Non-life insurance mathematics

Nils F. Haavardsson, University of Oslo and DNB Skadeforsikring

Overview

Important issues	Models treated	Curriculum	Duration (in lectures)
What is driving the result of a non-life insurance company?	insurance economics models	Lecture notes	0,5
How is claim frequency modelled?	Poisson, Compound Poisson and Poisson regression	Section 8.2-4 EB	1,5
How can claims reserving be modelled?	Chain ladder, Bernhuetter Ferguson, Cape Cod,	Note by Patrick Dahl	2
How can claim size be modelled?	Gamma distribution, log-normal distribution	Chapter 9 EB	2
How are insurance policies priced?	Generalized Linear models, estimation, testing and modelling. CRM models.	Chapter 10 EB	2
Credibility theory	Buhlmann Straub	Chapter 10 EB	1
Reinsurance		Chapter 10 EB	1
Solvency		Chapter 10 EB	1
Repetition			1

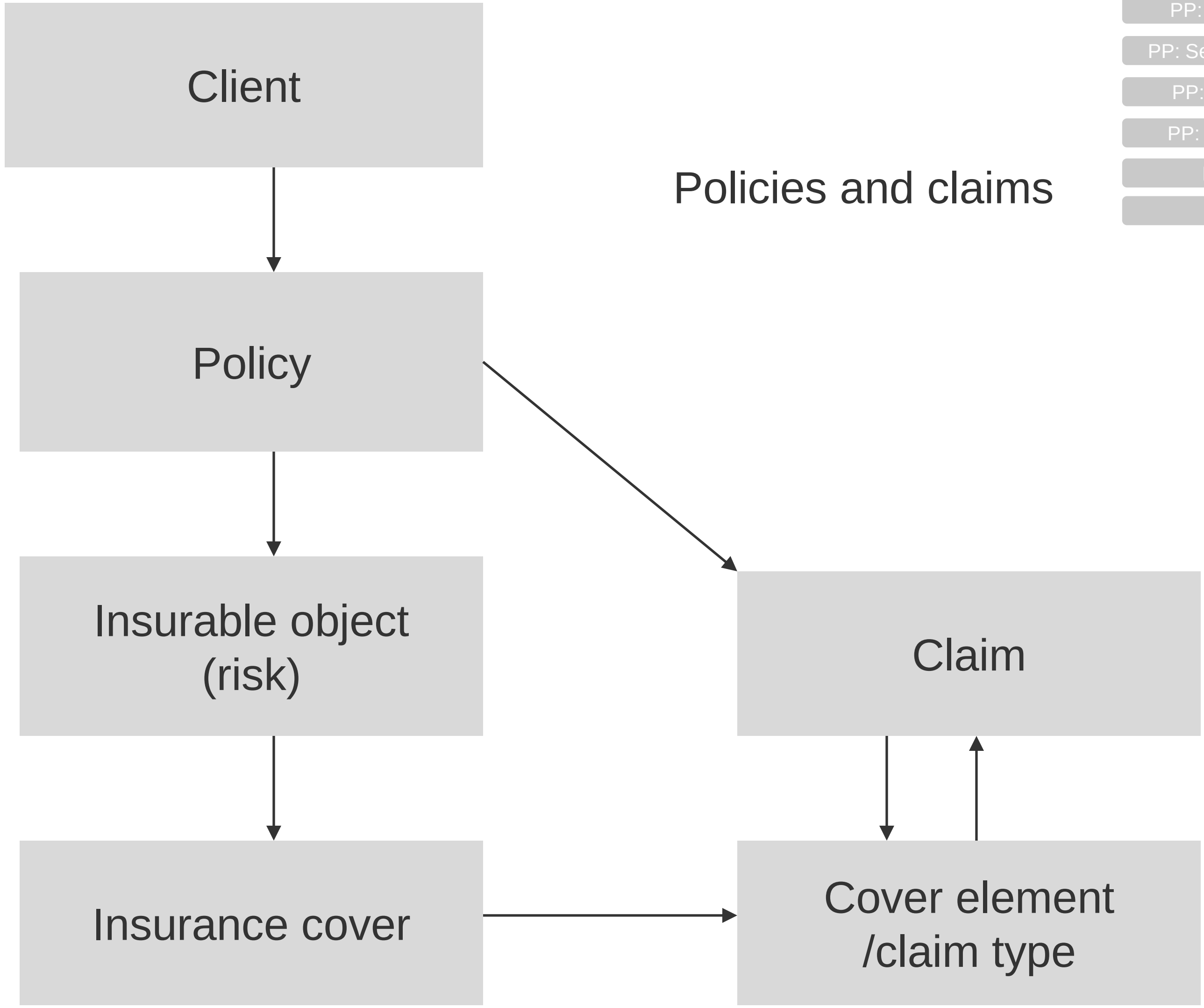
The pure premium is fundamental in pricing

Pure premium: select detail level	4 min
Pure premium: review potential risk drivers	17 min
Pure premium: select groups for each risk driver	5 min
Pure premium: Select large claims strategy for claim size	10 min
Pure premium: identify potential interactions	8 min
Pure premium: construct final model	22 min
Price assessment: take other considerations into account	14 min

The pure premium is fundamental in pricing

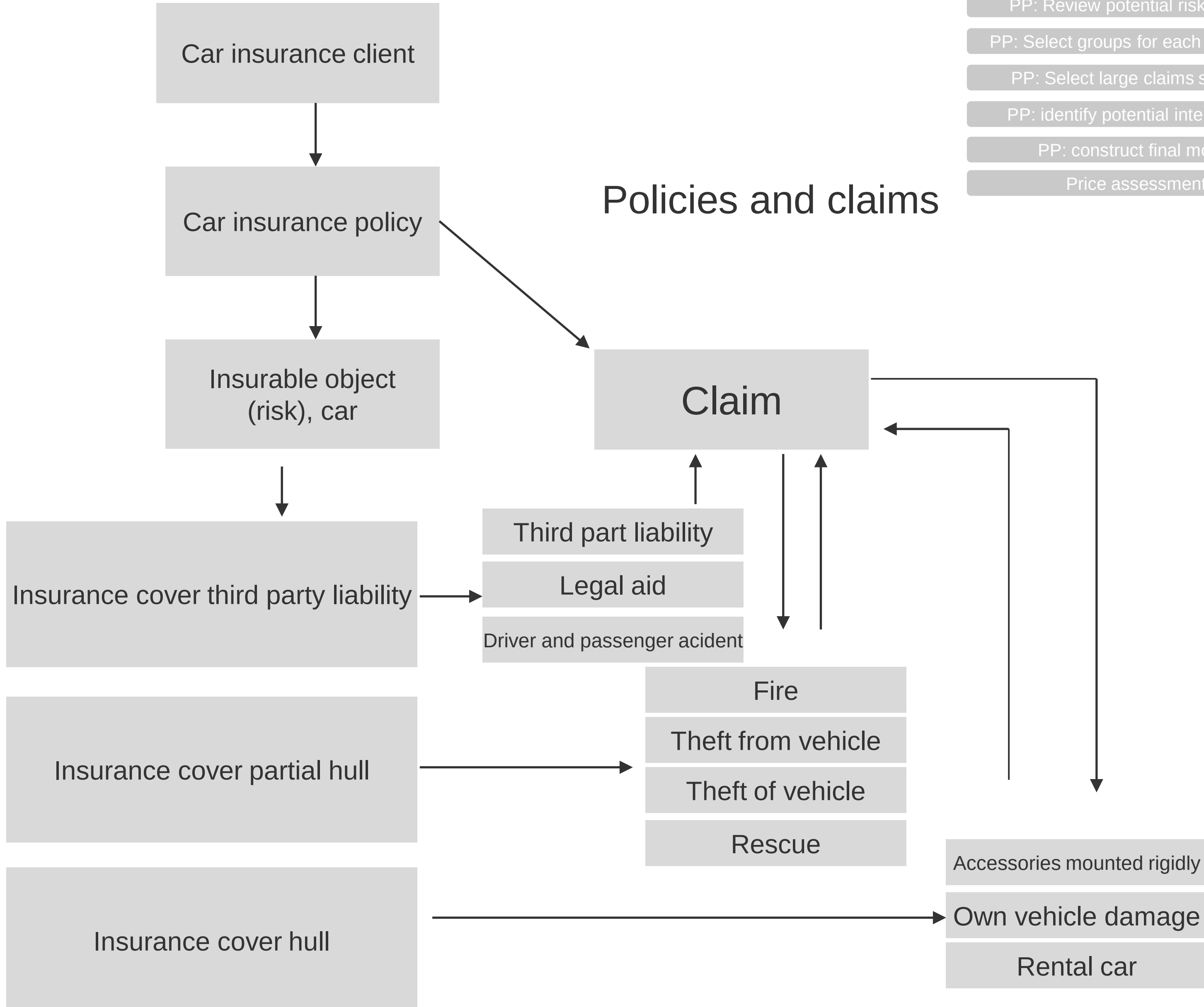
- Pure premium: select detail level
- Pure premium: review potential risk drivers
- Pure premium: select groups for each risk driver
- Pure premium: Select large claims strategy for claim size
- Pure premium: identify potential interactions
- Pure premium: construct final model
- Price assessment: take other considerations into account

Policies and claims



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

Policies and claims



Selection of detail level

- What is the ambition of the pure premium model?
- What is the price strategy of the company?
- What resources are available?
- What data are available?
- How long is the model intended to be operational?
- What characterize risk drivers at different levels of detail?
- Is the tail different at different levels?
- If high level of detail is selected, how should aggregation be done?

Review potential risk drivers

Object

Geography

Subject

- Standard
  - Electrical systems
  - Pipes
  - Roof
- Building type (concrete, tree etc)
- Size
- Building year
- Refurbishment status
- Electrical systems reviewed?

- Weather
- Climate
- Population density
- Infrastructure complexity
- Natural catastrophes
- Demography

- Policy holder
  - Age
  - Occupation
  - Risk averseness
  - Neatness
- Number of inhabitants in house
- Nature of use
  - Residence
  - Rent
  - Vacation home (i.e., inherited)

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

Review potential risk drivers

Object

Insurable object (risk)

