AKUR8 | CAS | PINNACLE

2021 CAS ANNUAL MEETING - San Diego, California

Machine Learning powered pricing: from GLMs to GAMs

9th of November

11:30 AM (Pacific Standard Time)

San Diego Ballroom

Guillaume Beraud
Chief Actuary & Co-Founder
Akur8

Gaétan Veilleux
Senior Consulting Actuary
Pinnacle Actuarial Resources

November 9, 2021

Gaetan Veilleux
Senior Consulting Actuary

PINNACLE
ACTUARIAL RESOURCES

Gaetan Veilleux, FCAS, MAAA, CSPA
Senior Consulting Actuary

30+ years of experience in the property/casualty insurance industry

20+ years of predictive analytics experience

Past president of the Casualty Actuaries of the Northwest

A Brief History of Insurance Analytics*

How actuaries have moved from using univariate analysis to predictive analytic approaches: the evolution of insurance analytics

** With apologies to Stephen Hawking*

Pre-1960 Classification Ratemaking

- Data aggregated
- Relativities determined one dimension at a time

- Simple homeowners example:
  - All brick dwellings → brick rate
  - All small dwellings → small rate
  - Small brick dwellings
    - Insufficient data
    - Includes some information (Brown)

Pre-1960 Classification Ratemaking

Construction	Exposures	Loss & LAE	Indicated PP	Indicated Relativity	Indicated Relativity to Base
Brick	280	195,500	698	1.359	1.000
Frame	235	69,000	294	0.572	0.421
Total	515	264,500	514	1.000	0.736

Sq Feet	Exposures	Loss & LAE	Indicated PP	Indicated Relativity	Indicated Relativity to Base
Large	295	192,000	651	1.267	1.975
Small	220	72,500	330	0.642	1.000
Total	515	264,500	514	1.000	1.558

1960s – Minimum Bias

- "Two Studies in Automobile Insurance" – Robert Bailey and LeRoy Simon, 1960
- "Insurance Rates with Minimum Bias" – Robert Bailey, 1963

Loss Costs

	Large	Small	Total
Brick	800	500	698
Frame	400	200	294
Total	651	330	514

Exposures

	Large	Small	Total
Brick	185	95	280
Frame	110	125	235
Total	295	220	515

Iteration:

		0	1	2	3
Brick	C_1		2.1275	2.1173	2.0993
Frame	C_2		0.9684	0.9496	0.9412
Large	S_1	1.970	1.9195	1.9349	1.9516
Small	S_2	1.000	1.1217	1.1334	1.1432

Iteration 3 normalized results:

Brick	1.000
Frame	0.448
Large	1.707
Small	1.000

Bailey and Simon

- Four criteria for an acceptable set of relativities
  - Should reproduce experience for each class overall
  - Should reflect relative credibility of the various groups
  - Should provide the minimum amount of departure from the raw data for the maximum number of people
  - Should produce a rate for each sub-group of risks, which is close enough to the experience that the differences could reasonably be caused by *chance*

1972 – Generalized Linear Models

JOURNAL OF THE
ROYAL STATISTICAL SOCIETY

SERIES A (GENERAL) Volume 135, No. 3, 1972

CONTENTS

ROYAL STATISTICAL SOCIETY, 21 BENTINCK ST., LONDON W1M 6AR

J. R. Statist. Soc. A,
(1972), **135**, *Part* 3, p. 370

Generalized Linear Models

By J. A. Nelder and R. W. M. Wedderburn

Rothamsted Experimental Station, Harpenden, Herts

Summary

The technique of iterative weighted linear regression can be used to obtain maximum likelihood estimates of the parameters with observations distributed according to some exponential family and systematic effects that can be made linear by a suitable transformation. A generalization of the analysis of variance is given for these models using log-likelihoods. These generalized linear models are illustrated by examples relating to four distributions; the Normal, Binomial (probit analysis, etc.), Poisson (contingency tables) and gamma (variance components).

The implications of the approach in designing statistics courses are discussed.

Keywords: ANALYSIS OF VARIANCE; CONTINGENCY TABLES; EXPONENTIAL FAMILIES; INVERSE POLYNOMIALS; LINEAR MODELS; MAXIMUM LIKELIHOOD; QUANTAL RESPONSE; REGRESSION; VARIANCE COMPONENTS; WEIGHTED LEAST SQUARES

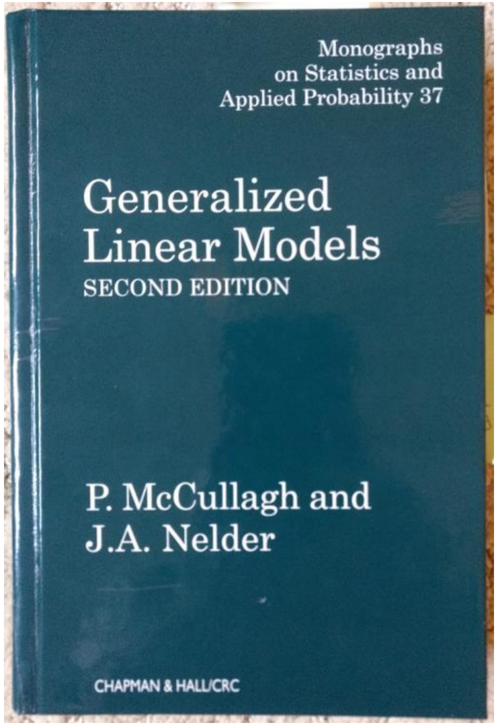
Introduction

Linear models customarily embody both systematic and random (error) components, with the errors usually assumed to have normal distributions. The associated analytic technique is least-squares theory, which in its classical form assumed just one error component; extensions for multiple errors have been developed primarily for analysis of designed experiments and survey data. Techniques developed for non-normal data include probit analysis, where a binomial variate has a parameter related to an assumed underlying tolerance distribution, and contingency tables, where the distribution is multinomial and the systematic part of the model usually multiplicative. In both these examples there is a linear aspect to the model; thus in probit analysis the parameter p is a function of tolerance Y which is itself linear on the dose (or some function thereof), and in a contingency table with a multiplicative model the logarithm of the expected probability is assumed linear on classifying factors defining the table. Thus for both, the systematic part of the model has a linear basis. In another extension (Nelder, 1968) a certain transformation is used to produce normal errors, and a different transformation of the expected values is used to produce linearity.

Mid-70s

- The first GLM packages
  - GENSTAT
  - GLIM software
    - Linear regression
    - Logistic and probit regression
    - Poisson regression
    - Log-linear models
    - Regression of skewed continuous distributions (sa Gamma)
- Quasi-likelihood (Wedderburn 1974)
- Overdispersion

1983 – GLMs

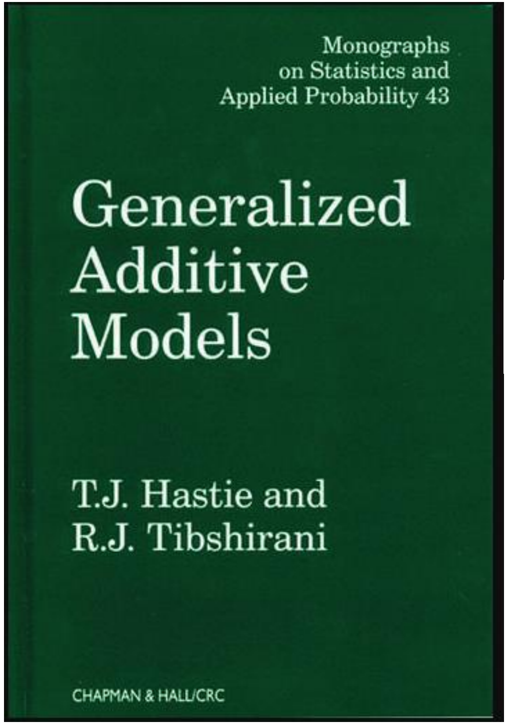


$$g[E(y|x)] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

1988 – Improving Minimum Bias

- "Minimum Bias with Generalized Linear Models" by Robert L. Brown, 1988
- Replace the bias function (the balance principal) with an expression from the likelihood function
- Assumes a distribution for quantity being modeled
- Solves for parameters to maximize its value
- A "statistical modeling" approach

1990 – GAMs introduced



$$g(E_Y(y|x)) = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p)$$

1992 – GLMs in Ratemaking

JIA 119 (1992) 457-543

J.I.A. **119**, III, 457–543

STATISTICAL MOTOR RATING: MAKING EFFECTIVE USE OF YOUR DATA

BY M. J. BROCKMAN, B.SC., F.I.A. AND T. S. WRIGHT, M.A., F.S.S., M.I.S.

[Presented to the Institute of Actuaries, 27 April 1992]

ABSTRACT

The paper gives details of statistical modelling techniques which can be used to estimate risk and office premiums from past claims data. The methods described allow premiums to be estimated for any combinaton of rating factors, and produce standard errors of the risk premium. The statistical package GLIM is used for analysing claims experience, and GLIM terminology is used and explained thoughout the paper.

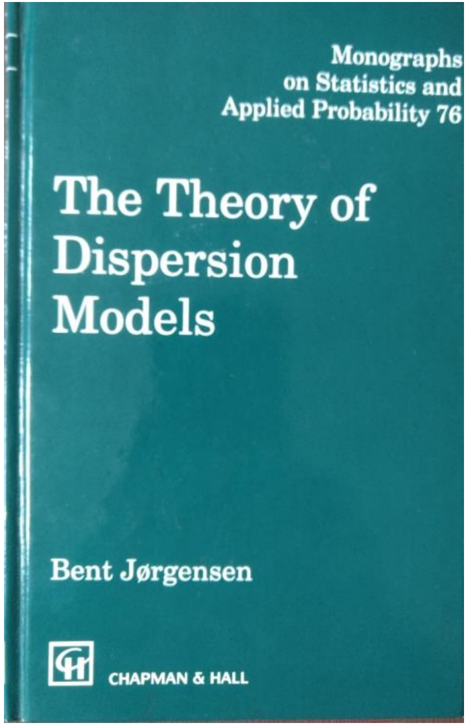
Arguments are put forward for modelling frequency and severity separately for different claim types. Fitted values can be used to estimate risk premiums, and the incorporation of expenses allows for the estimation of office premiums. Particular attention is given to the treatment of no claim discount.

