Swiss Re

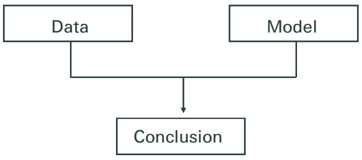
Tree-based Methods: Gaining new insights into (life) insurance data

Prof. Dr. Walter Olbricht, Dr. Ralf Krüger

Swiss Re

Background

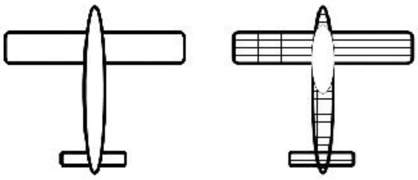
Basic model of statistical inference



Danger: "Pygmalion" phenomenon

Swiss Re

Background: armoring aircraft



Swiss Re

Background

Difficulties with some current approaches

Nature of actuarial data: Usually sampling error is not the prime concern, data are typically not small random samples with iid-error structure. They are often close to population data but show distinct substructures and heterogeneity.

Cox (Regression models and life tables, JRSS B, 1972, S. 187): *In other words, the applications are more likely to be in industrial reliability studies and in medical statistics than in actuarial science.*

Huber (Data analysis: what can be learned from the past 50 years, John Wiley & Sons, 2011, p. 12-13): *What really forces the issue is that larger data sets almost invariably are composite: they are less homogeneous and have more complex internal structure...*

Swiss Re

Background

Indeed, it is remarkably difficult to find homogeneous samples matching the ubiquitous "i.i.d. random variables" of theoretical statistics...

Already in 1940, Deming had admonished the statistical profession that as a whole it was paying much too little attention to the need for dealing with heterogeneous data and with data that arise from conditions not in statistical control (randomness)...

The problems cause by heterogeneous and highly structured data are difficult; I think they have been eschewed precisely because they go beyond tactics and require strategic thinking. Moreover, these problems cannot be harnessed through mathematical formalism, not even the theoretical ones among them.

Swiss Re

Background

- **Nonlinearity:** Nonadditive effects are not uncommon.

  Examples:
  - regular physical exercise and absenteeism in young and old age
  - profession in disability insurance in different age groups

  Modelling by interaction effects is only a partial remedy.

- **Complexity:** Many techniques are hard to understand.

  Effects:
  - makes it difficult to assess results
  - inhibits incorporation of contextual knowledge
  - precludes plausibility checks by practitioners

Swiss Re

Background

Example: average damage for four large subgroups of equal size

	Smoker	Nonsmoker
Male	1000 950 1000	700 750 700
Female	500 550 450	400 350 450

-- original data
-- linear model
-- nonlinear tree

Swiss Re

Background

Linear model:

DAMAGE = $\mu + \theta_1$SEX + θ_2SMOKING

μ = basic level (for female nonsmokers) ≈ 350

θ_1 = additive effect for „SEX = male" ≈ 400

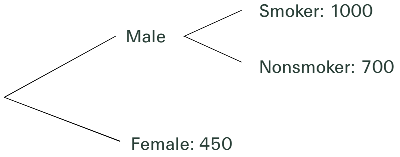
θ_2 = additive effect for „SMOKING = yes" ≈ 200

(The numbers are least squares estimates.)

Swiss Re

Background

Nonlinear tree:



- better fit
- plausibility checks and incorporation of contextual knowledge possible
- we can „think" in terms of this classification

Swiss Re

Background

Aim

An approach with the following properties:

- close to the data, rather data-analytic than inferential
- nonlinear
- transparent and easily interpretable
- corresponding to the way in which humans think about substructures i. e. by refinements