Insurance cover

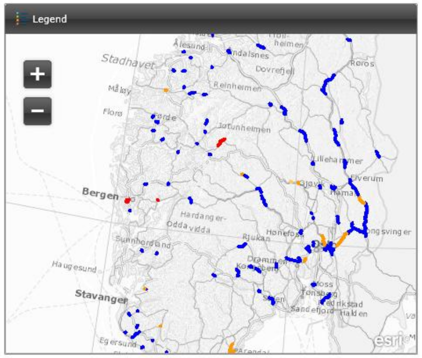
Fire

- Cumulus

A B C

Water

- Likelihood of flood



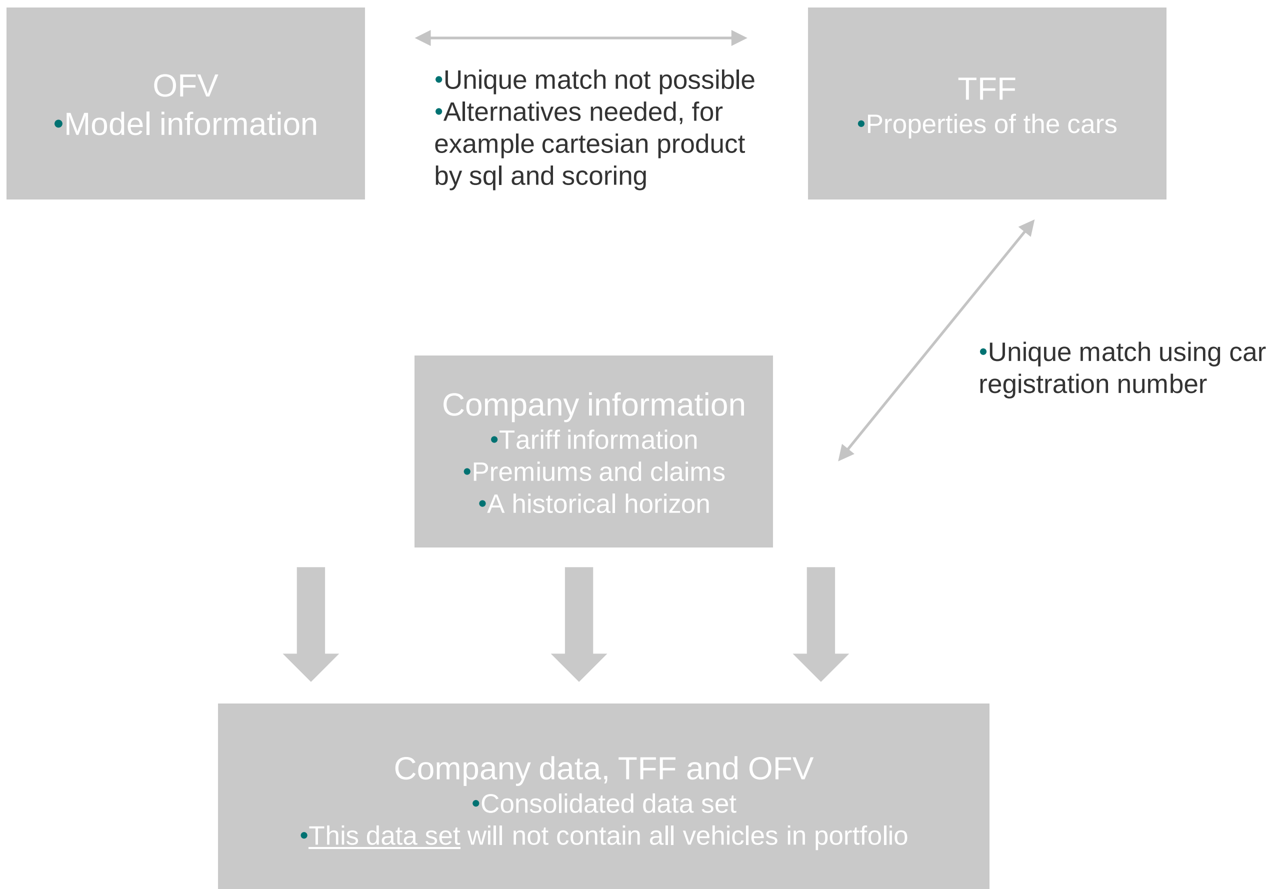
Review potential risk drivers

- What data sources are available in-house?
- What data sources are publically available?
- What data sources can be purchased?
- What is the quality of the data?
- Are there maps publically available or purchasable?

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

Data from different sources must be matched

Example from car insurance



Data set construction

- How long time horizon is available?
- How long time horizon is representative?
- How has the portfolio changed during the time that has passed?
- Is this change important?
- Has the portfolio grown substantially in the chosen time horizon?
- Is there objects that should be disregarded?
- Has there been unwanted risk inflow or outflow in the period?
  - Assume some bad apples were in the portfolio for some time
  - Assume that these are now removed
  - Assume that the company routines have been improved, making a revert unlikely
  - Should these bad apples then be removed before the analysis begins?
- Are there periods with untypical behavour? (Example: frost 2010)
  - How should these be treated?
- The final data set used in the analysis will by many players be considered as the truth (although we know better)