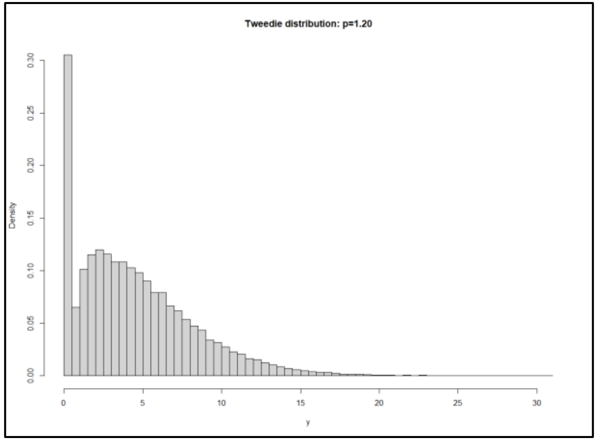
The paper also discusses possible uses of the modelled premiums. These include the construction of 'standardised' one way tables and the analysis of experience by postal code and model of vehicle. Also discussed is the possibility of using the results for assessing the impact of competition, and for finding 'niche' markets in which an insurer can operate both competitively and profitably.

KEYWORDS

General Insurance; Motor; Pricing; Statistical Analysis

1997 – The Tweedie Distribution





1999 – GLMs and Minimum Bias

A SYSTEMATIC RELATIONSHIP BETWEEN MINIMUM BIAS AND GENERALIZED LINEAR MODELS

STEPHEN MILDENHALL

Abstract

The minimum bias method is a natural tool to use in parameterizing classification ratemaking plans. Such plans build rates for a large, heterogeneous group of insureds using arithmetic operations to combine a small set of parameters in many different ways. Since the arithmetic structure of a class plan is usually not wholly appropriate, rates for some individual classification cells may be biased. Classification ratemaking therefore requires measures of bias, and minimum bias is a natural objective to use when determining rates.

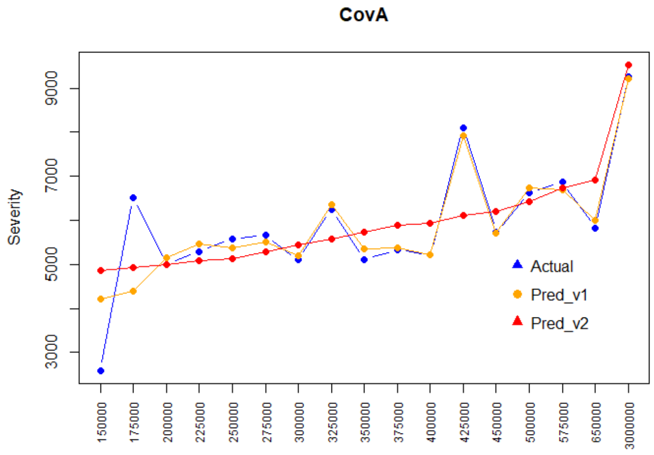
This paper introduces a family of linear bias measures and shows how classification rates with minimum (zero) linear bias for each class are the same as those obtained by solving a related generalized linear model using maximum likelihood. The examples considered include the standard additive and multiplicative models used by the Insurance Services Office (ISO) for private passenger auto ratemaking and general liability ratemaking (see ISO [11] *and Graves and Castillo* [8], *respectively).*

Knowing how to associate a generalized linear model

The "links" between specific Minimum Bias techniques and specific GLM forms.

2000 – Present

- Rapid adoption of GLMs
- GLMs quickly become the standard tool for rating models
- Expansion of GLM usage to address other business questions

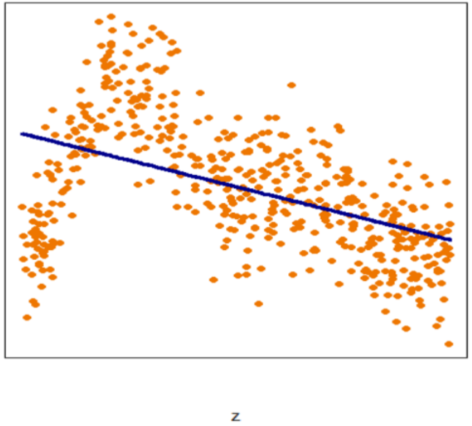


GLM Popularity

- Easy to develop
- Intuitive results
- Easy to visualize
- Explainable
- Wide variety of diagnostics
- Closed form results
- Regulatory acceptance (eventually)
- Transparency
- Wealth of insurance specific material on the use of GLMs

GLM Limitations

- Tendency to “torture” the data at times
- Poor choice for non-linear relationships
- Need for additional flexibility



Generalized Additive Models - GAMs

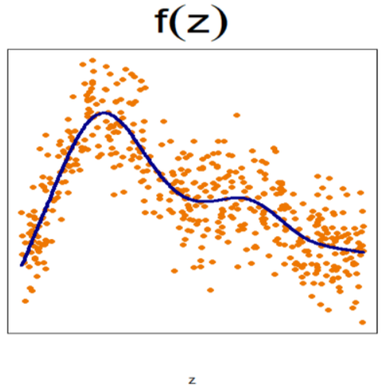
Generalized Linear Models **GLM**

$$g(E_Y(y|x)) = \beta_0 + \beta_1 x_1 + \ldots \beta_p x_p$$

Generalized Additive Models **GAM**

$$g(E_Y(y|x)) = \beta_0 + f_1(x_1) + f_2(x_2) + \ldots + f_p(x_p)$$

- GAMs are an extension of GLMs
- Allows for non-linear effects through the functions
- Regulatory concerns



Gaetan Veilleux

309.807.2333

gveilleux@pinnacleactuaries.com

PINNACLE
ACTUARIAL RESOURCES

AKUR8

CAS Annual Meeting

November 9, 2021

Guillaume Béraud-Sudreau

Chief Actuary & Co-Founder of Akur8

Biography

Guillaume is the Chief Actuary and Co-Founder of Akur8.

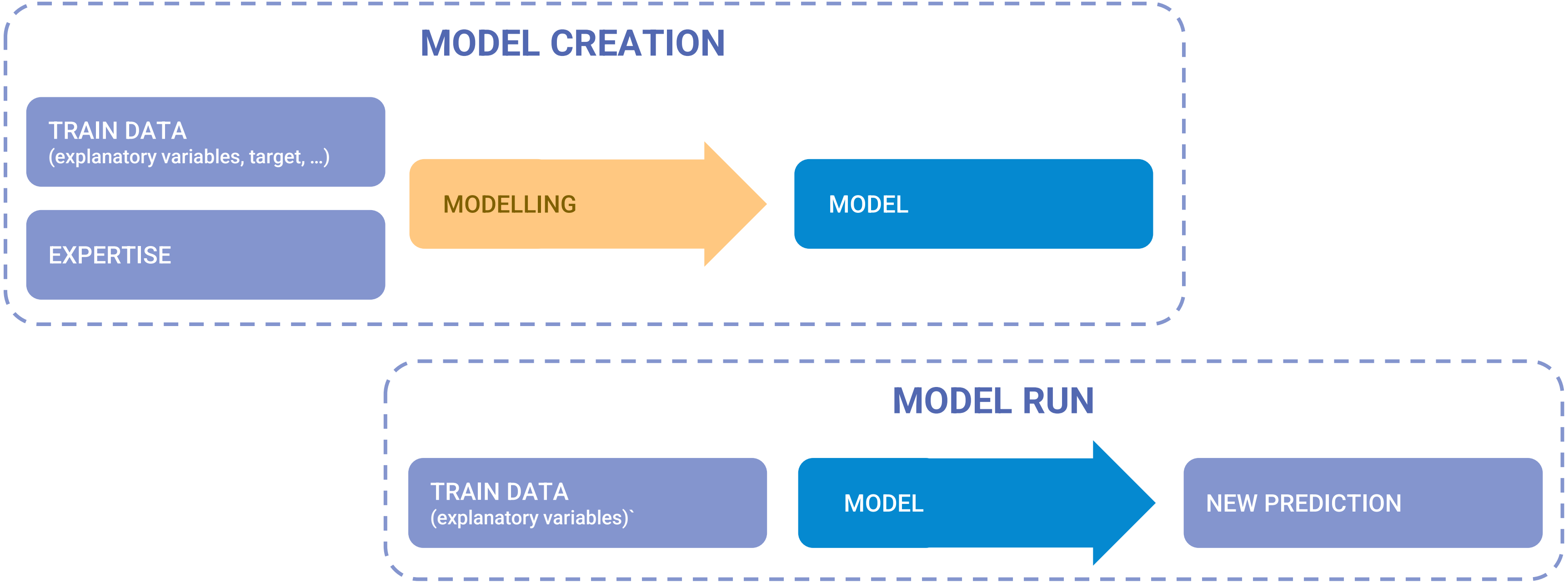
He has both a data science and an actuarial background. Guillaume started researching the potential of AI for insurance pricing as Head of Pricing R&D at Axa Global Direct, before being incubated at Kamet Ventures and founding Akur8.

Guillaume is a Fellow of the French Institute of Actuaries and holds Master degrees in Actuarial Science, Cognitive Science and Engineering from Institut des Actuaires - CNAM, Ecole normale supérieure, and Télécom Paris.