Suggestion: tree-based methods

Swiss Re

Tree-based methods

Example: Prediction of the perfomance of students

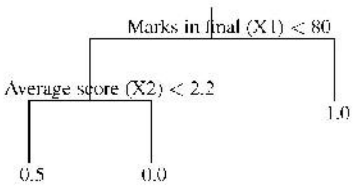
Marks in statistics final (X1)	Average school score (X2)	Success (S of F) (Y1)	Value (0.0, 0.5,1.0) (Y2)
99	3.2	S	1.0
85	1.6	S	1.0
84	2.4	S	1.0
81	1.9	S	1.0
79	3.5	F	0.0
78	2.5	F	0.0
66	1.4	S	0.5
89	1.2	S	1.0
44	2.4	F	0.0
25	2.5	F	0.0
40	2.0	S	0.5
90	3.0	S	1.0
35	3.5	F	0.0

Aim: We wish to predict success (classification problem) or a numerical value (regression problem) on the basis of marks (X1) and school score (X2).

Swiss Re

Tree-based methods

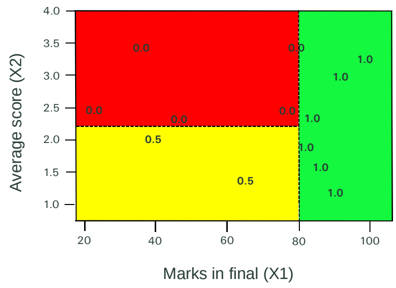
Performance of students: regression tree



Swiss Re

Tree-based methods

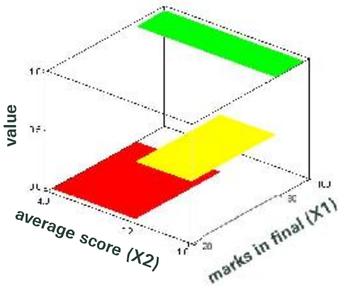
Performance of students: graphical representation for regression



Swiss Re

Tree-based methods

Performance of students: 3-d-graphical representation for regression



Swiss Re

Tree-based methods

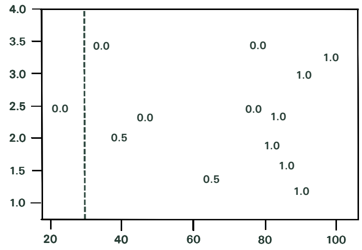
The regression problem: approach

- We use the residual sum of squares (RSS) as measure of deviance (impurity index), i. e. the sum of the squared distances between observations and predicted values.
- The optimal predicted value for each rectangle is the mean taken over the elements in that rectangle.

Swiss Re

Tree-based methods

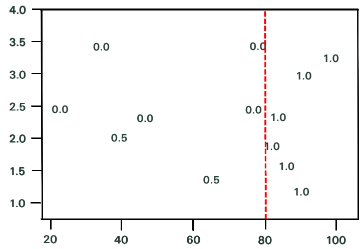
The regression problem: partitioning



Swiss Re

Tree-based methods

The regression problem: partitioning



Swiss Re

Tree-based methods

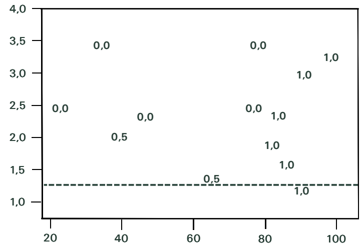
The regression problem: partitioning

No.	Split	MV1	MV2	RSS1	RSS2	RSS
0	kein		0.54		2.73	2.73
1	30.0	0.00	0.58	0.00	2.42	2.42
2	37.5	0.00	0.64	0.00	2.05	2.05
3	42.0	0.17	0.65	0.17	2.03	2.19
4	55.0	0.13	0.72	0.19	1.56	1.74
5	72.0	0.20	0.75	0.30	1.50	1.80
6	78.5	0.17	0.86	0.33	0.86	1.19
7	80.0	0.14	1.00	0.36	0.00	0.36
8	82.5	0.25	1.00	1.00	0.00	1.00
9	84.5	0.33	1.00	1.50	0.00	1.50
10	87.0	0.40	1.00	1.90	0.00	1.90
11	89.5	0.45	1.00	2.23	0.00	2.23
12	94.5	0.50	1.00	2.50	0.00	2.50