PP: Select detail level

PP: Review potential risk drivers

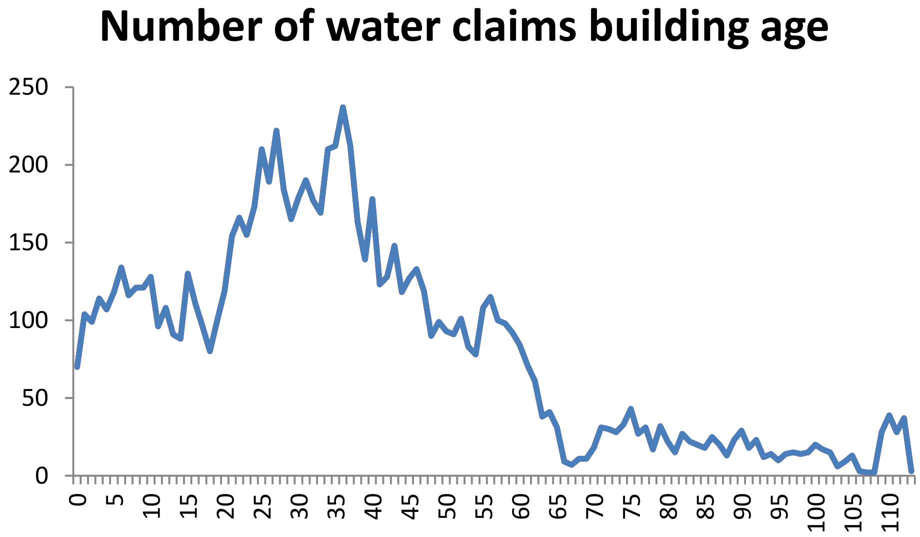
PP: Select groups for each risk driver

PP: Select large claims strategy

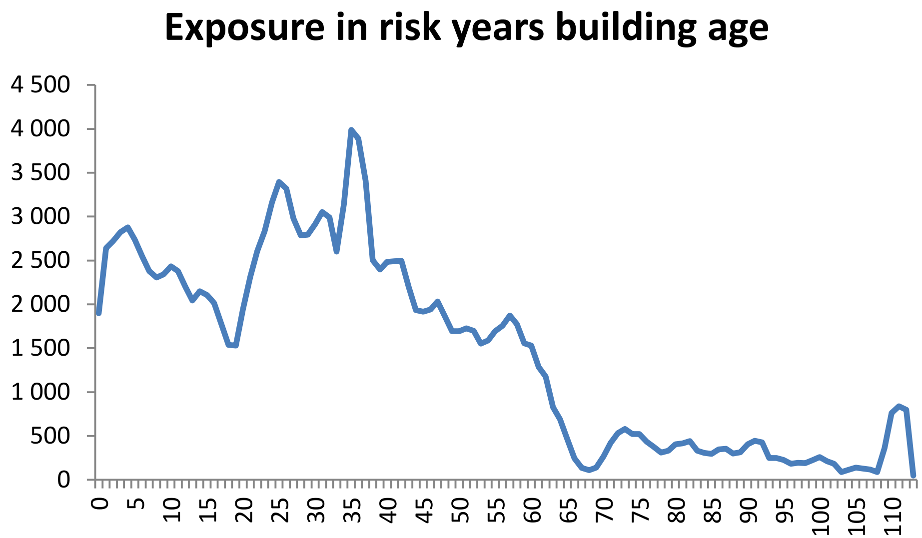
PP: identify potential interactions

PP: construct final model

Price assessment



- 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



PP: Select detail level

PP: Review potential risk drivers

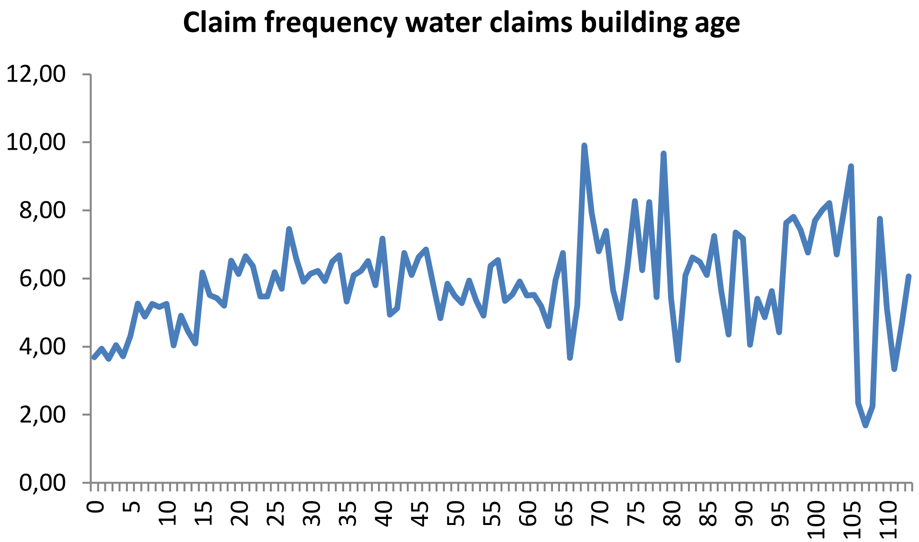
PP: Select groups for each risk driver

PP: Select large claims strategy

PP: identify potential interactions

PP: construct final model

Price assessment



PP: Select detail level

PP: Review potential risk drivers

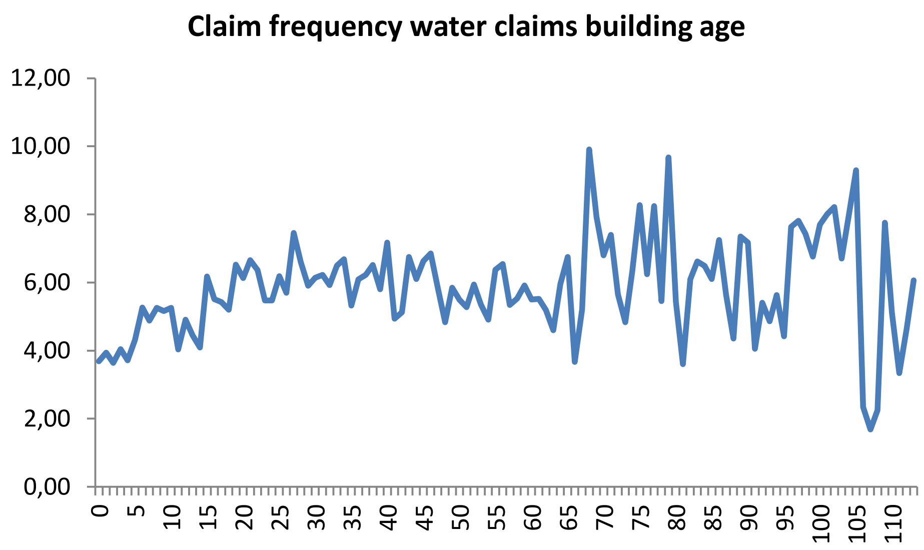
PP: Select groups for each risk driver

PP: Select large claims strategy

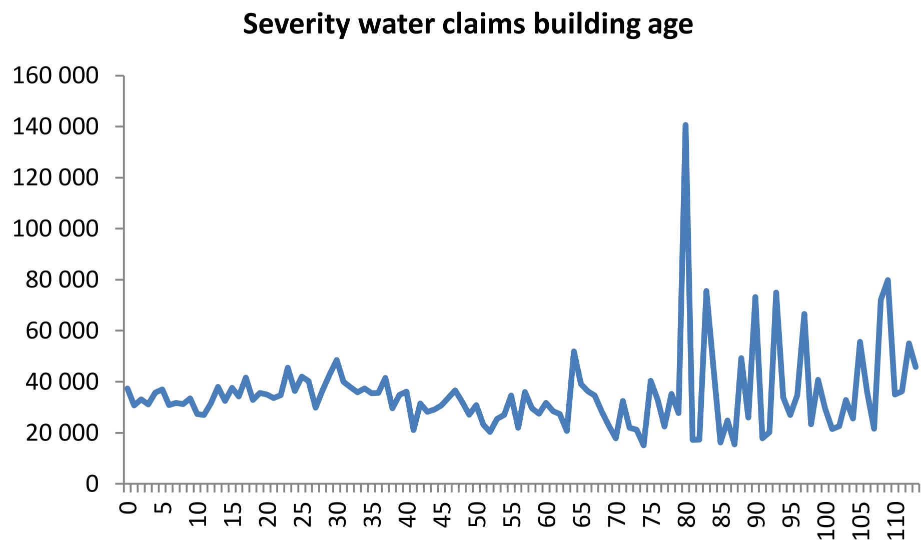
PP: identify potential interactions

PP: construct final model

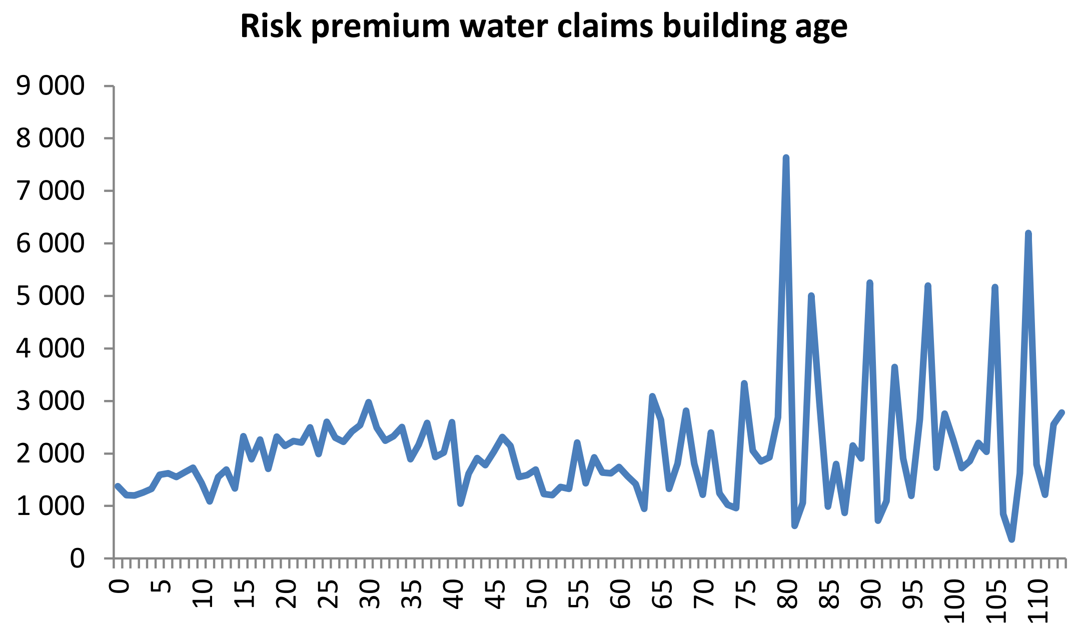
Price assessment



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



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



PP: Select detail level

PP: Review potential risk drivers

PP: Select groups for each risk driver

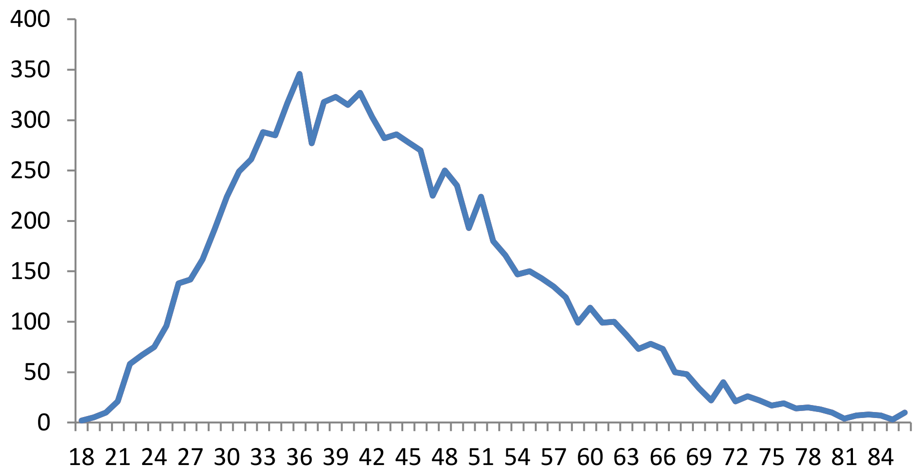
PP: Select large claims strategy

PP: identify potential interactions

PP: construct final model

Price assessment

Number of water claims policy holder age



PP: Select detail level

PP: Review potential risk drivers

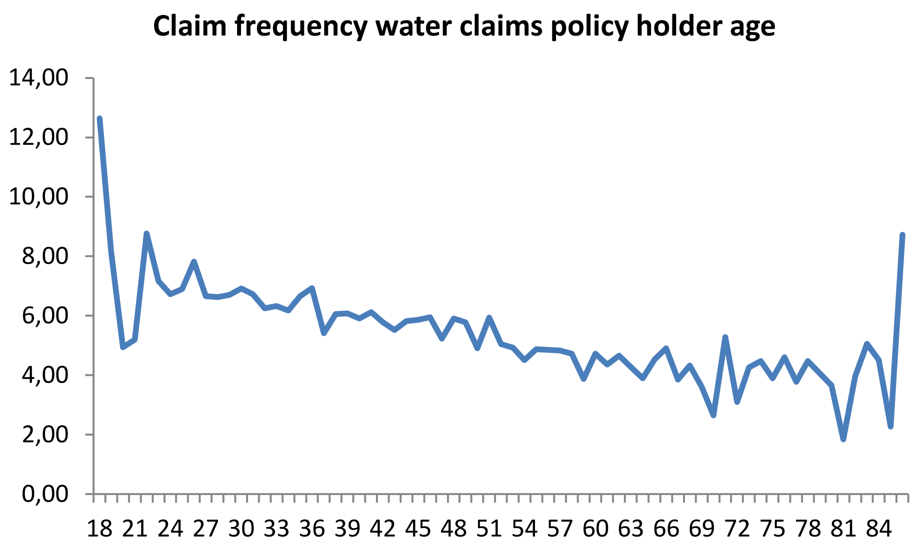
PP: Select groups for each risk driver

PP: Select large claims strategy

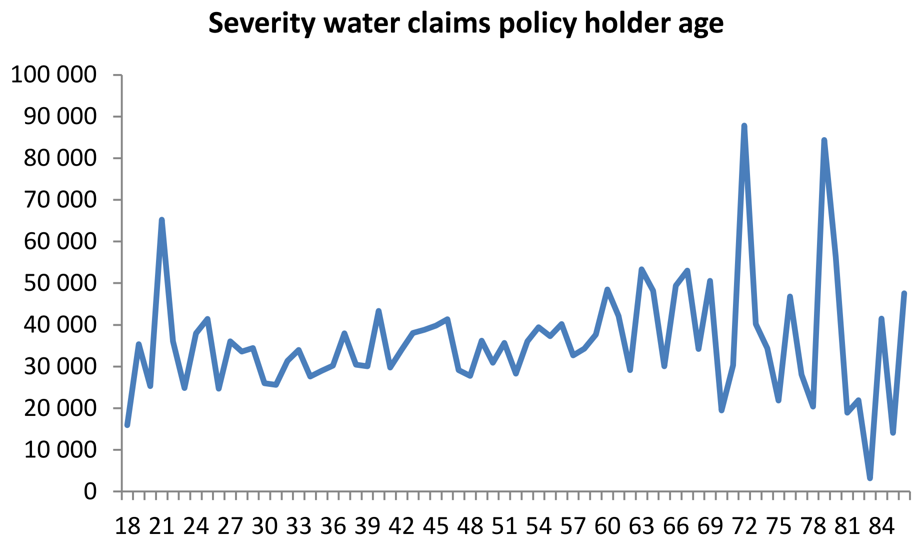
PP: identify potential interactions

PP: construct final model

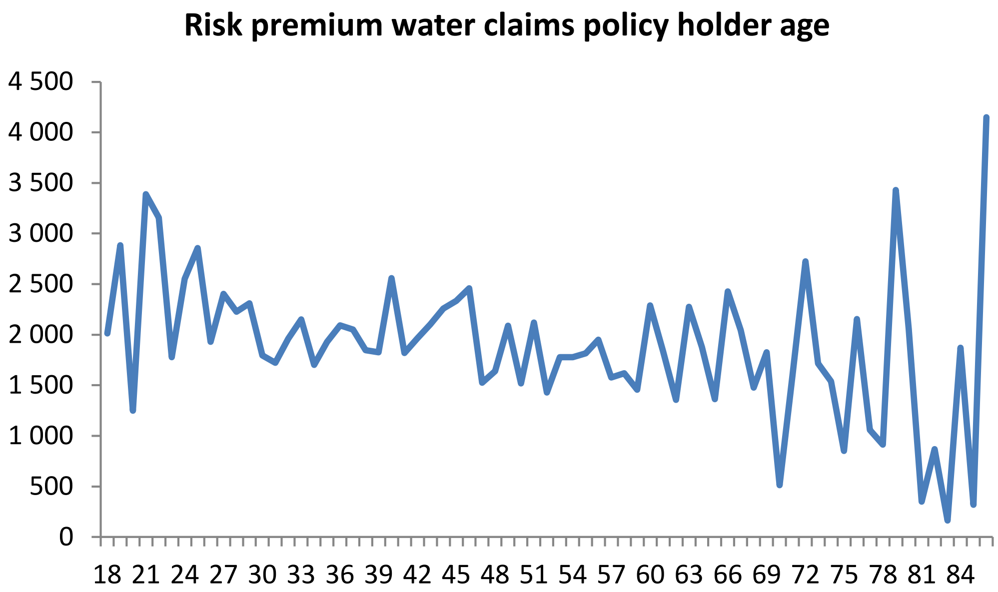
Price assessment



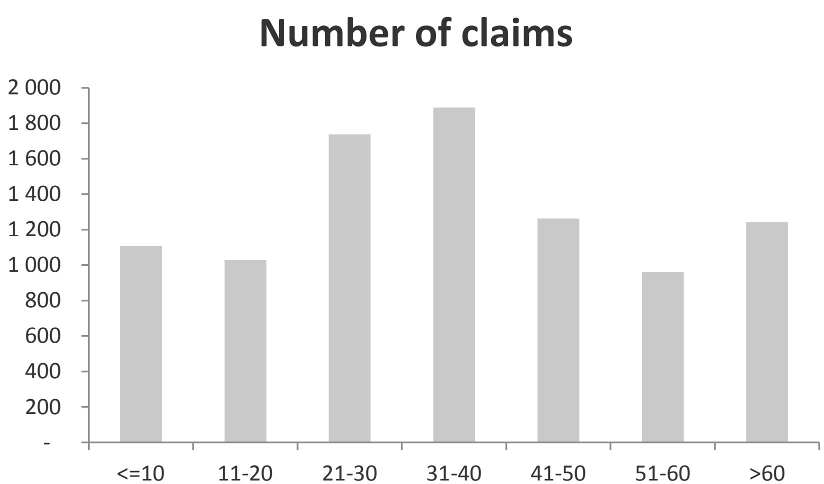
PP: Select detail level
PP: Review potential risk drivers
PP: Select groups for each risk driver
PP: Select large claims strategy
PP: identify potential interactions
PP: construct final model
Price assessment

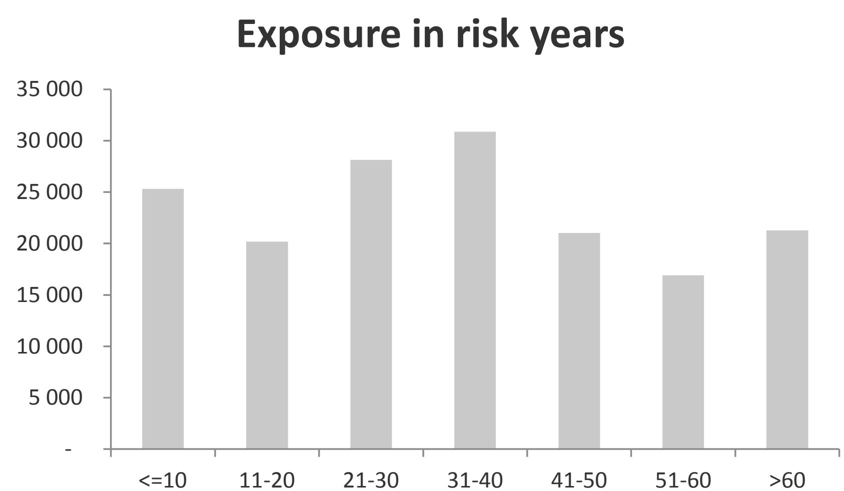


PP: Select detail level
PP: Review potential risk drivers
PP: Select groups for each risk driver
PP: Select large claims strategy
PP: identify potential interactions
PP: construct final model
Price assessment