AKUR8

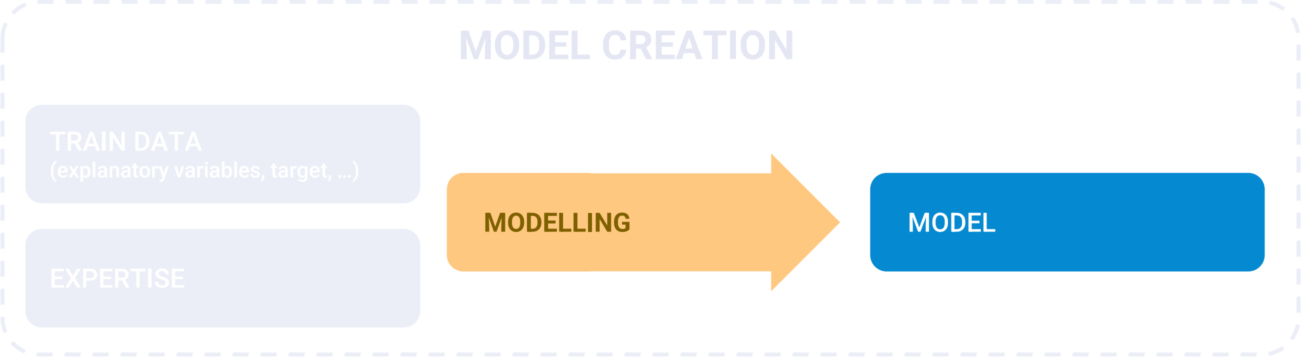
Modelling and Models

The Modelling Approach
Creating & using models



The Modelling Approach

Creating & using models



- What is a **transparent model**?
- What is **machine-learning**?
- Why are they **opposed**?
- What happens when the two concepts are **combined**?

NEW PREDICTION

The choice between black-box ML and traditional GLMs (presentation by Swiss Re)

Model approach comparison

GLM vs. other ML-methods

	XGBoost	Random Forest	GLM
Automatic Feature selection	✓	✓	✗
Model Runtime	Longer	Medium	Short
Performance (AUC)	High	Medium	Medium
Interpretable results	✗	✗	✓

- Different modelling techniques display different performance along key measurement criteria
- Setting clear expectations a priori helps to select the preferred one

Swiss Re

Presented by Swiss Re at the Swiss Association of Actuaries on the 06/10/2021: https://www.actuaries.ch/de/fach-arbeitsgruppen/fachgruppe-data-science/Ds%20AMs

October 2021 12

Model creation & structure

	Creation Process	Result
GBM		**Black Box** (Trees Ensemble)
Random Forest		**Black Box** (Trees Ensemble)
Neural Network		**Black Box** (Neural Network)
Data-Prep + GLM		**Transparent** (Data-Prep + LM.)
GAM (manual)		**Transparent** (GAM)

Classic Actuarial approach

AKUR8

Direct Models Visualization

While **model interpretability techniques** can be applied to any model, a **direct model understanding** is restricted to the specific class of models

To be understood, models must be:

- **Reductible:** the models can be **splitted** and **visualized** piece-by-piece
- **Parsimonious:** the model must incorporate a **limited number of effects** to be analizable

This class of models restrict human-understandable models to:

→ **Simple rules**
→ **Shallow tree**
→ **Generalized Additive Models** (including GLMs), with **parsimonious interactions**

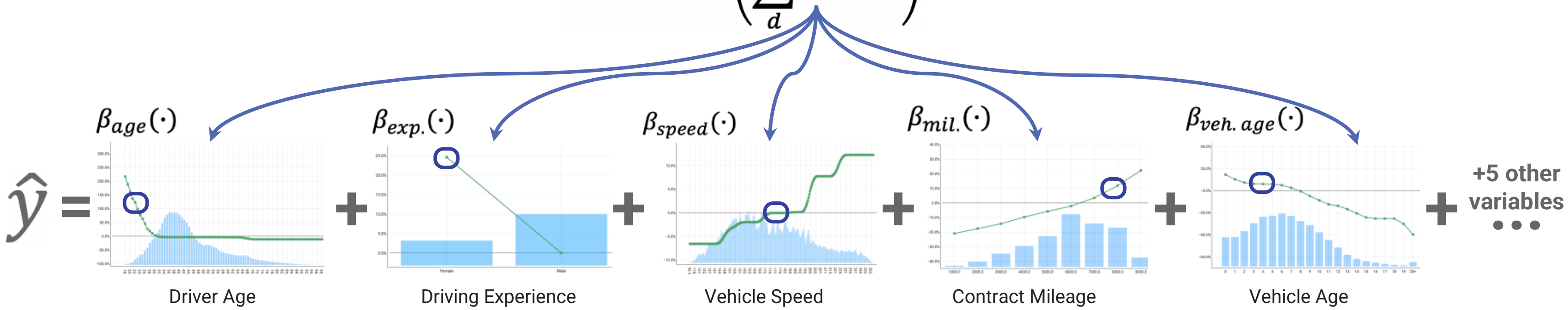
Direct Models Visualization

If a model can be decomposed, it can be visualized

Actuaries have been focusing during the past 20 years on the GAM modeling, because it allows the modeler to decompose the model's effects $\beta_j(X_j)$ and:

- Validate the effects
- "Force" the effects if no exposure is available

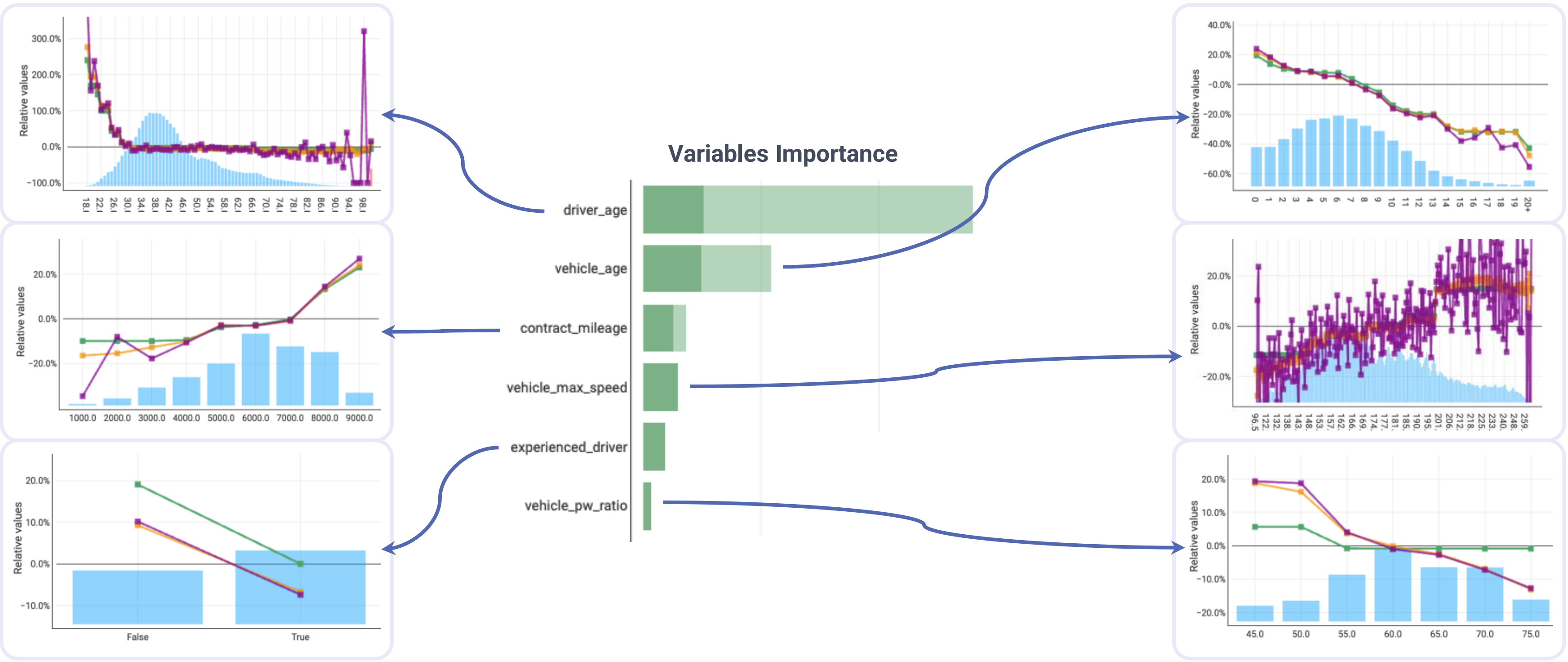
The GAM models are defined by their shape: $$\hat{y}(X) = g^{-1}\left(\sum_d \beta_d(X_d)\right)$$



Here the model itself is visualized and fully understood by a human.

Analysing a GAM

Only a limited number of variables play a role; each variable's impact is fully known



GLMs or GAMs

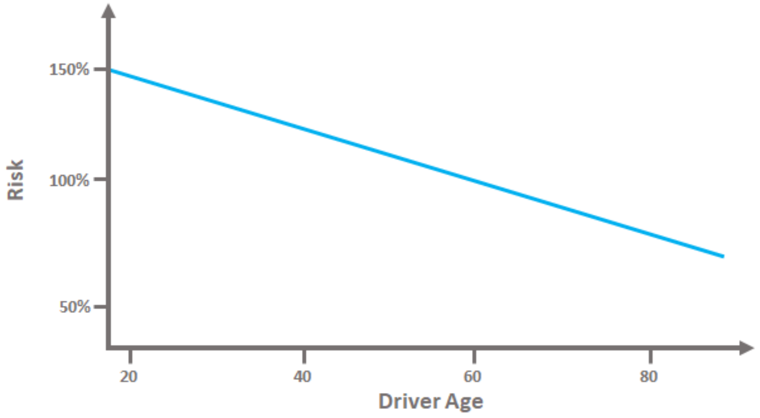
Linear or Additive

AKUR8

Linear models, GLMs and GAMs

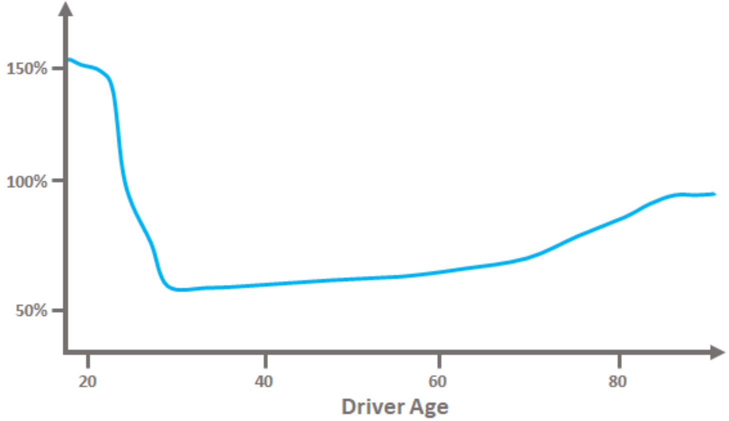
Linear Model

- Simple and well know technique
- First regression created & learned
- Captures the linear relations in the data
- Simultaneously select the variables and fit the trends