Swiss Re

Tree-based methods

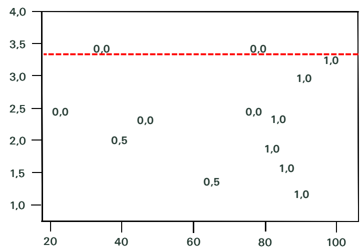
The regression problem: partitioning



Swiss Re

Tree-based methods

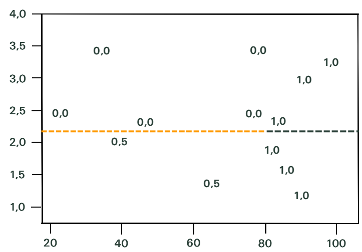
The regression problem: partitioning



Swiss Re

Tree-based methods

The regression problem: partitioning



Swiss Re

Tree-based methods

The regression problem: partitioning

No.	Split	MV1	MV2	RSS1	RSS2	RSS
0	kein		0.54		2.73	2.73
1	1.30	1.00	0.50	0.00	2.50	2.50
2	1.50	0.75	0.50	0.13	2.50	2.63
3	1.75	0.83	0.45	0.17	2.23	2.39
4	1.95	0.88	0.39	0.19	1.89	2.08
5	2.20	0.80	0.38	0.30	1.88	2.18
6 and 7	2.45	0.71	0.33	0.93	1.33	2.26
8 and 9	2.75	0.56	0.50	1.72	1.00	2.72
10	3.10	0.60	0.33	1.90	0.67	2.57
11	3.35	0.64	0.00	2.05	0.00	2.05

Swiss Re

Tree-based methods

The regression problem: partitioning

Hence split at „Marks in final < 80".
Now the same approach is applied to the left subset with seven elements.
For „Marks in final" RSS is always at least 0.3.
The results for „Average score" are:

No.	Split	MV1	MV2	RSS1	RSS2	RSS
0	kein		0.14		0.36	0.36
1	1.70	0.50	0.08	0.00	0.21	0.21
2	2.20	0.50	0.00	0.00	0.00	0.00
3	2.45	0.33	0.00	0.17	0.00	0.17
4 and 5	3.00	0.20	0.00	0.30	0.00	0.30

Hence we split at „Average score < 2.20".

Swiss Re

Tree-based methods

Technical aspects

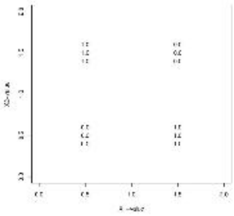
- Usually only *binary splits* are considered (no severe restriction).
- For an ordered (ordinal or continuous) predictor variable with m distinct values, the number of splits is $m-1$. A nominal predictor variable with m categories requires $2^{m-1}-1$ non-trivial splits. (In case of a 0-1-response variable, this can be reduced to $m-1$, if the predictor categories are ordered according to the proportion of 1's.)
- Due to the stepwise subdivision, tree-based methods are often referred to as *recursive partitioning*. It is, in general, not possible to find the *globally* optimal rectangular subdivision.

Swiss Re

Tree-based methods

Technical aspects

- For a 0-1-response variable, using the Gini-index leads to the same result as the RSS-criterion. There are still many open questions concerning the splitting criteria.
- There are many open problems concerning the stopping rules. The recursive nature implies a certain ‚short-sightedness'. Hence *pruning* is used: One grows a detailed tree and cuts it down afterwards by combining subgroups.



Swiss Re

Tree-based methods

Technical aspects

- There is usually not one final tree, but a whole range of potential candidates. However, a pronounced *‚internal structure'* (if it exists) will usually show up.

  Recommendation: Not just one tree, but *‚working with trees'*.
- Great danger of the approach: Spurious structure may be misinterpreted as a real feature.