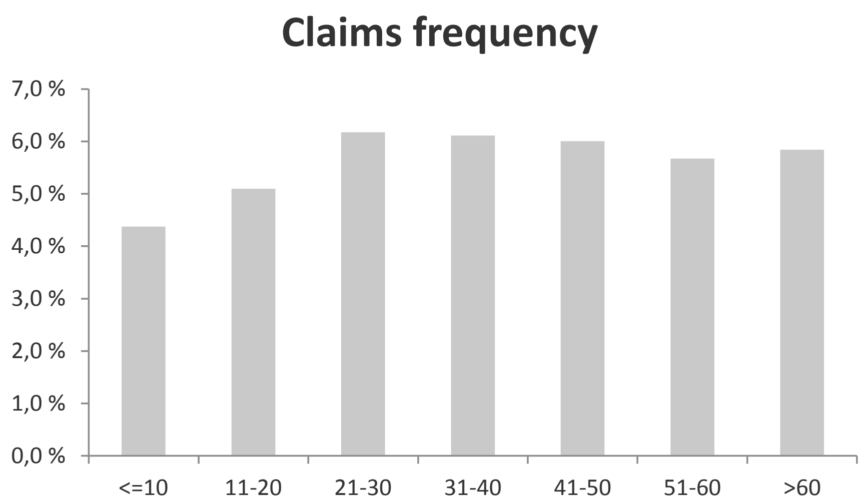


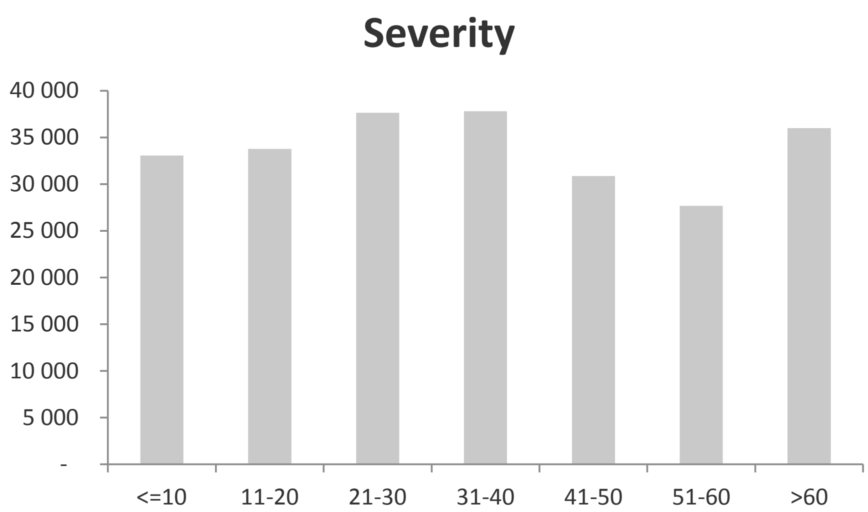
- The groups should be quite big (not too few observed claims in a group)
- There should not be too many groups
- The size of the different groups should not differ too much
- If the model contains many variables, this possibly affects the number of groups. The number of groups should then attempted to be reduced. The oposite should be considered if the model contains few variables.
- A continous function (typically smooth) could also be considered.
- The graphs below shows a grouping for building age



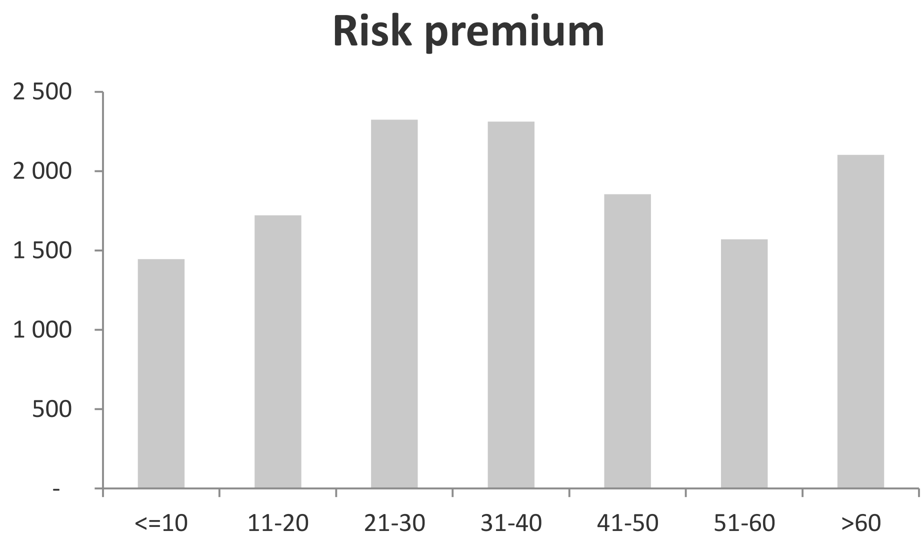


- The groups should be quite big (not too few observed claims in a group)
- There should not be too many groups
- The size of the different groups should not differ too much
- If the model contains many variables, this possibly affects the number of groups. The number of groups should then attempted to be reduced. The oposite should be considered if the model contains few variables.
- A continous function (typically smooth) could also be considered.
- The graphs below shows a grouping for building age

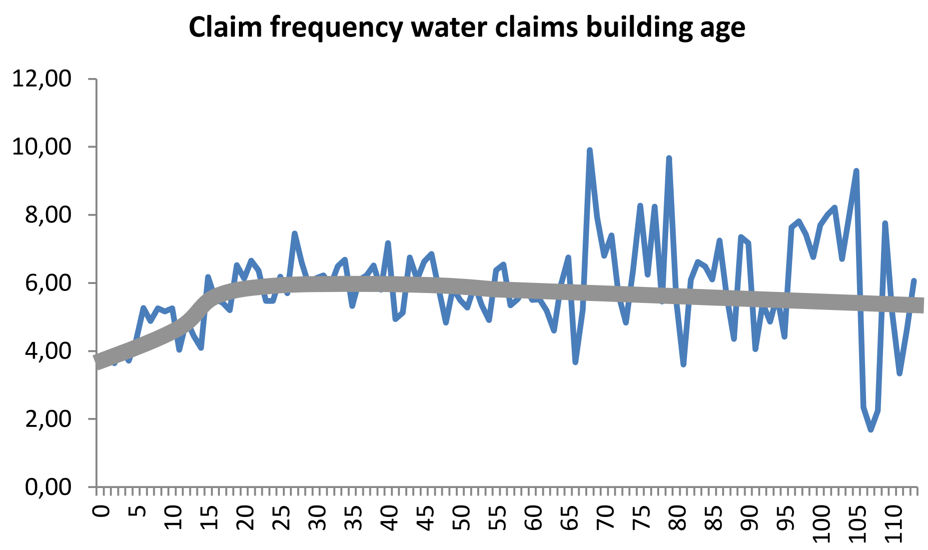




- The groups should be quite big (not too few observed claims in a group)
- There should not be too many groups
- The size of the different groups should not differ too much
- If the model contains many variables, this possibly affects the number of groups. The number of groups should then attempted to be reduced. The oposite should be considered if the model contains few variables.
- A continous function (typically smooth) could also be considered.
- The graphs below shows a grouping for building age



- The groups should be quite big (not too few observed claims in a group)
- There should not be too many groups
- The size of the different groups should not differ too much
- If the model contains many variables, this possibly affects the number of groups. The number of groups should then attempted to be reduced. The oposite should be considered if the model contains few variables.
- A continous function (typically smooth) could also be considered.
- The graphs below shows a grouping for building age



- What do to with the large claims in the claim size regression?
- Example:
  - Drammen had 3 large fires in 2010
  - We choose to believe that these occured at random (i.e., none of these were bad apples etc)
  - Should the residents of Drammen be punished for these fires in their insurance premium?
- Suggestion:
- We disregard large claims in the data when regression coefficients are estimated
- The proportion of large claims are distributed to all policies as a fixed add on (in percent) – this could be thought of as «large claims premium»
- What is a suitable large claim threshold?
- At what detail level should the large claim threshold be set? (client? Or policy? Or cover?)