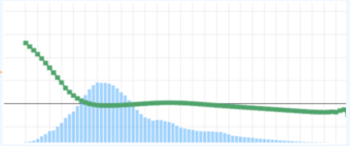
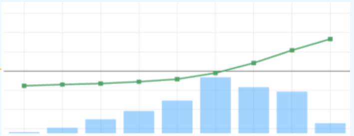
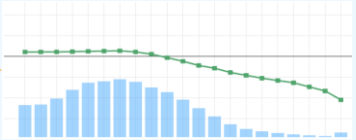
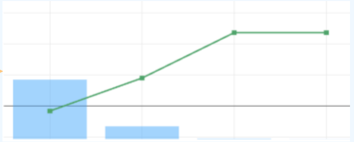


Additive Model

- Much more powerful models
- Captures non-linear effects
- Incorrectly called "GLMs"
- Requires both variables selection and fitting



Creating a GAM with variables transformations

Original Variables	Heavy Data-Preparation → Transformed Variables	GLM Modeling → Coefficients	Aggregation into a GAM → Functional Effects
Driver Age	Driver Age	-2.50	
	Driver Age ^{2}	0.10	
	Driver Age ^{3}	-0.02	
Annual Mileage	Annual Mileage	1.20	
	Annual Mileage ^{2}	0.30	
	Annual Mileage ^{3}	-0.01	
Vehicle Age	Vehicle Age	0.70	
	Vehicle Age ^{2}	-1.10	
	Vehicle Age ^{3}	-0.04	
Nb. of Past Claims	Past Claims=0	-10.00	
	Past Claims=1	20.00	
	Past Claims=2+	50.00	

Building the GAMs manually

Only a limited number of variables play a role; each variable's impact is fully known

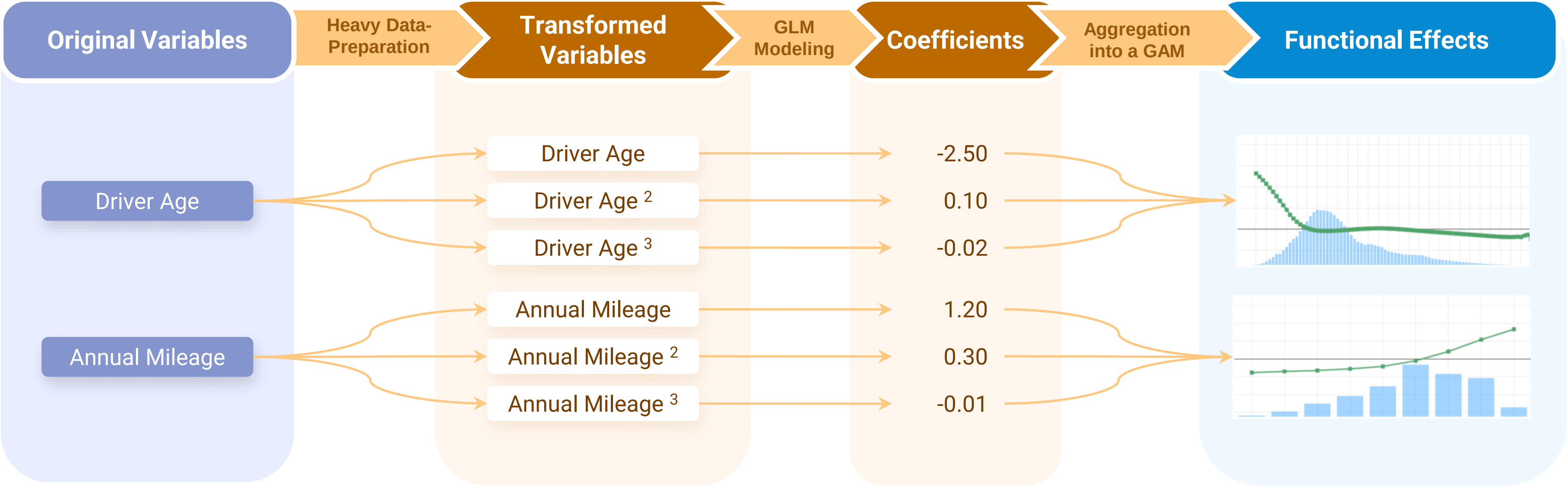
Generalized Additive Models are **transparent** by structure.

So, as modelers can **understand** and **interact** with them, it is possible to create them manually (unlike ensemble of trees, which have to be created by machines).

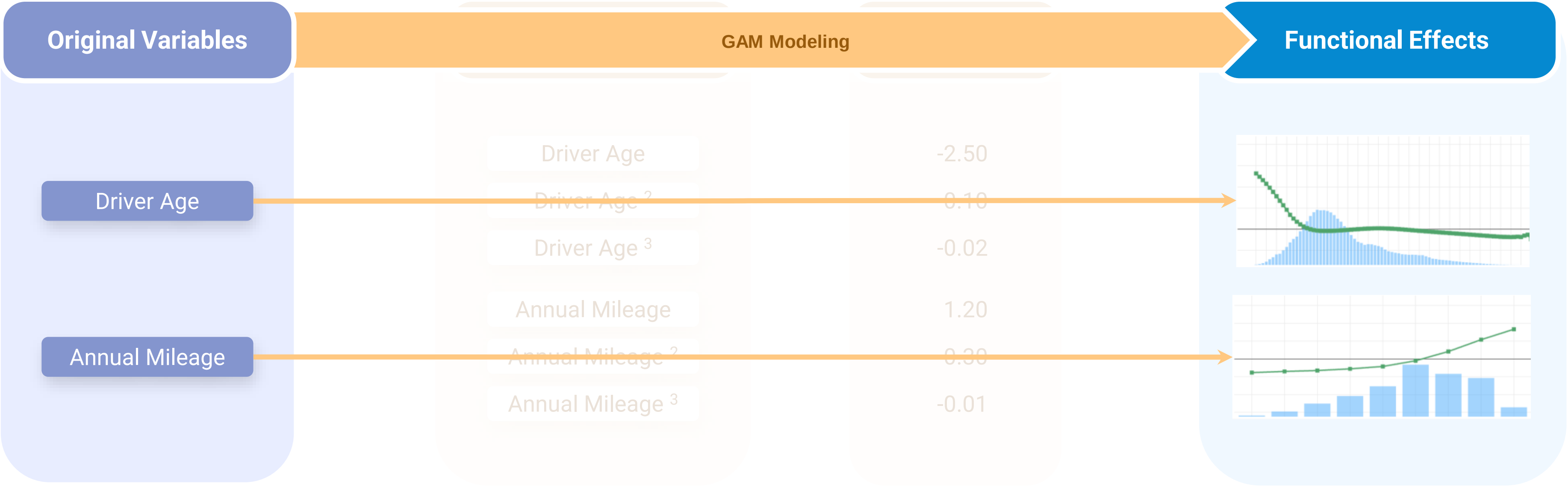
However, building GAMs through variables transformations and linear modeling leads to **severe limitations**!

- ✕ The right **set of variables** need to be **selected manually** by the modeler
- ✕ The right **transformations** need to be **manually created** by the modeler
- ✕ They are **limited to linear combinations** of the basis of variables created.
- ✕ Complexity is limited as creating too many transformations leads to **overfitting**.

Creating a GAM model through variable transformations...



… or creating a GAM with Machine Learning ?



Classic ML approach

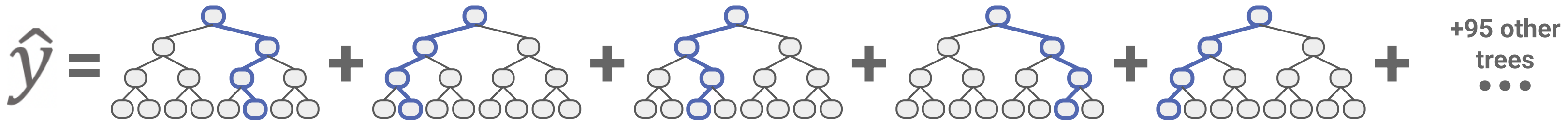
Black-Box models

Black-box models can be analysed

Most ML models are black-boxes: they **can't be directly understood, but can be analysed.**

For instance, a Gradient Boosting generates predictions from an ensemble of decision trees: $\hat{y}(X) = g^{-1}\left(\sum_t \mathrm{T}_t(X)\right)$

Each tree T_t leverages all the dimensions of the data, generating interactions between the variables.



GBMs are really great because **they just work**:
it is straightforward to produce automatically good models.

As a GBM typically involves hundreds of trees of depth 2 to 6 (generating 2 to 6-ways interactions),
this model is **not directly understandable** by a human.

For this reason, powerful model-analysis tools have been developed.

Global Parameters and Model Parameters

Models creation is automated:

- The user defines **global parameters** and **data**.
- The algorithm **fits** on the data and produces the **model**.

The model itself is often less looked-at than the global parameters.

For instance, when building a GBM, a user will find the global parameters maximizing the back-test results (through a k-fold), not the best model.

