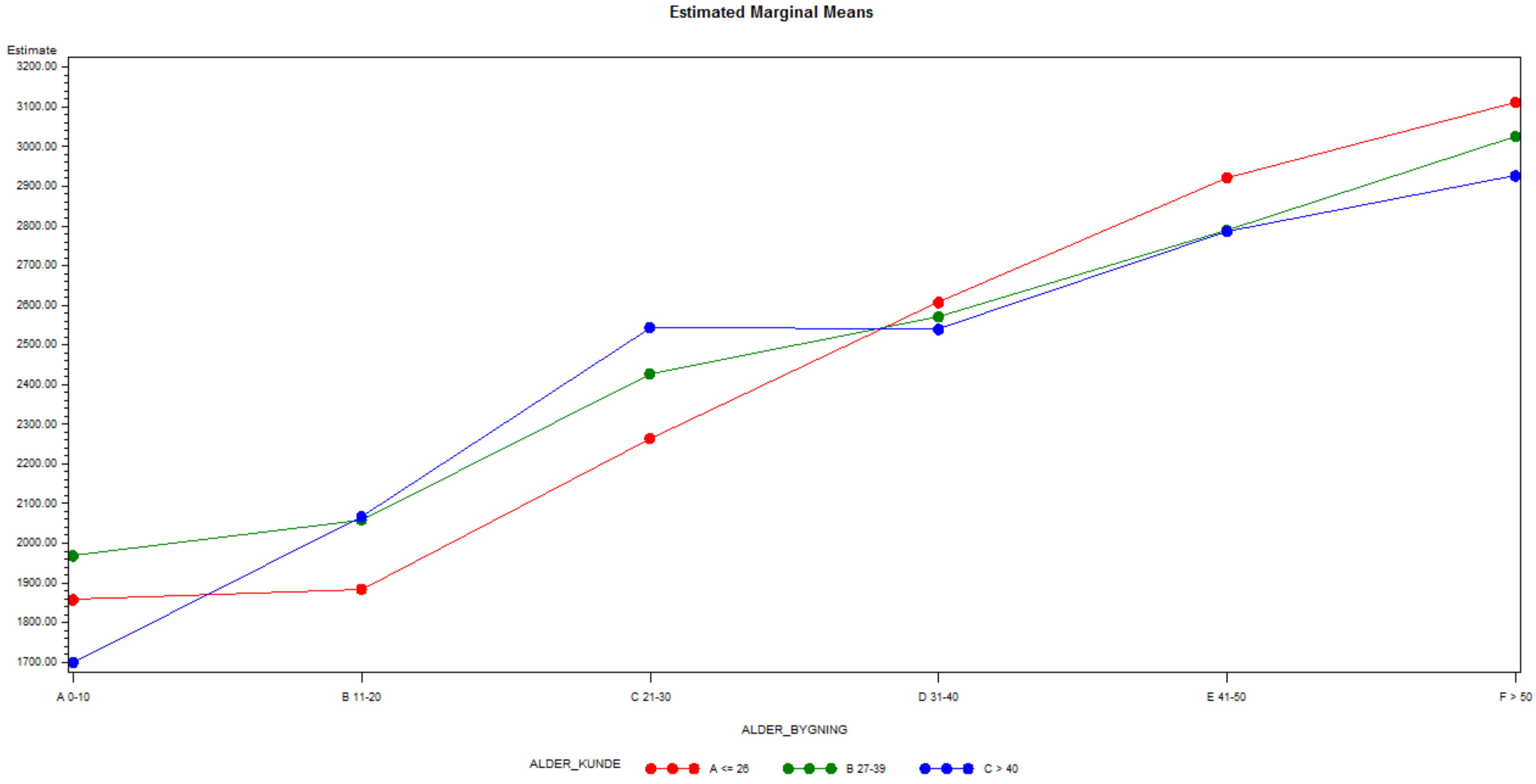


Model important interactions

- Definition:
  - Consider a regression model with two explanatory variables A and B and a response Y. If the effect of A (on Y) depends on the level of B, we say that there is an *interaction* between A and B
- Example (house owner):
  - The risk premium of new buildings are lower than the risk premium of old buildings
  - The risk premium of young policy holders is higher than the risk premium of old policy holders
  - The risk premium of young policy holders in old buildings is particularly high
  - Then there is an interaction between building age and policy holder age

Model important interactions

PP: Select detail level
PP: Review potential risk drivers
PP: Select groups for each risk driver
PP: Select large claims strategy
PP: identify potential interactions
PP: construct final model
Price assessment



How to evaluate a regression model

- QQplot
- Akaike Information Criterion (AIC)
- Scaled deviance
- Cross validation
- Type 3 analysis
- Results interpretations

QQplot

- The quantiles of the model are plotted against the quantiles from a standard Normal distribution
- If the model is good, the point in the QQ plot will lie approximately on the line y=x

AIC

- Akaike Information criterion is a measure of goodness of fit that balances model fit against model simplicity
- AIC has the form
  - AIC=-2LL+2p where
- LL is the log likelihood evaluated at the value fo the estimated parameters
- p is the number of parameters estimated in the model
- AIC is used to compare model alternatives m_1 and m_2
- If AIC of m_1 is less than AIC of m_2 then m_1 is better than m_2

Scaled deviance

- Scaled deviance = 2(l(y,y)-l(y,muh)) where
  - l(y,y) is the maximum achievable log likelihood and l(y,muh) is the log likelihood at the maximum estimates of the regression parameters
- Scaled deviance is approximately distributed as a Chi Square random variables with n-p degrees of freedom
- Scaled deviance should be close to 1 if the model is good

Cross validation

Estimation

Validation

- Model is calibrated on for example 50% of the portfolio

- Model is then validated on the remaining 50% of the portfolio
- For a good model that predicts well there should not be too much difference between the modelled number of claims in a group and the observed number of claims in the same group

Type 3 analysis

- Does the degree of variation explained by the model increase significantly by including the relevant explanatory variable in the model?
- Type 3 analysis tests the fit of the model with and without the relevant explanatory variable
- A low p value indicates that the relevant explanatory variable improves the model significantly

Results interpretation

- Is the intercept reasonable?
- Are the parameter estimates reasonable compared to the results of the one way analyses?

How to use a regression model

- Smoothing of estimates