PP: Select detail level

PP: Review potential risk drivers

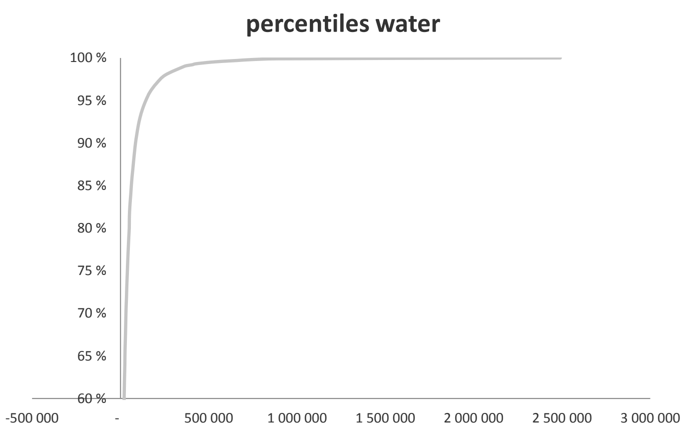
PP: Select groups for each risk driver

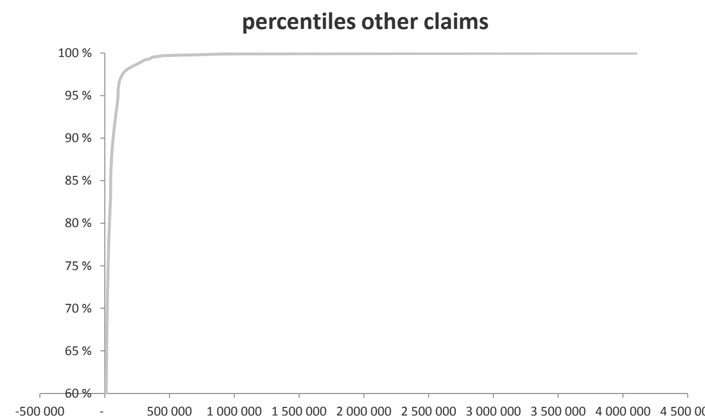
PP: Select large claims strategy

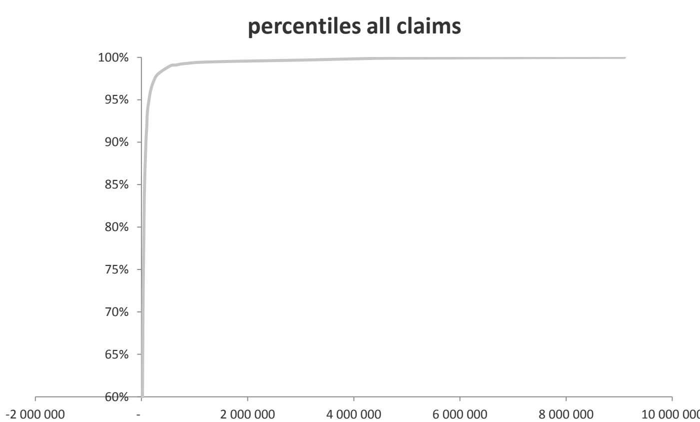
PP: identify potential interactions

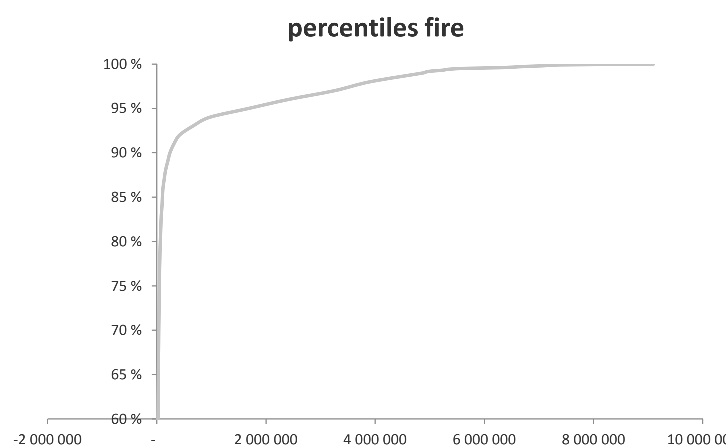
PP: construct final model

Price assessment









PP: Select detail level

PP: Review potential risk drivers

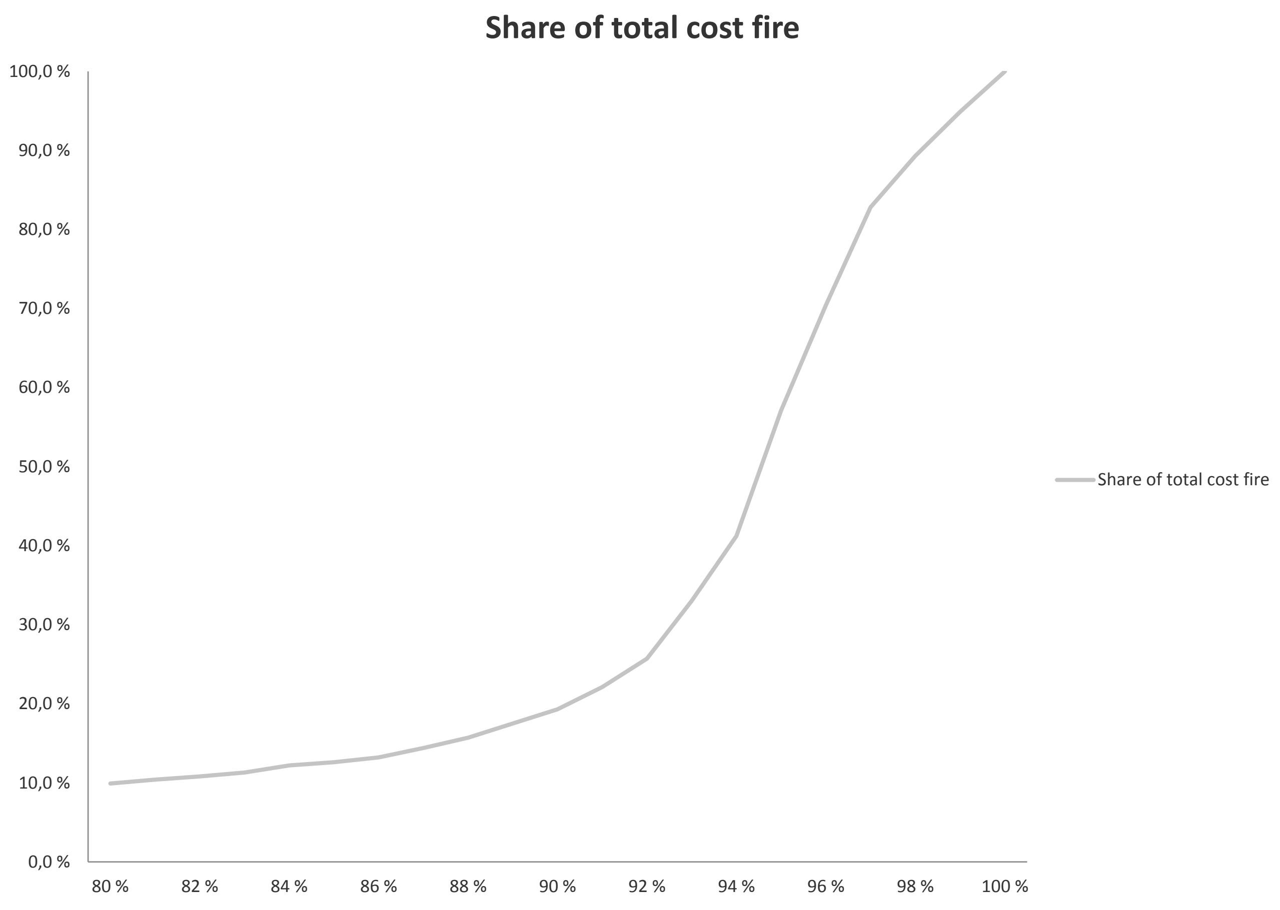
PP: Select groups for each risk driver

PP: Select large claims strategy

PP: identify potential interactions

PP: construct final model

Price assessment



PP: Select detail level

PP: Review potential risk drivers

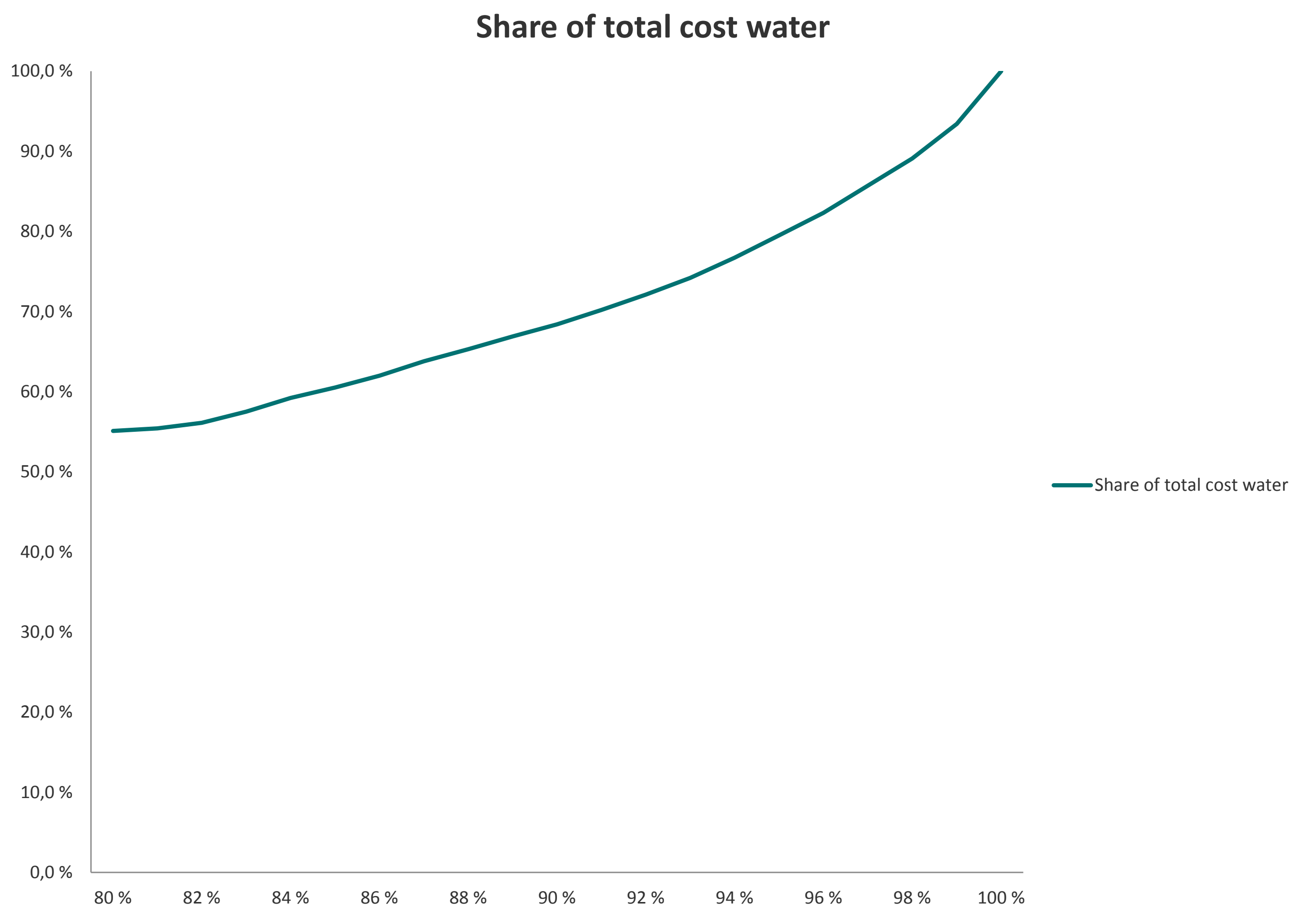
PP: Select groups for each risk driver

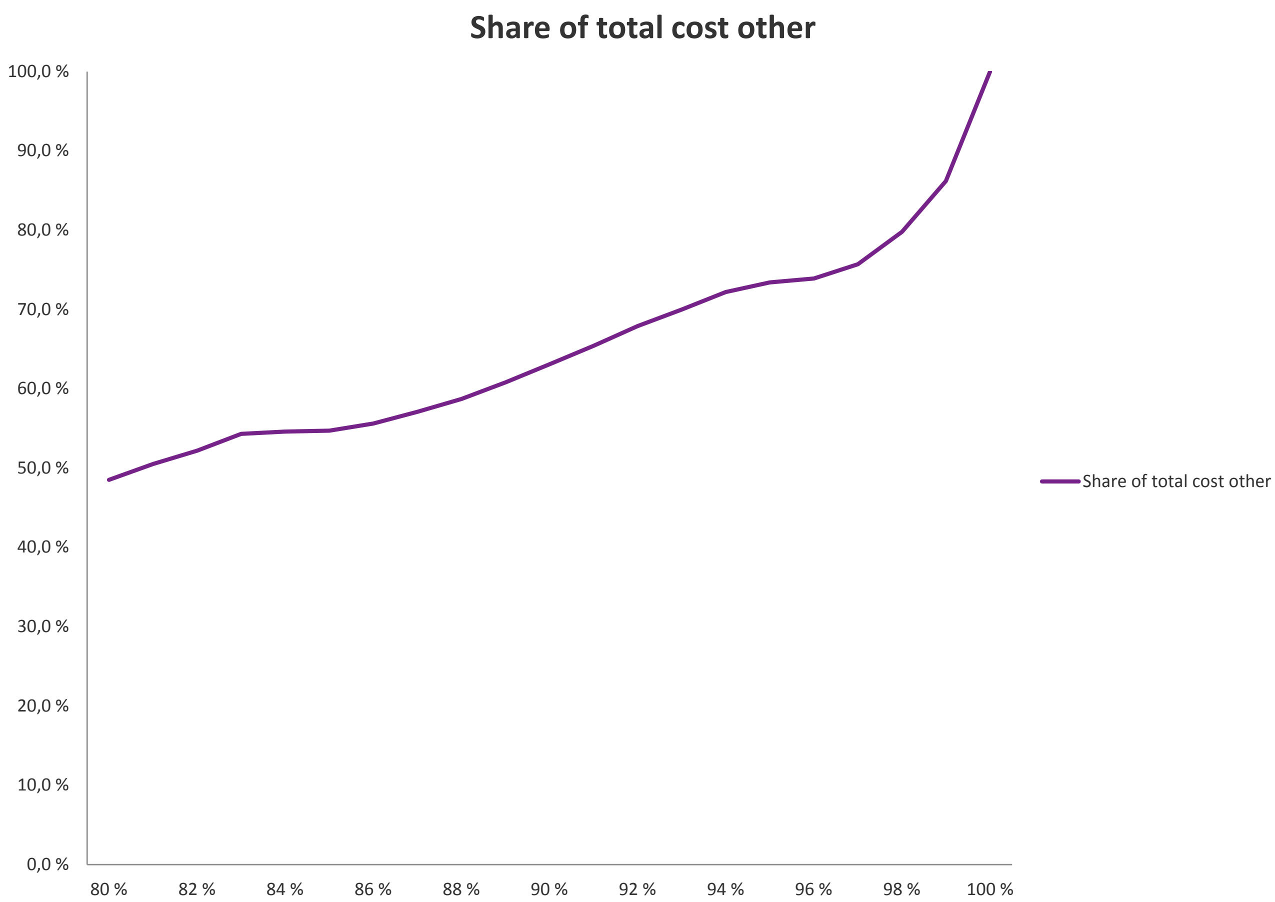
PP: Select large claims strategy

PP: identify potential interactions

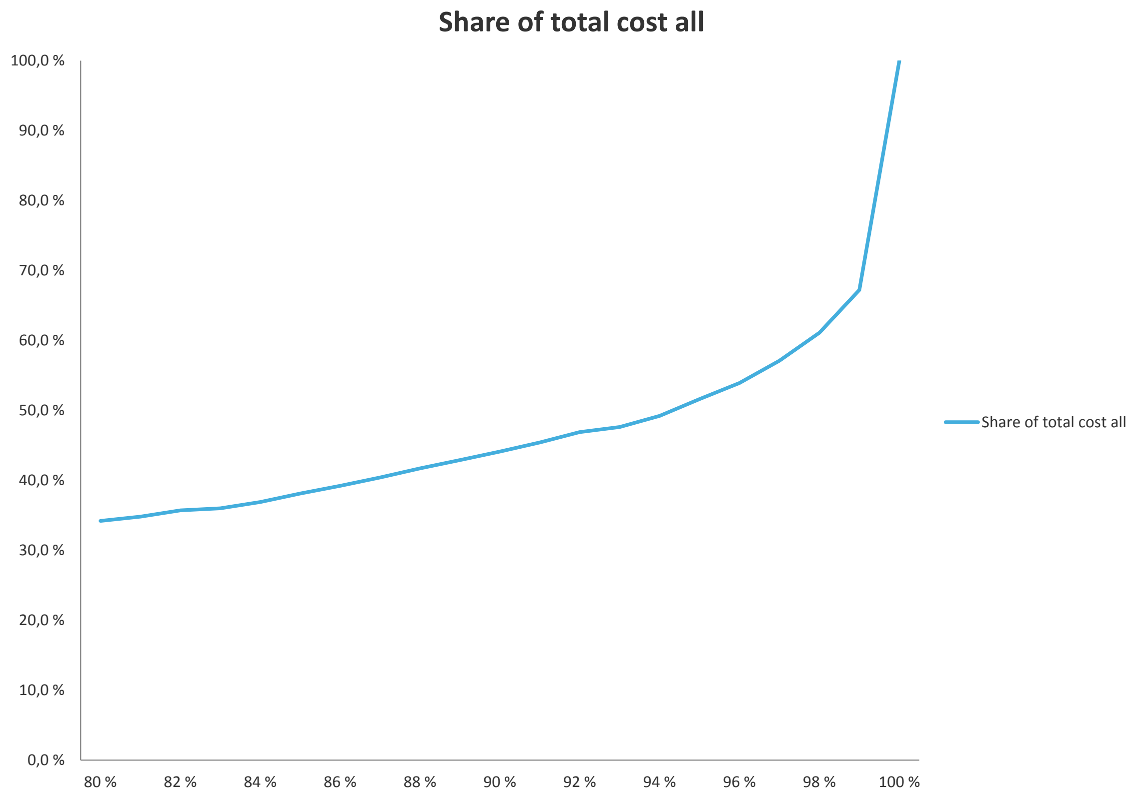
PP: construct final model

Price assessment




Share of total cost other
100,0 %
90,0 %
80,0 %
70,0 %
60,0 %
50,0 %
40,0 %
30,0 %
20,0 %
10,0 %
0,0 %
80 %
82 %
84 %
86 %
88 %
90 %
92 %
94 %
96 %
98 %
100 %
Share of total cost other