GLOBAL PARAMETERS

- Number of trees
- Trees depth
- Learning Rate

DATA

- Explanatory Variables
- Target Variable

FIT

MODEL PARAMETERS

- Ensemble of trees
  (split points, split variables, leaves estimates)

Finding the best Global Parameters

The Grid-Search approach

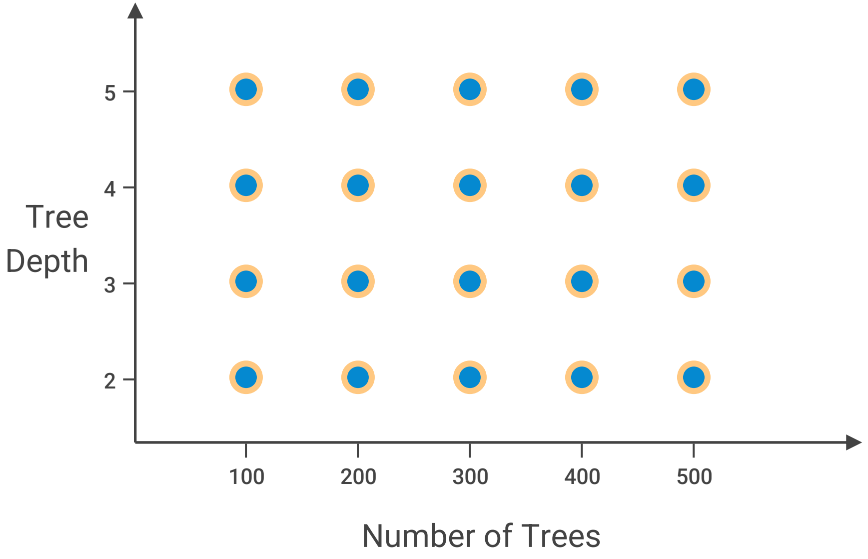
The grid-search approach seeks at finding the best **Global Parameters**.

Based on a modeling data-set, models are fitted with different global parameters, and their performance is measured.

The models themselves are not looked at: only the out-of-sample performance is considered.

The set of global parameters leading to the best performances is considered as the best one.

They are used to fit a model on the entire data-set: this model will be the one used in production.
It is possible to follow this whole process without ever looking at the selected model.



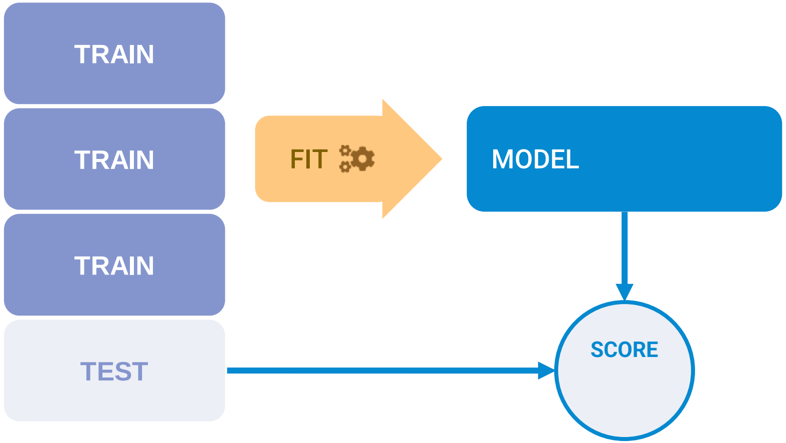
Testing a models performance

The K-Fold approach

As the grid-search approach of modeling is “blind”, it relies a lot on performances measures.

To make sure the performance measure is as precise as possible, a k-fold approach is used: the data is split in K subsets (typically 4) and K models are created on all the subsets but one. The performances of these models are tested on the last subset.

This approach is very efficient, and works perfectly well independently of the model. However, it requires a completely automated model creation process.



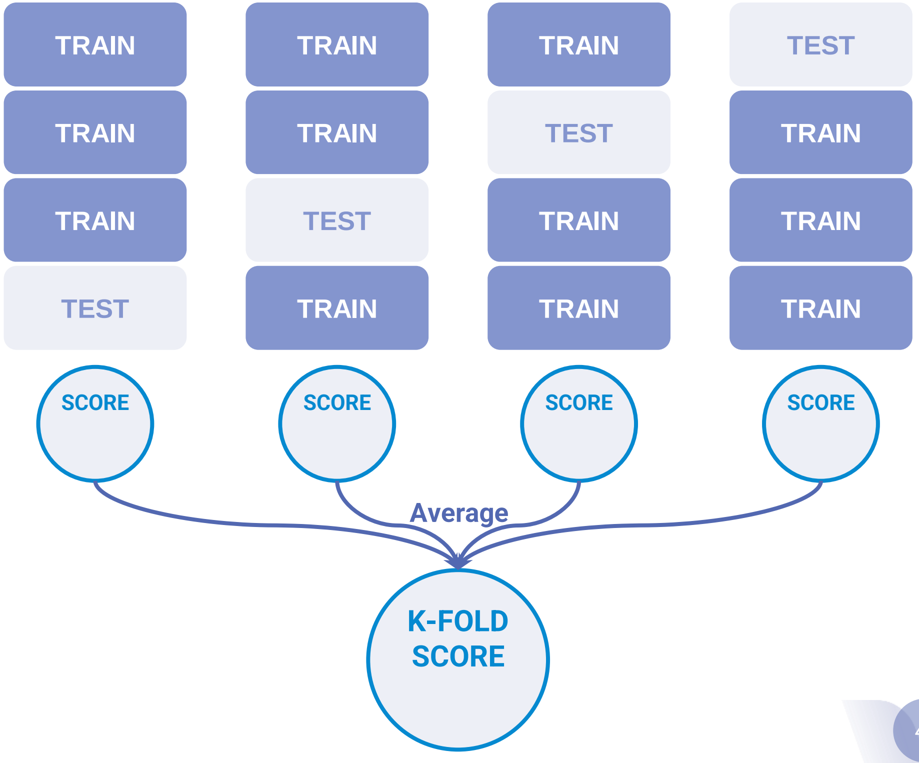
Testing a models performance

The K-Fold approach

As the grid-search approach of modeling is "blind", it relies a lot on measures of performances.

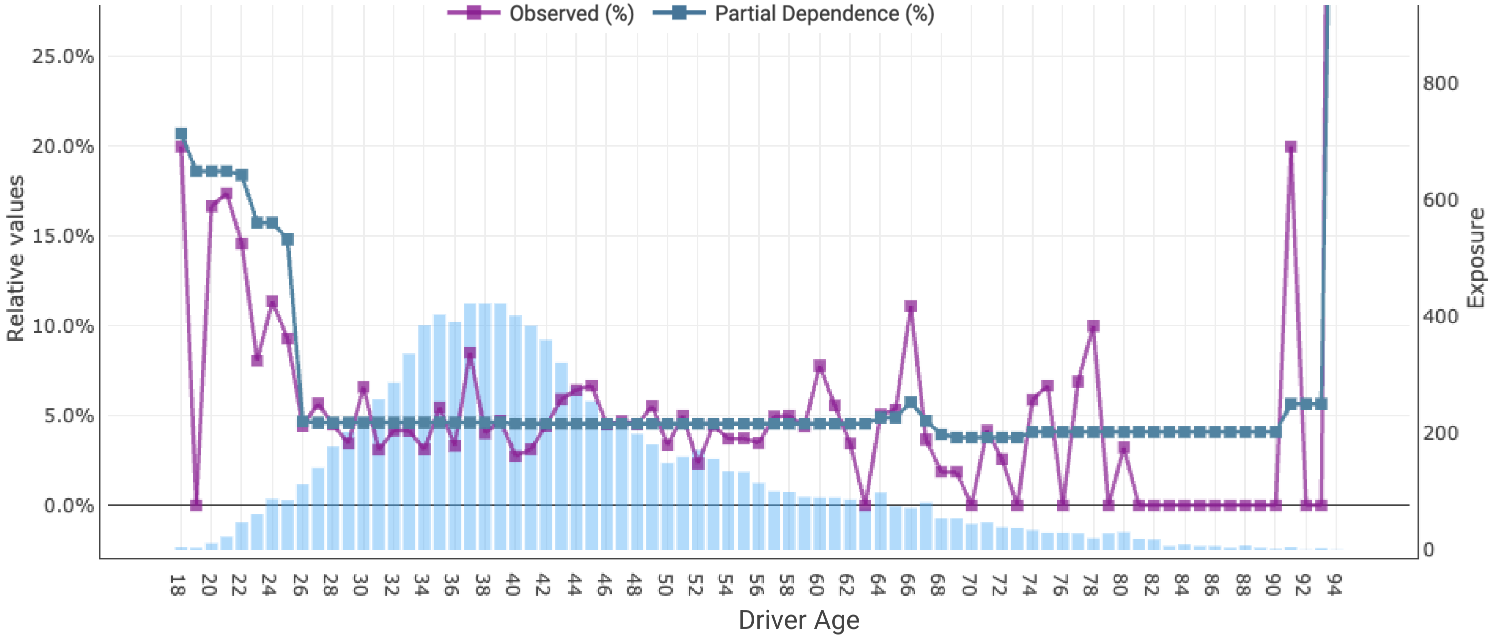
To make sure the performance measure is as precise as possible, a k-fold approach is used: the data is split in K subsets (typically 4) and K models are created on all the subsets but one. The performances of these models are tested on the last subset.

This approach is very efficient, and works perfectly well independently of the model. However, it requires a completely automated model creation process.