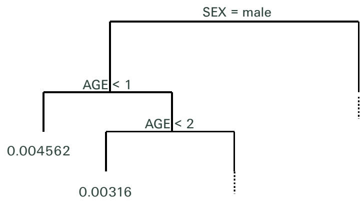
  Remedy: (Culture of) Validation

  Recommendation: Independent data set (better than cross-validation)

Swiss Re

Application to life insurance

DAV 2008 T as a tree
(Figures are mortality rates at the respective age)



Swiss Re

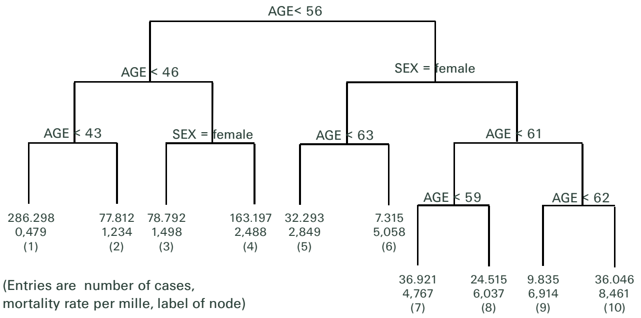
Application to life insurance

"Oracle test"

- A data set was taken from the Swiss Re data monitoring pool by combining some (partial) portfolios for some years. Variables are: Response variable (alive=0, dead=1), SEX (male=1, female=2), AGE
- The first years are used as learning set; the others as independent test set.
- A tree was generated on the basis of the learning set. On the same basis a/e-factors for adapting the DAV 2008 T to the data set were derived.
- The tree prediction and the 'classical' prediction (based on DAV 2008 T) for the independent data set (future) are compared to each other and to the actual development.

Swiss Re

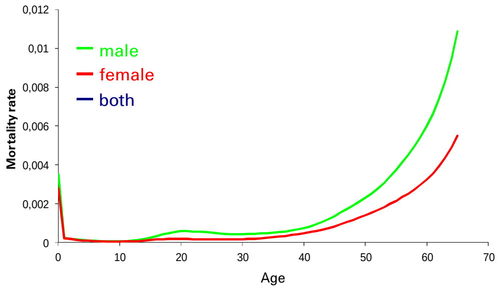
Application to life insurance



(Entries are number of cases, mortality rate per mille, label of node)

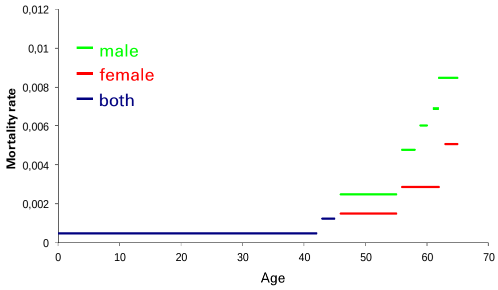
Swiss Re

Application to life insurance



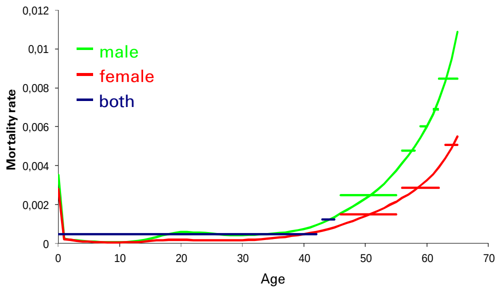
Swiss Re

Application to life insurance



Swiss Re

Application to life insurance



Swiss Re

Application to life insurance

"Oracle test"