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



PP: Select detail level

PP: Review potential risk drivers

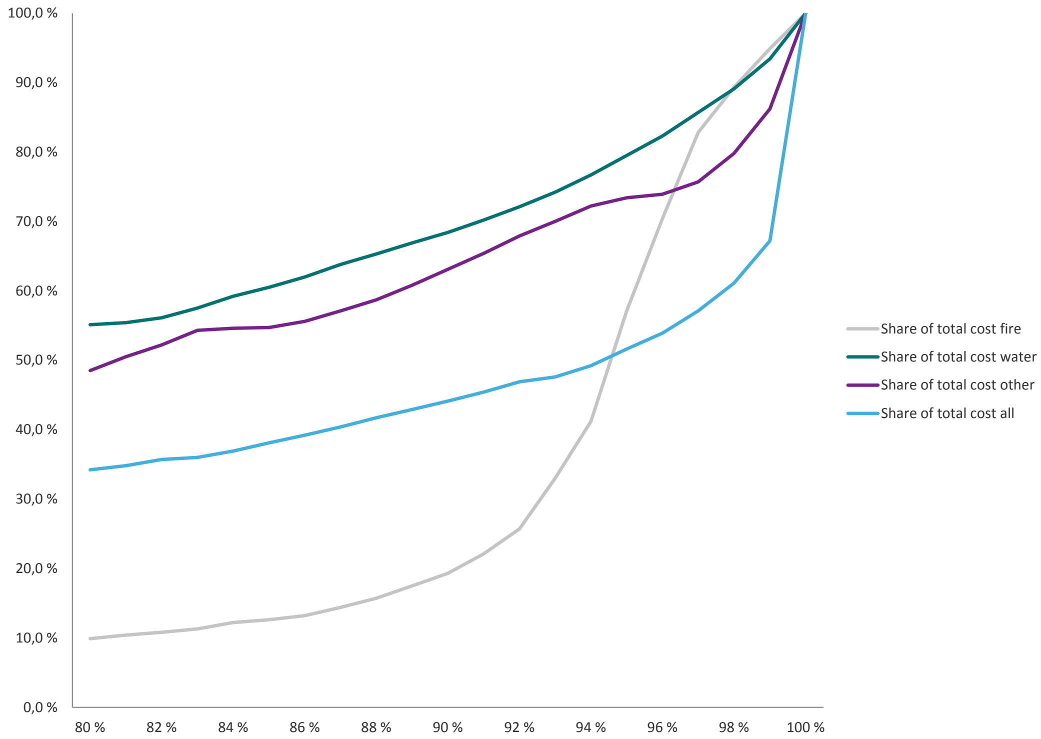
PP: Select groups for each risk driver

PP: Select large claims strategy

PP: identify potential interactions

PP: construct final model

Price assessment



	fire	water	other	all
99,1 %	4 927 015	377 682	300 000	656 972
99,2 %	5 021 824	406 859	307 666	726 909
99,3 %	5 226 985	424 013	344 006	871 552
99,4 %	5 332 034	464 769	354 972	1 044 598
99,5 %	5 576 737	511 676	365 925	1 510 740
99,6 %	6 348 393	576 899	409 618	2 330 786
99,7 %	6 647 669	663 382	462 719	3 195 813
99,8 %	7 060 421	740 187	724 682	3 832 486
99,9 %	7 374 623	891 226	1 005 398	4 733 953
100,0 %	9 099 312	2 490 558	4 100 000	9 099 312

- 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

PP: Select detail level

PP: Review potential risk drivers

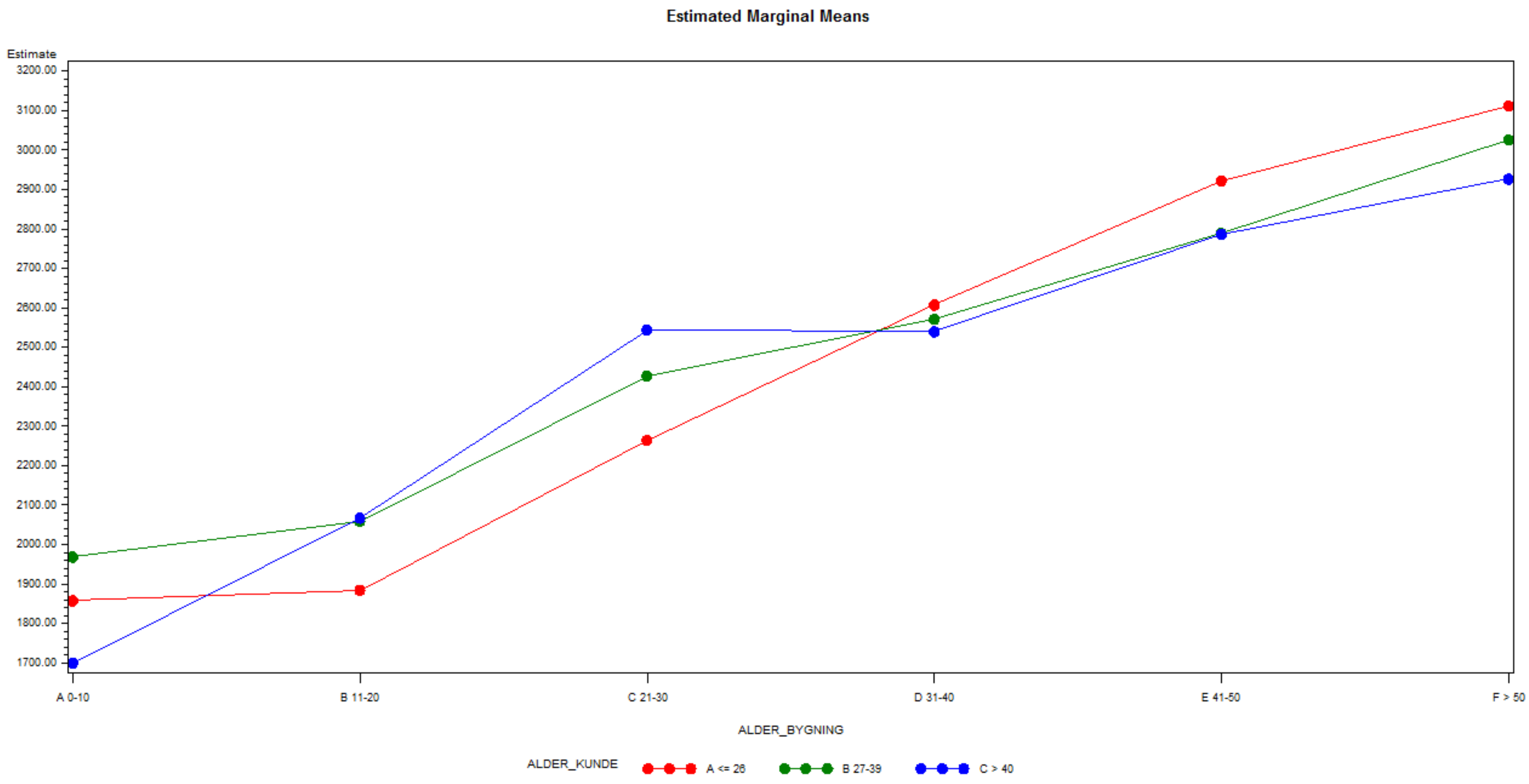
PP: Select groups for each risk driver

PP: Select large claims strategy

PP: identify potential interactions

PP: construct final model

Price assessment

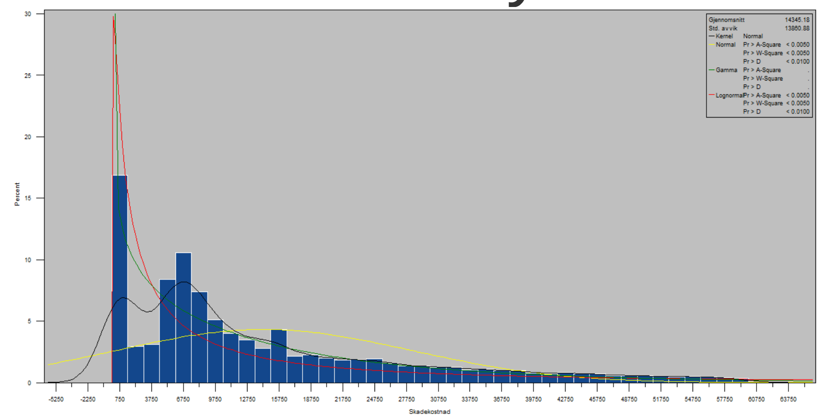


General guidelines

- Start performing marginal analysis of response against each candidate explanatory variable
- If the response does not vary with the candidate explanatory variable, it is high on the discard list
- If two candidate explanatory variables are sufficiently correlated one of them should be discarded. Choose the one least correlated with the other candidate explanatory variables
- Model principle: *Occam's razor:* Things should not be multiplied unnecessarily
- We do not want a model that fits the data perfectly, but one that is good at predicting next year's outcome
- Assume you consider adding a new variable X in your model: Does the model explain more of the variation in pure premium by introducing X?
- Here we only discuss very briefly some topics of Model selection
- The course STK 4160 - Statistical Model Selection is highly recommended!!

Example: car insurance

- Hull coverage (i.e., damages on own vehicle in a collision or other sudden and unforeseen damage)
- Time period for parameter estimation: 2 years
- Covariates initially:
  - Driving length
  - Car age
  - Region of car owner
  - Category of vehicle
  - Bonus of insured vehicle
  - Age of user
- Model for frequency and severity are fitted.