Example of black-box analysis

PDP : understand the global impact

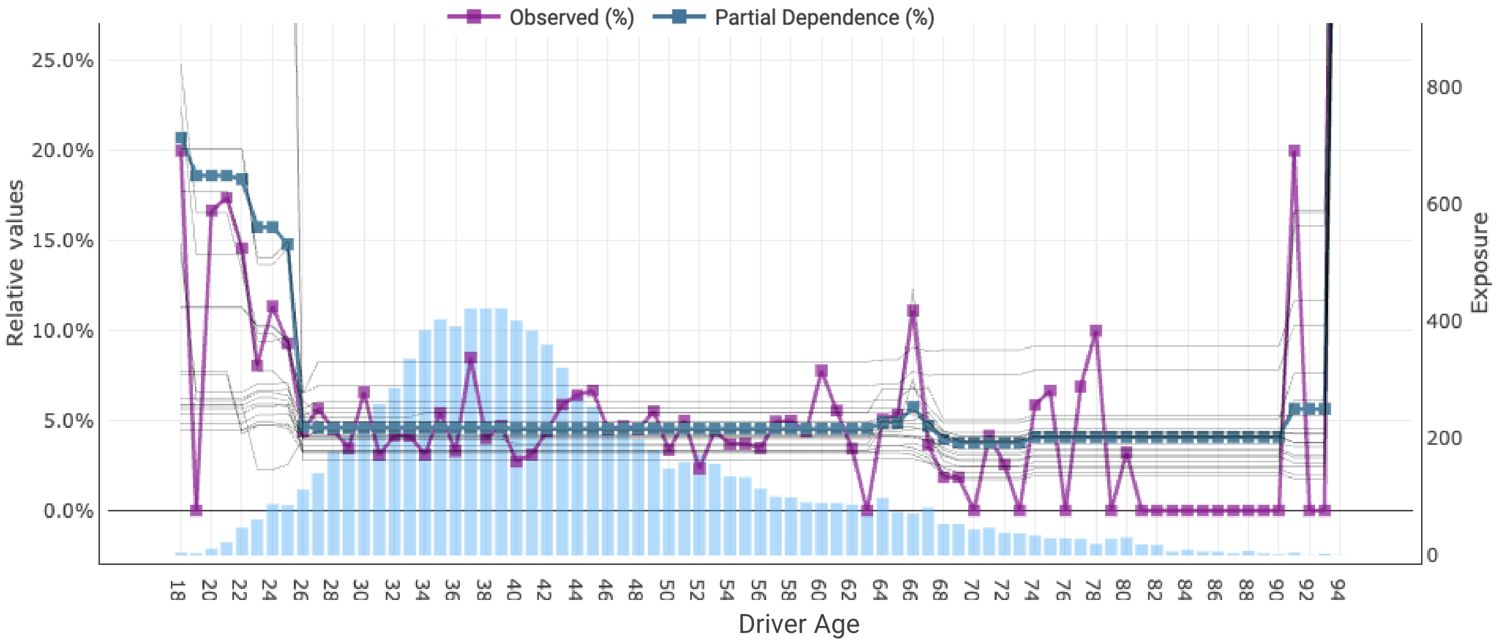


For example: a Partial Dependence Plot (PDP)) and Individual Conditional Expectation (ICE) showing the impact of a driver's age.

*Screenshot from Akur8's interface

Example of black-box analysis

ICE: visualize the conditional impacts

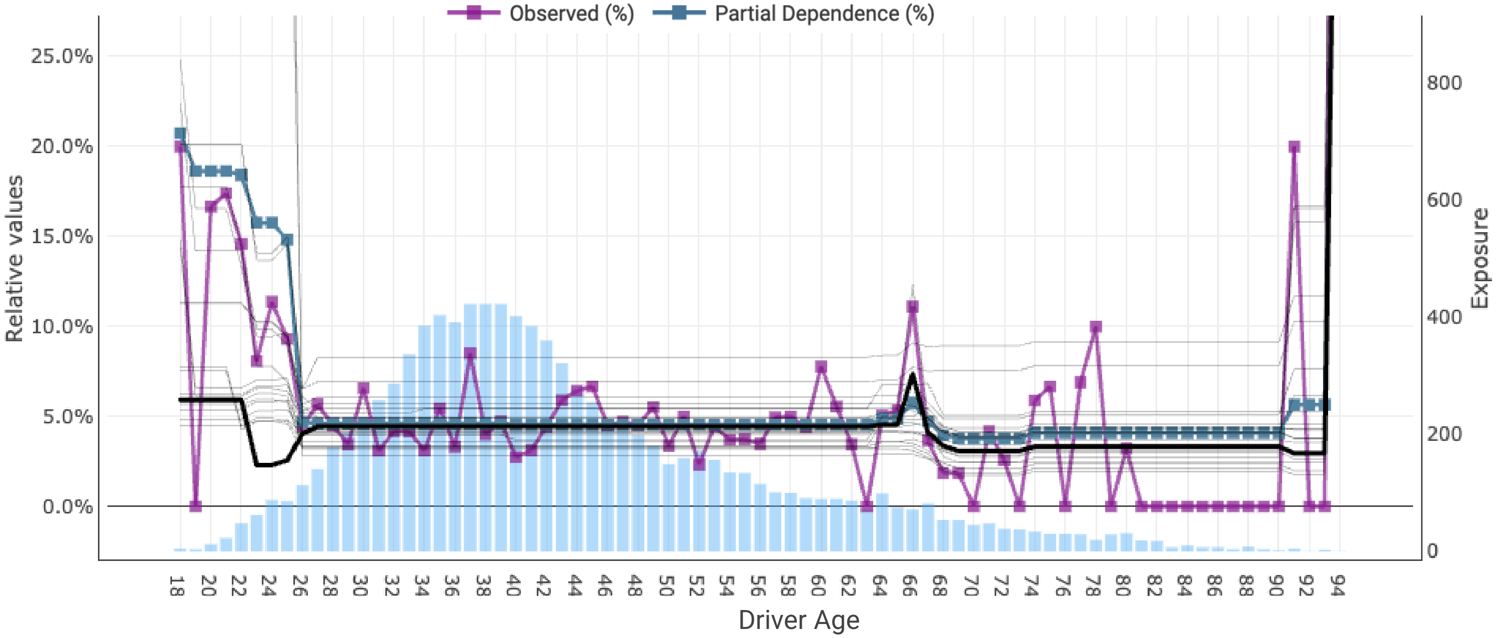


For example: a Partial Dependence Plot (PDP)) and Individual Conditional Expectation (ICE) showing the impact of a driver's age.

*Screenshot from Akur8's interface

Example of black-box analysis

ICE: visualize the conditional impacts



For example: a Partial Dependence Plot (PDP)) and Individual Conditional Expectation (ICE) showing the impact of a driver's age.

*Screenshot from Akur8's interface

The Dilemma

Trees Ensembles and GAMs

Strengths and Limits

Strengths associated with **Trees ensembles** models are related to their **creation process**.

Strengths associated with **GAMs** are related to their **models structure**.

Trees Ensembles	GAM
Models structure • Sum of small effect of all the variables • Trees depth	**Models structure** • Sum of effects of single variables.
Models Understanding • Via reverse-engineering or local analysis	**Models Understanding** • Direct visualization.
Models Creation • Machine learning	**Models Creation** • Human-creation • Machine-learning

Models creation & structure

	Creation Process	Result
GBM		Black Box (Trees Ensemble)
Random Forest		Black Box (Trees Ensemble)
Neural Network		Black Box (Neural Network)
Data-Prep + LM		Transparent (Data-Prep + LM.)
GAM (manual)		Transparent (GAM)
GAM (automated)		Transparent (GAM)

Mixing ML & Actuarial approaches

AKUR8

Global Parameters and Model Parameters

Applying ML to GAMs

It is possible to design an algorithm fitting GAMs, based on **2 global parameters**:

- **Level of smoothness**: how significant should the selected effects be?
- **Level of parsimony**: how many variable should be included in the model?

We developed this algorithm: Models can be **generated automatically** for many values of the global parameters (machine-learning Grid-Search approach), **tested on back-tests** and **fully analysed**.

GLOBAL PARAMETERS

- Smoothness level
- Parsimony level

DATA

- Explanatory Variables
- Target Variable

FIT

MODEL PARAMETERS

- Effect functions values
  (one function per selected variable)

The Fitting Process

Optimizing the Likelihood with Constraints

Maximum of Likelihood

Maximize the Likelihood of the observations.

This is the standard approach used in GLMs modeling, where the probability of observing the target given the predictors and a loss function (the likelihood) is optimized.

Smoothness Constraint

Similar to a **credibility approach**: all effects are supposed to be null.
This hypothesis is tested for every level and, **if the effect is significant** enough, it is included in the model.

More or less sensitive models are obtained by modulating the significance threshold: models selecting only significant effects will be very smooth and **robust**, models with more permissive threshold will be more **sensitive**.

Parsimony Constraint

In order to **improve the readability** of the models created, all the **least significant variables are removed** from the model.

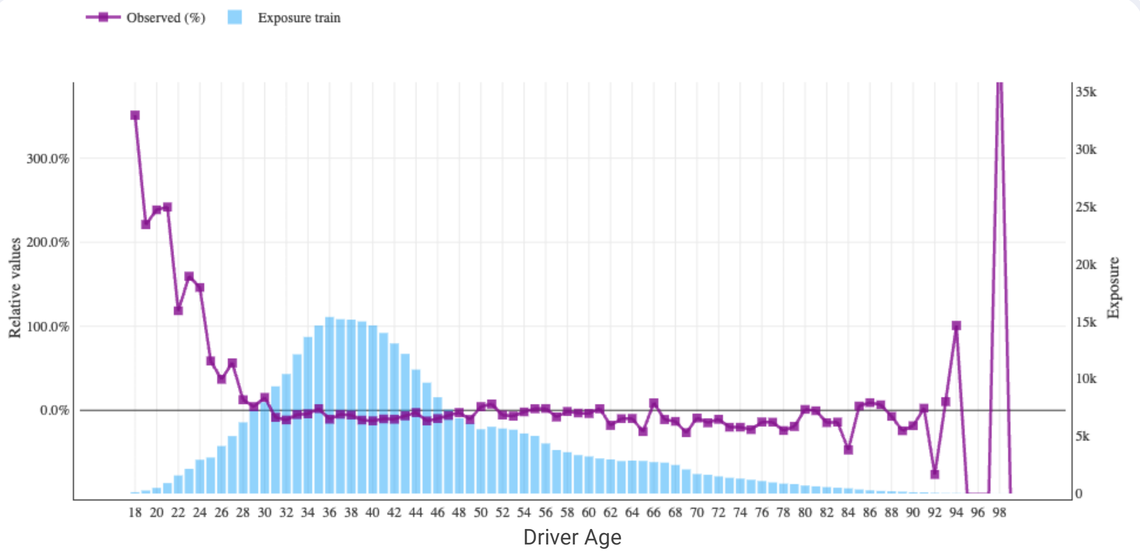
These are the variables that would provide the lowest gains in likelihood if included in the model.

This approach provides an optimal subset of variables to be included in the model.

1. Controlling the smoothness: Signal and Noise

Raw data contains both **signal** and **noise.**

A trade-off needs to be found between robustness and sensitivity.