Node	Learning set			Independent test set			
	no. of elements in node	no. of deaths in node	estimated mortality rate (per mille)	no. of elements in node	no. of deaths in node	tree prediction	classical prediction (DAV 2008 T)
1	286 298	137	0.479	254 995	143	122	127
2	77 812	96	1.234	75 882	60	94	79
3	78 792	118	1.498	79 202	146	119	116
4	163 197	406	2.488	155 912	361	388	389
5	32 293	92	2.849	33 163	119	94	96
6	7 315	37	5.058	7 440	26	38	36
7	36 921	176	4.767	41 759	163	199	188
8	24 515	148	6.037	20 708	118	125	118
9	9 835	68	6.914	8 354	59	58	55
10	36 046	305	8.461	33 525	219	284	299
Total	753 024	1 583		710 940	1 414	1 521	1 503

Swiss Re

Application to disability insurance

"Oracle test"

- A data basis was put together by mixing real (partial) portfolios from four successive years.
- Two years were chosen as learning set, the other two years form the independent test set.
- A tree was generated on the basis of the learning set. On the same basis e/a-factors for adapting the DAV 1997 I to the data set were derived.
- The tree prediction and the 'classical' prediction (based on DAV 1997 I) for the independent data set (future) are compared to each other and to the actual development.

Swiss Re

Application to disability insurance

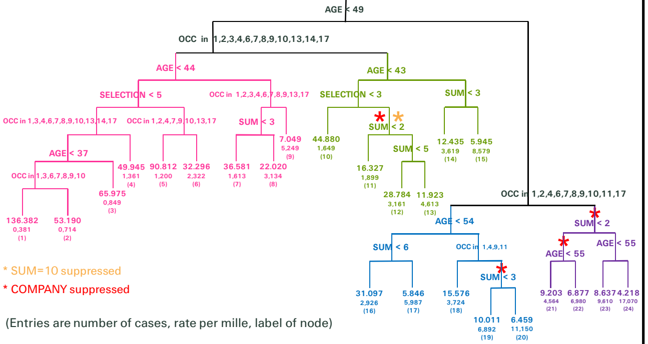
Variables

- Response variable (no disability=0, disability=1)
- COMPANY (A, ... , F)
- SEX (m, f)
- AGE (15, ... , 65)
- OCC (occupational class) (1, ... , 17)
- SELECTION (in years) (1, 2, 3, ... , 6+)
- SUM (sum insured in 10 classes) (1, ... , 10)

Swiss Re

Application to disability insurance

DISABILITY ~ COMPANY + SEX + AGE + OCC + SELECTION + SUM



* SUM=10 suppressed

* COMPANY suppressed

(Entries are number of cases, rate per mille, label of node)

Swiss Re

Application to disability insurance

"Oracle test"

	absolute nodes 1-15	absolute nodes 16-24	absolute total	relative nodes 1-15	relative nodes 16-24	relative total
observed	887	540	1'427			
classical	748	597	1'345	84%	111%	94%
tree	834	483	1'317	94%	89%	92%

Swiss Re

Specific Aspects

Tree-based and classical prediction relative to the actually observed claims (in %):

Company	A	B	C	D	E	F	Total
Tree-based	97	94	78	92	116	91	**92**
Classical	57	80	29	105	103	103	**94**

For a more detailed model it is also possible to derive company specific trees.

Swiss Re

Specific Aspects

Non-linearity:

Incidence rates	Age < 49	Age ≧ 49	
Office workers / white collar	0.9	3.7	(min. # cases: 121)
Retail occupations	1.4	7.4	
Craft workers	3.0	9.9	(min. # cases: 41)
Plant & machinery operators ...	1.1	8.4	

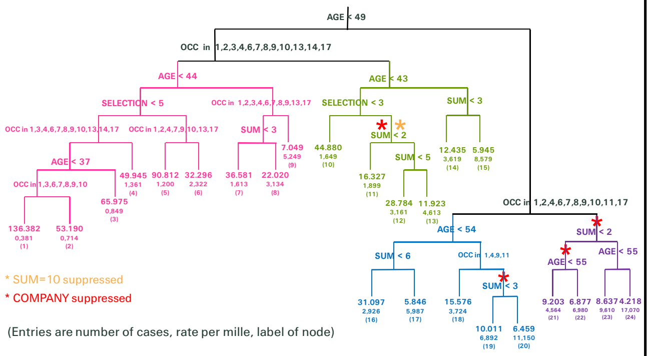
(figures in per mille)