Results for frequency

Model	Akaike	BIC	Difference in % between estimated (calibrated) portfolio and predicted portfolio at total level	Difference in % between estimated (calibrated) portfolio and predicted portfolio p10	Difference in % between estimated (calibrated) portfolio and predicted portfolio p50	Difference in % between estimated (calibrated) portfolio and predicted portfolio p90
Age of car, age of user, bonus, region, category of vehicle, driving length	5476	5610	6,51	5,02	8,03	86,1
Age of car, age of user, bonus, region, driving length	4506	4606	2,01	1,98	5,18	13,37
Age of car, bonus, region, driving length	3062	3122	0,55	0,54	2,33	9,36

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

Frequency

Variables	Class	Model	Portfolio	Diff.
Age of car	Total	12 234	12 302	0,55
Age of car	A <= 5	5 614	5 453	2,96
Age of car	B 6-10	4 739	4 762	0,49
Age of car	C 11-15	1 647	1 621	1,61
Age of car	D > 15	234	466	49,80
CurrNCD_Cd	Total	12 234	12 302	0,55
CurrNCD_Cd	A < 70%	2 319	2 576	9,96
CurrNCD_Cd	B 70%	1 892	1 833	3,19
CurrNCD_Cd	C 75%	8 023	7 893	1,65
Region	Total	12 234	12 302	0,55
Region	Agder	466	460	1,38
Region	Akershus Østfold	2 988	2 920	2,33
Region	Buskerud Hedmark Oppland	1 843	1 887	2,34
Region	Hordaland	981	942	4,12
Region	M og R Rogaland S og F	1 184	1 148	3,13
Region	Nord	1 942	1 920	1,14
Region	Oslo	1 844	2 021	8,77
Region	Telemark Vestfold	986	1 004	1,75
DriveLength	Total	12 234	12 302	0,55
DriveLength	12000	3 559	3 589	0,83
DriveLength	16000	2 980	2 898	2,82
DriveLength	20000	2 123	2 053	3,39
DriveLength	25000	825	803	2,75
DriveLength	30000	545	509	7,16
DriveLength	8000	1 903	2 149	11,47
DriveLength	99999	300	301	0,43

Source	Num DF	Den DF	F Value	Pr > F	Chi-square	Pr>Chi-sq	Method
ALDER_BIL	3	487	5,68	0,0008	17,05	0,0007	LR
CurrNCD_Cd	2	487	12,58	<.0001	25,16	<.0001	LR
KundeFylkeNavn	7	487	2,57	0,0131	17,99	0,012	LR
Side1Verdi4	6	487	11,29	<.0001	67,73	<.0001	LR

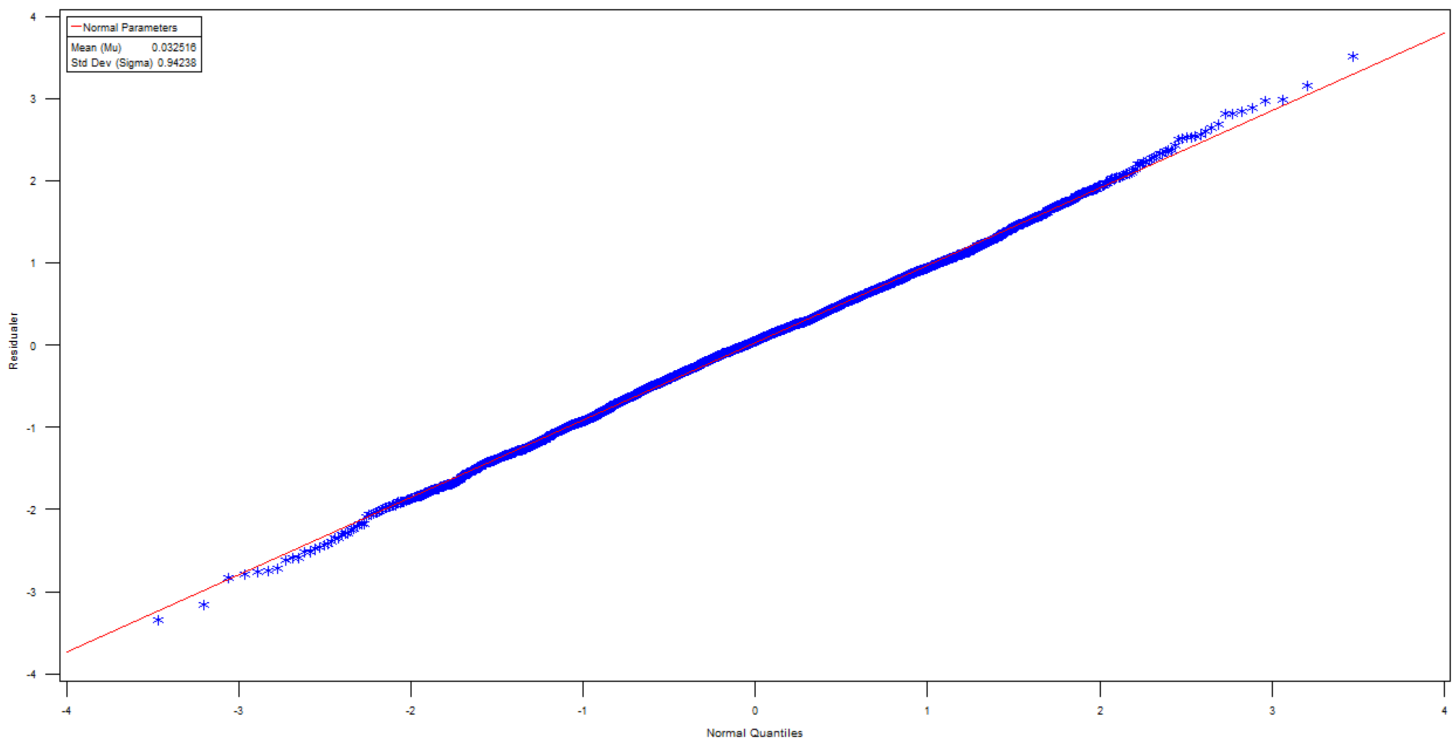
Criterion	Deg. fr.	Value	Value/DF
Deviance	487	911,0755	1,8708
Scaled Deviance	487	221,7220	0,4553
Pearson Chi-Square	487	2 001,1265	4,1091
Scaled Pearson X2	487	487,0000	1,0000
Log Likelihood	_	8 512,5578	_
Full Log Likelihood	_	- 1 512,1335	_
AIC (smaller is better)	_	3 062,2671	_
AICC (smaller is better)	_	3 063,8308	_
BIC (smaller is better)	_	3 142,5712	_

- 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

Frequency

Variables	Class	Estimate	Std. Err.	Confidence Limit	Confidence	Chi-square	Pr>Chi-sq	Policy Years
Intercept		- 1,9487	0,0545	- 2,0556	- 1,8418	1 276,26	<.0001	.
ALDER_BIL	A <= 5	0,0000	0,0000	0,0000	0,0000	.	.	66 017
ALDER_BIL	B 6-10	- 0,0990	0,0403	- 0,1781	- 0,0200	6,03	0,0141	63 239
ALDER_BIL	C 11-15	- 0,1806	0,0584	- 0,2951	- 0,0661	9,55	0,002	25 066
ALDER_BIL	D > 15	0,2811	0,1516	- 0,0160	0,5781	3,44	0,0637	2 333
CurrNCD_Cd	A < 70%	0,2423	0,0484	0,1474	0,3372	25,04	<.0001	25 293
CurrNCD_Cd	B 70%	0,1072	0,0523	0,0047	0,2097	4,20	0,0405	23 328
CurrNCD_Cd	C 75%	0,0000	0,0000	0,0000	0,0000	.	.	108 033
KundeFylkeNavn	Agder	- 0,0759	0,1020	- 0,2759	0,1241	0,55	0,4568	6 282
KundeFylkeNavn	Akershus Østfold	0,0000	0,0000	0,0000	0,0000	.	.	37 135
KundeFylkeNavn	Oppland	- 0,0720	0,0603	- 0,1903	0,0462	1,43	0,2324	24 420
KundeFylkeNavn	Hordaland	- 0,0749	0,0748	- 0,2216	0,0718	1,00	0,3168	13 488
KundeFylkeNavn	M og R Rogaland S og F	- 0,0521	0,0701	- 0,1894	0,0853	0,55	0,4576	16 181
KundeFylkeNavn	Nord	0,0095	0,0595	- 0,1072	0,1261	0,03	0,8739	24 928
KundeFylkeNavn	Oslo	0,1625	0,0607	0,0435	0,2815	7,17	0,0074	20 417
KundeFylkeNavn	Telemark Vestfold	- 0,0915	0,0746	- 0,2378	0,0547	1,50	0,22	13 804
Side1Verdi4	12000	0,0000	0,0000	0,0000	0,0000	.	.	50 403
Side1Verdi4	16000	0,1233	0,0506	0,0241	0,2225	5,93	0,0148	37 415
Side1Verdi4	20000	0,2233	0,0564	0,1128	0,3339	15,68	<.0001	24 276
Side1Verdi4	25000	0,3530	0,0801	0,1961	0,5099	19,45	<.0001	8 241
Side1Verdi4	30000	0,4680	0,0950	0,2819	0,6542	24,30	<.0001	4 959
Side1Verdi4	8000	- 0,0756	0,0579	- 0,1891	0,0378	1,71	0,1911	28 934
Side1Verdi4	99999	0,5829	0,1251	0,3378	0,8280	21,72	<.0001	2 426

Results for frequency



Severity

- 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

Variables	Class	Model	Portfolio	Diff.
Age of car	Total	160 132 806	167 040 879	4,14
Age of car	A 0-2	28 433 215	28 243 815	0,67
Age of car	B 3-5	47 997 569	49 623 495	3,28
Age of car	C 6-10	62 122 992	59 271 047	4,81
Age of car	D 11-15	17 057 608	16 523 520	3,23
Age of car	E > 15	4 521 422	13 379 002	66,21
CurrNCD_Cd	Total	160 132 806	167 040 879	4,14
CurrNCD_Cd	A <70%	35 277 257	41 678 869	15,36
CurrNCD_Cd	B 70%	22 110 029	24 275 857	8,92
CurrNCD_Cd	C 75%	102 745 520	101 086 153	1,64
Drive Length	Total	160 132 806	167 040 879	4,14
Drive Length	A 8000	29 850 226	34 353 069	13,11
Drive Length	B 12000	44 293 334	46 007 077	3,72
Drive Length	C 16000	35 836 672	37 401 861	4,18
Drive Length	D >=20000	50 152 573	49 278 871	1,77
Region	Total	160 132 806	167 040 879	4,14
Region	Agder	6 915 540	6 244 053	10,75
Region	Akershus Østfold	36 706 077	34 709 242	5,75
Region	Oppland	25 612 549	25 227 169	1,53
Region	Hordaland	12 396 522	13 583 705	8,74
Region	M og R Rogaland S og F	14 653 460	15 056 007	2,67
Region	Nord	27 883 283	27 564 765	1,16
Region	Oslo	24 622 501	32 499 740	24,24
Region	Telemark Vestfold	11 342 873	12 156 196	6,69

Source	Deg. fr.	Chi-square	Pr>Chi-sq	Method
ALDER_BIL	4	17,99	0,0012	LR
CurrNCD_Cd	2	33,02	<.0001	LR
KundeFylkeNavn	7	23,84	0,0012	LR
Side1Verdi4	3	8,14	0,0431	LR