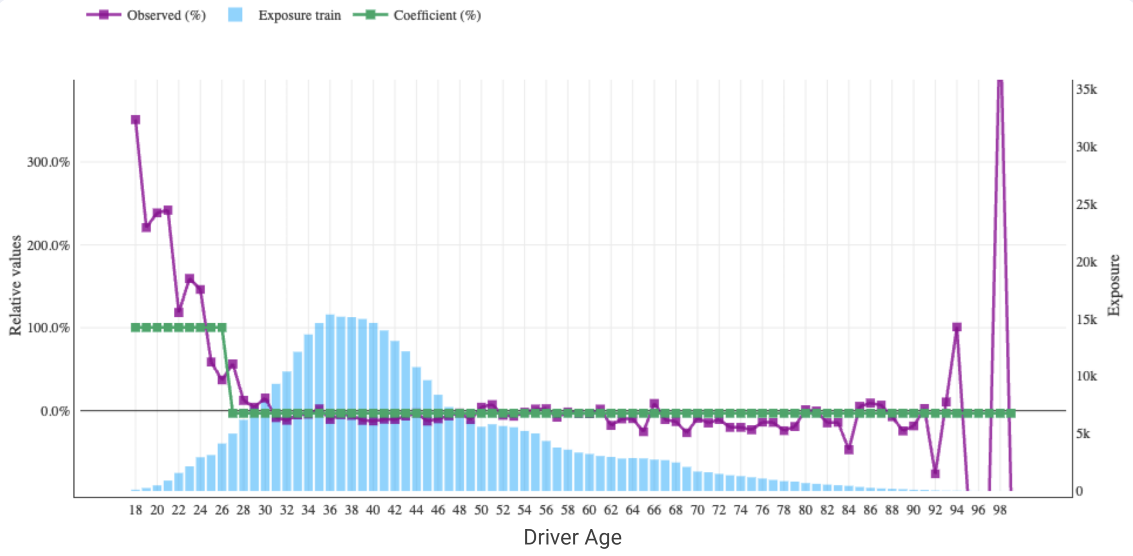


*Screenshot from Akur8's interface

1. Controlling the smoothness: Signal and Noise

Robust model

Missing part of the predictive signal

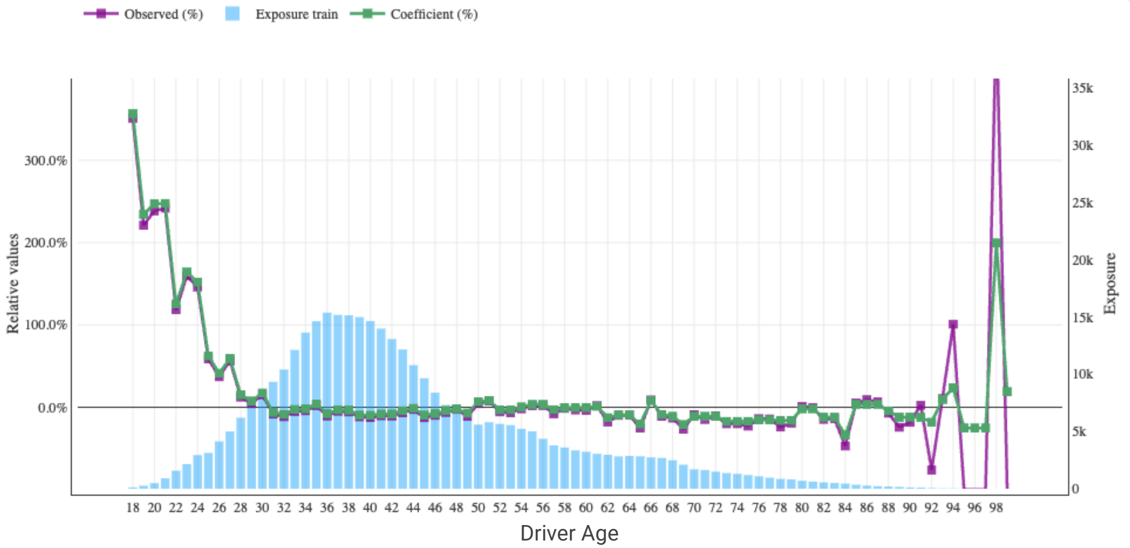


*Screenshot from Akur8's interface

1. Controlling the smoothness: Signal and Noise

Over-fitted model

Capturing noise

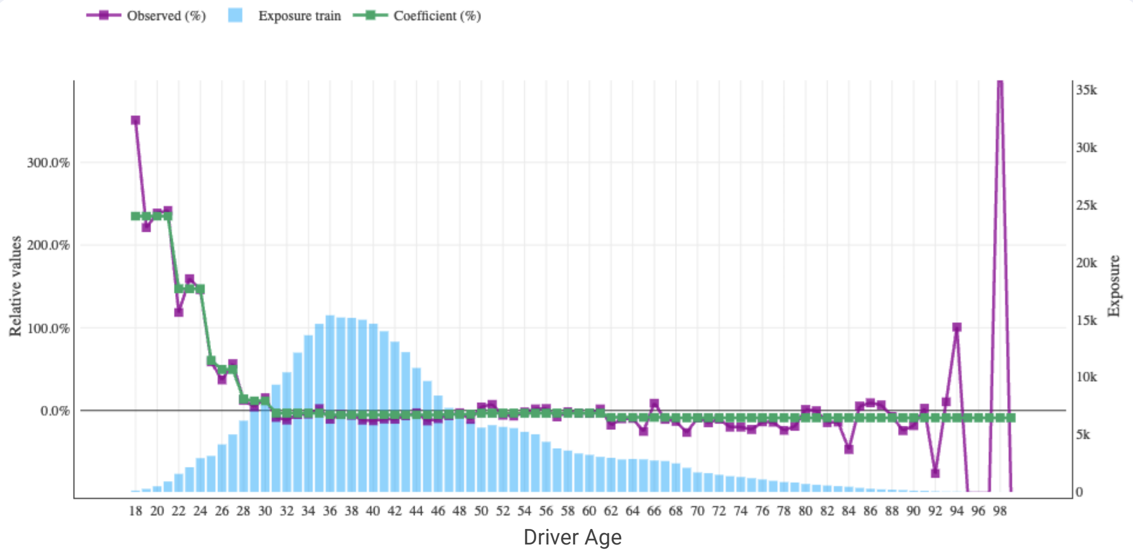


*Screenshot from Akur8's interface

1. Controlling the smoothness: Signal and Noise

Efficient model

Good trade-off, capturing signal and rejecting noise



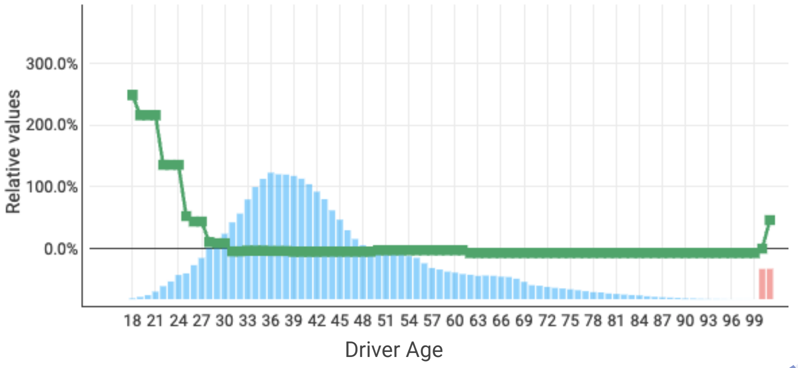
*Screenshot from Akur8's interface

1. Controlling the smoothness: Signal and Noise

What is overfitting?

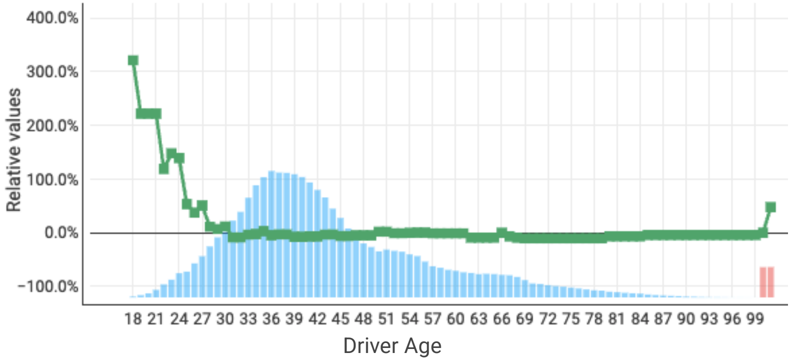
Which model should be selected?

Out-of-sample **Gini: 20.5%**
Robust model



Model on the left might lead to **better results** once **deployed in production**.

Out-of-sample **Gini: 21%**
Noisy Model

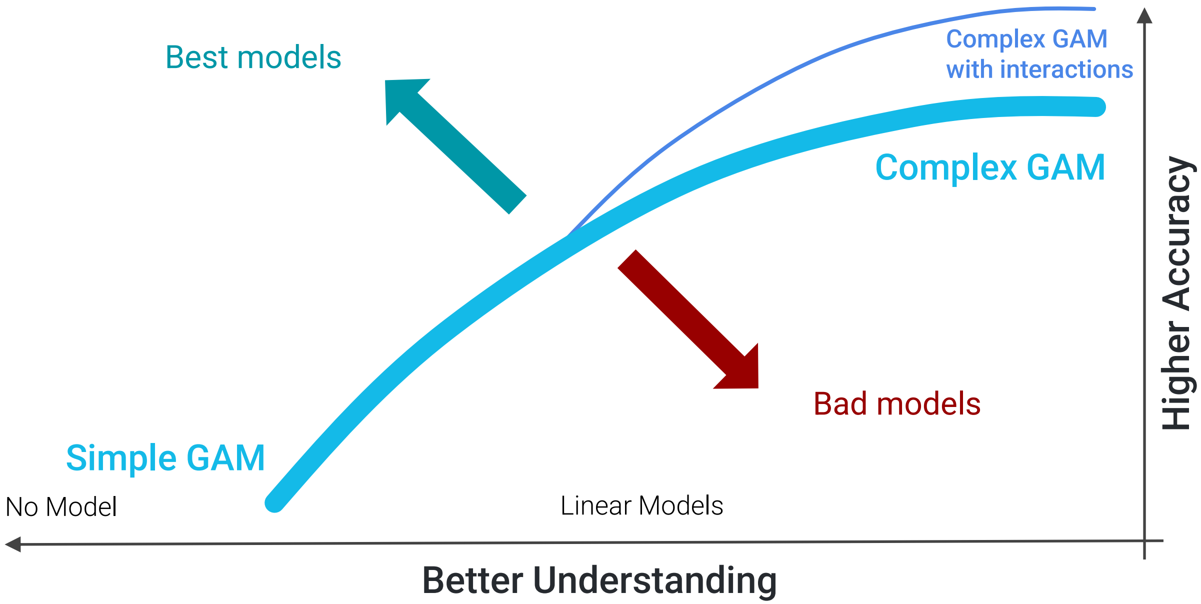


Model on the right has **stronger results on the back-test** but does **not inspire much trust**.