Non-linear structures occur
No universal shape of incidence rates

Swiss Re

Specific Aspects

DISABILITY ~ COMPANY + SEX + AGE + OCC + SELECTION + SUM



* SUM=10 suppressed

* COMPANY suppressed

(Entries are number of cases, rate per mille, label of node)

Swiss Re

Specific Aspects

Impact of gender:

Occupational class	Age < 49		Age ≧ 49	
	M	F	M	F
Office workers / white collar	0.8	1.0	3.5	4.2
Retail occupations	1.4	1.5	7.8	6.2
Teaching, social & cultural professionals	0.8	1.1	5.1	4.3
Other healthcare workers	2.2	1.5	7.1	6.2
Catering & hospitality workers	2.7	2.4	7.6	7.3

(min. # cases: 17, figures in per mille)

Influence of the variable "Sex" appears much less than in classical modellings

Swiss Re

Specific Aspects

Impact of gender:

A fair and meaningful comparison needs to control for other variables by weighing the rates for males and females in each node by the number of policies in each node („standardisation").

Data set	Non-standardised		standardised	
	M	F	M	F
Learning set	2.347	1.530	2.063	2.026
Independent test set	2.478	1.824	2.223	2.336
Total	2.408	1.668	2.138	2.174

(figures in per mille)

No difference in incidence rates between males and females

Swiss Re

Specific Aspects

Impact of gender:

		Learning set		Independent test set		Total	
Node	**# elements**	**male (%)**	**female (%)**	**male (%)**	**female (%)**	**male (%)**	**female (%)**
1	136,382	12	34	11	32	12	33
2	53,190	8	6	8	6	8	6
3	65,975	8	11	8	10	8	11
4	49,945	9	3	9	3	9	3
5	90,812	13	13	14	14	14	13
6	32,296	4	6	4	7	4	6
7	36,581	5	5	5	5	5	5
8	22,020	4	2	4	2	4	2
9	7,049	1	1	1	1	1	1
10	44,880	7	5	7	4	7	4
11	16,327	3	1	3	1	3	1
12	28,784	5	3	5	3	5	3
13	11,923	2	1	2	1	2	1
14	12,435	2	1	2	1	2	1
15	5,945	1	0	1	0	1	0
Total	**614,544**						

Swiss Re

Specific Aspects

Impact of gender:

		Learning set		Independent test set		Total	
Node	**# elements**	**male (%)**	**female (%)**	**male (%)**	**female (%)**	**male (%)**	**female (%)**
16	31,097	4	4	4	4	4	4
17	5,846	1	0	1	0	1	0
18	15,576	3	1	2	1	2	1
19	10,011	2	1	2	1	2	1
20	6,459	1	0	1	0	1	0
21	9,203	2	1	1	1	1	1
22	6,877	1	0	1	0	1	0
23	8,637	2	0	2	0	2	0
24	4,218	1	0	1	0	1	0
Total	**97,924**						

Swiss Re

Conclusions

- Tree-based methods seem ideally suited for the actuarial field.
- They are feasible and offer an interesting alternative or a complement to traditional approaches.
- In particular, they can help to uncover the ‚internal structure' of the data.
- It is recommended not to use just one tree, but to ‚work with trees'.
- Particular emphasis *must* be put on proper validation.
- They appear promising for less transparent situations.

Swiss Re

References

Olbricht W (2012) Tree-based methods: a useful tool for life insurance.
Eur Actuar J 2:129-147

Bauer M, Krüger R, Olbricht W (2013) Tree-based methos: an application to disability probabilities
Eur Actuar J 3:491-513

Swiss Re

Legal notice