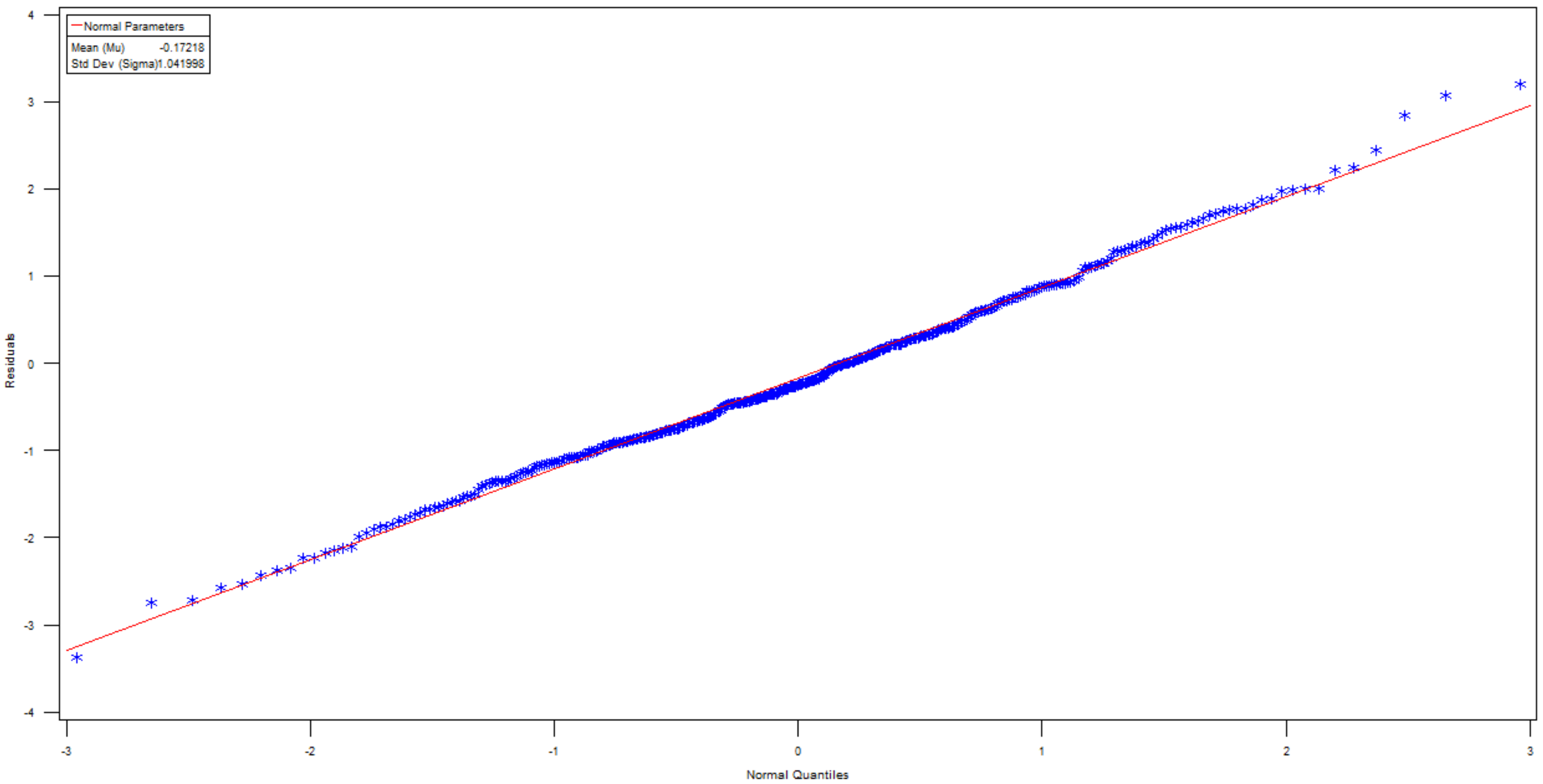
Criterion	Deg. fr.	Value	Value/DF
Deviance	389	682,6904	1,7550
Scaled Deviance	389	430,8548	1,1076
Pearson Chi-Square	389	728,0355	1,8716
Scaled Pearson X2	389	459,4726	1,1812
Log Likelihood	_	- 4 310,8979	_
Full Log Likelihood	_	- 4 310,8979	_
AIC (smaller is better)	_	8 657,7959	_
AICC (smaller is better)	_	8 659,5633	_
BIC (smaller is better)	_	8 729,9102	_

Severity

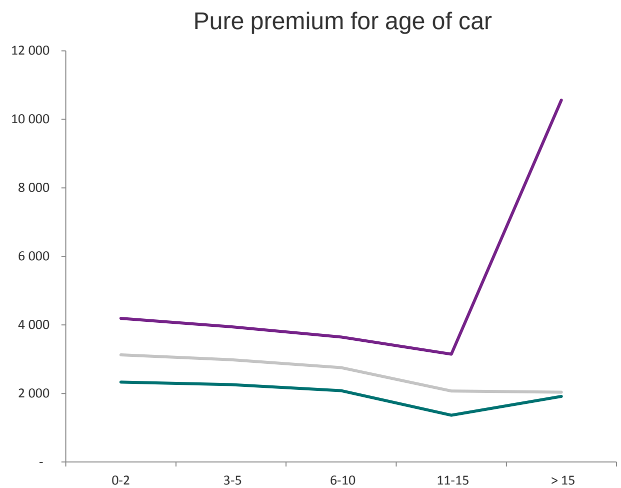
Variables	Class	Estimate	Std. Err.	dence Limit	onfidence Limit	Chi-square	Pr>Chi-sq	Policy Years
Intercept		9,9688	0,0482	9,8743	10,0633	42 728,47	<.0001	.
Age of car	A 0-2	0,0271	0,0468	- 0,0646	0,1188	0,34	0,562	24 274
Age of car	B 3-5	- 0,0186	0,0393	- 0,0957	0,0585	0,22	0,6358	41 744
Age of car	C 6-10	0,0000	0,0000	0,0000	0,0000	.	.	63 238
Age of car	D 11-15	- 0,2029	0,0514	- 0,3037	- 0,1021	15,57	<.0001	25 066
Age of car	E > 15	0,0500	0,1815	- 0,2463	0,4651	0,36	0,5467	2 333
CurrNCD_Cd	A <70%	0,2471	0,0436	0,1616	0,3326	32,08	<.0001	25 284
CurrNCD_Cd	B 70%	0,0201	0,0470	- 0,0721	0,1123	0,18	0,6693	23 331
CurrNCD_Cd	C 75%	0,0000	0,0000	0,0000	0,0000	.	.	108 039
Drive Length	A 8000	0,2099	0,0521	0,0077	0,2121	4,44	0,035	28 935
Drive Length	B 12000	0,0000	0,0000	0,0000	0,0000	.	.	50 404
Drive Length	C 16000	- 0,0354	0,0446	- 0,1227	0,0520	0,63	0,4276	37 415
Drive Length	D >=20000	- 0,0191	0,0422	- 0,1018	0,0636	0,20	0,6508	39 900
Region	Agder	0,1491	0,0874	- 0,0222	0,3203	2,91	0,088	6 282
Region	Akershus Østfold	0,0000	0,0000	0,0000	0,0000	.	.	37 135
Region	Oppland	0,0542	0,0509	- 0,0456	0,1540	1,13	0,2869	24 420
Region	Hordaland	0,0815	0,0695	- 0,0548	0,2178	1,37	0,2411	13 488
Region	M og R Rogaland S og F	0,1491	0,0605	0,0305	0,2677	6,07	0,0138	16 181
Region	Nord	0,1751	0,0511	0,0749	0,2752	11,73	0,0006	24 928
Region	Oslo	0,1218	0,0559	0,0122	0,2313	4,75	0,0294	20 417
Region	Telemark Vestfold	- 0,0868	0,0634	- 0,2112	0,0376	1,87	0,1713	13 804

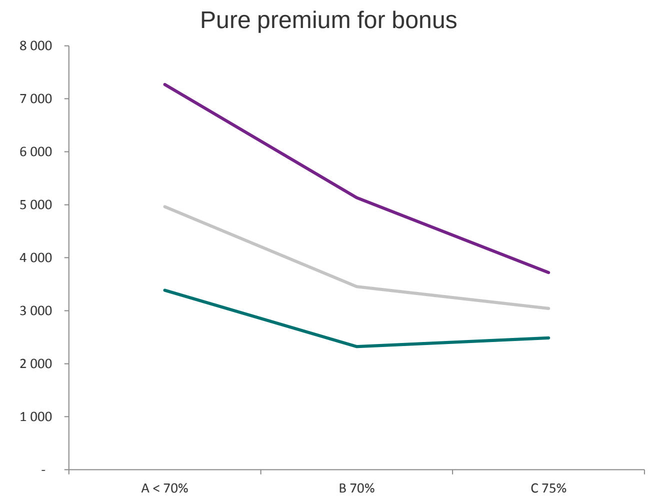
Results for claim size

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

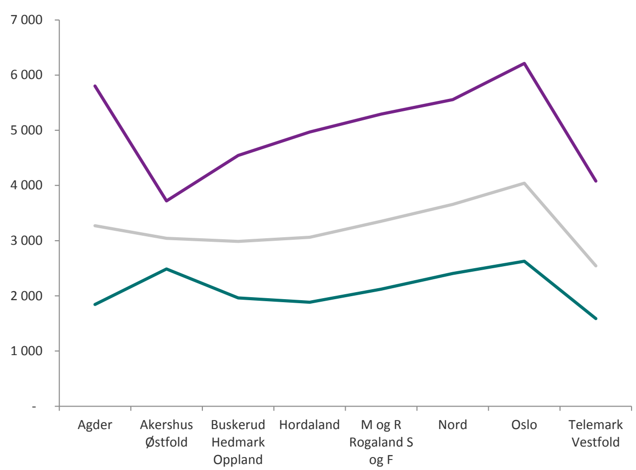


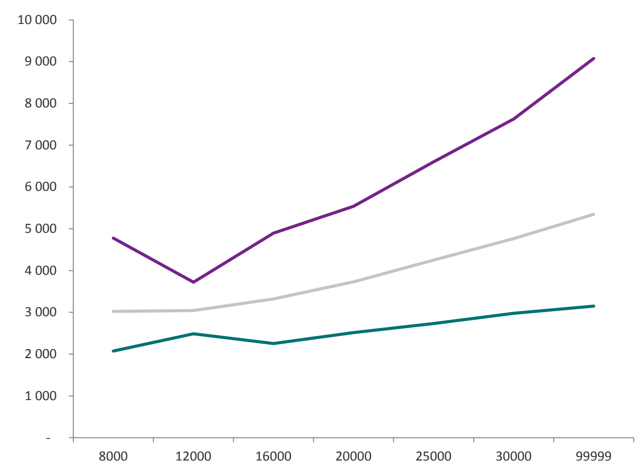
- PP: Select detail level
- PP: Review potential risk drivers
- PP: Select groups for each risk driver
- PP: Select large claims strategy
- PP: identify potential interactions
- PP: construct final model
- Price assessment





Pure premium
95% Lower Confidence Limit
95% Upper Confidence Limit





Some model selection principles

- Forward selection
- Backward elimination
- Stepwise regression
- Type 1 analysis
- Type 3 analysis

- **Forward Selection (FORWARD)**
- The forward-selection technique begins with no variables in the model. For each of the independent variables, the FORWARD method calculates statistics that reflect the variable's contribution to the model if it is included. The -values for these statistics are compared to the SLENTRY= value that is specified in the MODEL statement (or to 0.50 if the SLENTRY= option is omitted). If no statistic has a significance level greater than the SLENTRY= value, the FORWARD selection stops. Otherwise, the FORWARD method adds the variable that has the largest statistic to the model. The FORWARD method then calculates statistics again for the variables still remaining outside the model, and the evaluation process is repeated. Thus, variables are added one by one to the model until no remaining variable produces a significant statistic. Once a variable is in the model, it stays.
- **Backward Elimination (BACKWARD)**
- The backward elimination technique begins by calculating statistics for a model, including all of the independent variables. Then the variables are deleted from the model one by one until all the variables remaining in the model produce statistics significant at the SLSTAY= level specified in the MODEL statement (or at the 0.10 level if the SLSTAY= option is omitted). At each step, the variable showing the smallest contribution to the model is deleted.
- **Stepwise (STEPWISE)**
- The stepwise method is a modification of the forward-selection technique and differs in that variables already in the model do not necessarily stay there. As in the forward-selection method, variables are added one by one to the model, and the statistic for a variable to be added must be significant at the SLENTRY= level. After a variable is added, however, the stepwise method looks at all the variables already included in the model and deletes any variable that does not produce an statistic significant at the SLSTAY= level. Only after this check is made and the necessary deletions are accomplished can another variable be added to the model. The stepwise process ends when none of the variables outside the model has an statistic significant at the SLENTRY= level and every variable in the model is significant at the SLSTAY= level, or when the variable to be added to the model is the one just deleted from it.

- A Type 1 analysis consists of fitting a sequence of models, beginning with a simple model with only an intercept term, and continuing through a model of specified complexity, fitting one additional effect on each step. Likelihood ratio statistics, that is, twice the difference of the log likelihoods, are computed between successive models. This type of analysis is sometimes called an analysis of deviance since, if the dispersion parameter is held fixed for all models, it is equivalent to computing differences of scaled deviances. The asymptotic distribution of the likelihood ratio statistics, under the hypothesis that the additional parameters included in the model are equal to 0, is a chi-square with degrees of freedom equal to the difference in the number of parameters estimated in the successive models. Thus, these statistics can be used in a test of hypothesis of the significance of each additional term fit.
- Type 1 analysis has the general property that the results depend on the order in which the terms of the model are fitted. The terms are fitted in the order in which they are specified in the MODEL statement.

Price assessment

- The statistician needs help!
  - Product experts, price experts, strategy experts
- What is the company strategy?
- What is the market price?
- Customer focus:
  - What is the expected lifetime value for a given customer?
  - How likely is it that a new client will purchase the product?
  - What is the expected cross sales potential for a given customer?
  - What is the value of the customer for other players? (example: insurance sold in a bank)
- Discounts should reflect company strategy and the customer focus points above

- 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

Some price related tasks for an analyst in an insurance company