*Screenshot from Akur8's interface

2. **Parsimony has a cost** (but it is worth it)

Understanding / Accuracy trade-off

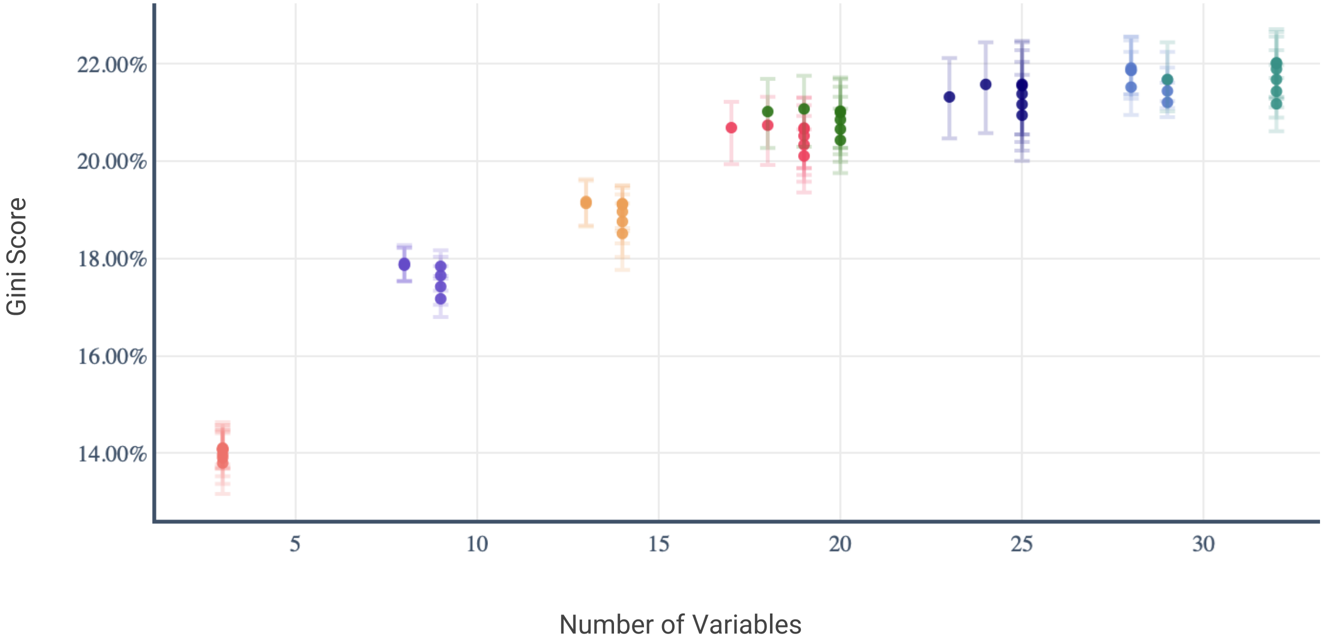


Black-box models
(GBMs, RF, NN…)

The accuracy is measured on a back-test; actual results when moving to productions will not be

2. Parsimony has a cost (but it is worth it)

Grid-search result



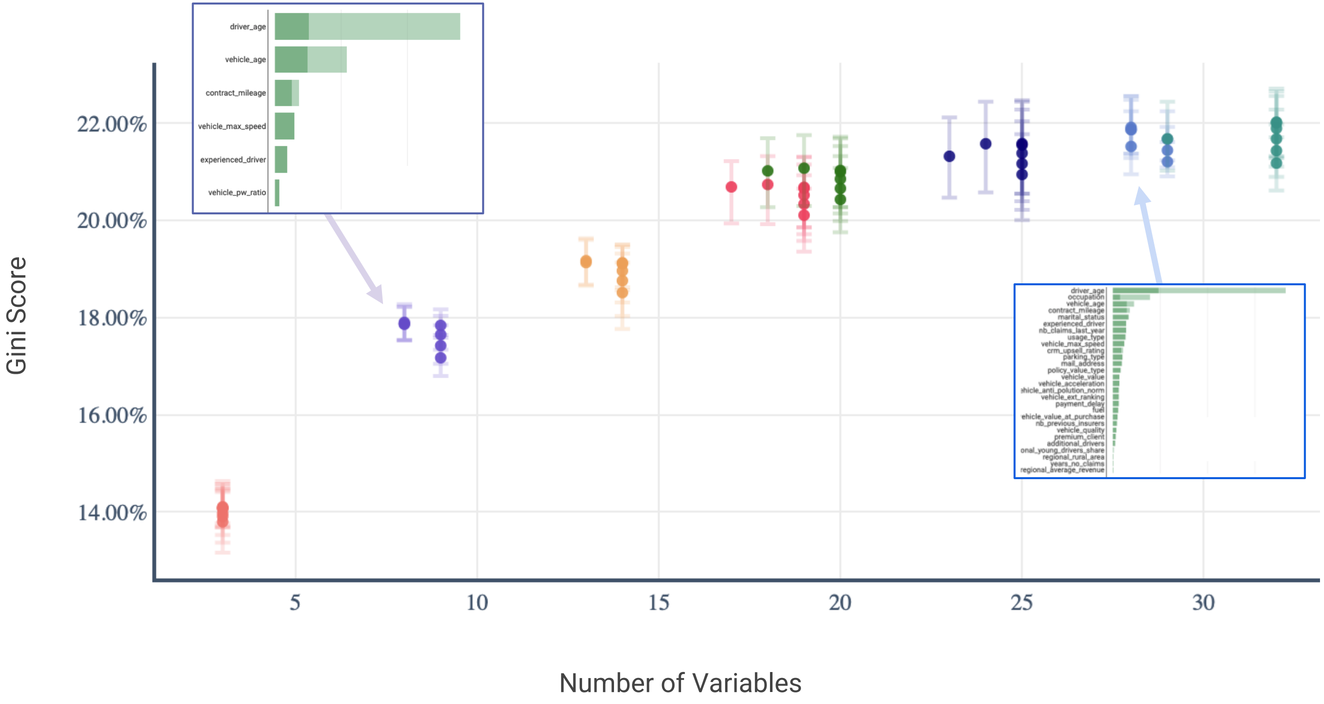
Grid-search results: each **point** represents one **model**.

The gain in models quality and the fading marginal improvement are clearly visible.

*Screenshot from Akur8's interface

2. Parsimony has a cost (but it is worth it)

Grid-search result



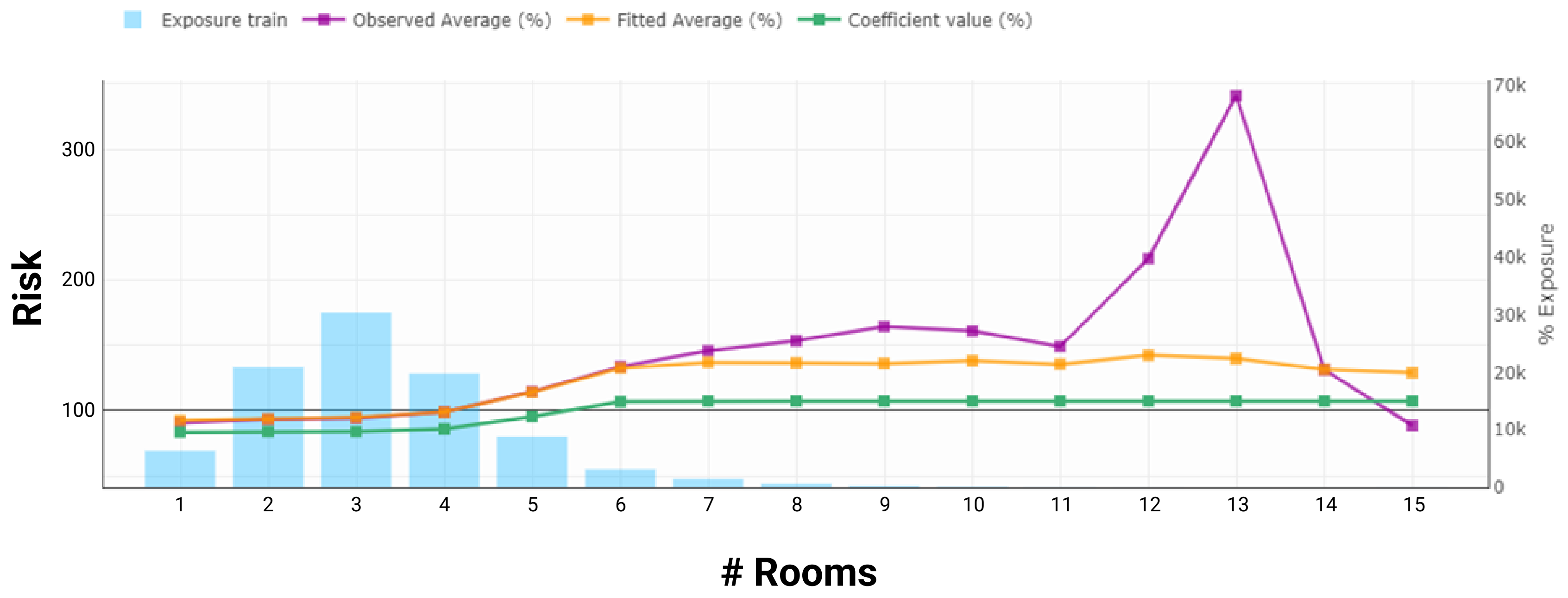
Grid-search results:
each **point** represents one **model**.

The gain in models quality and the fading marginal improvement are clearly visible.

*Screenshot from Akur8's interface

3. Interact with the models

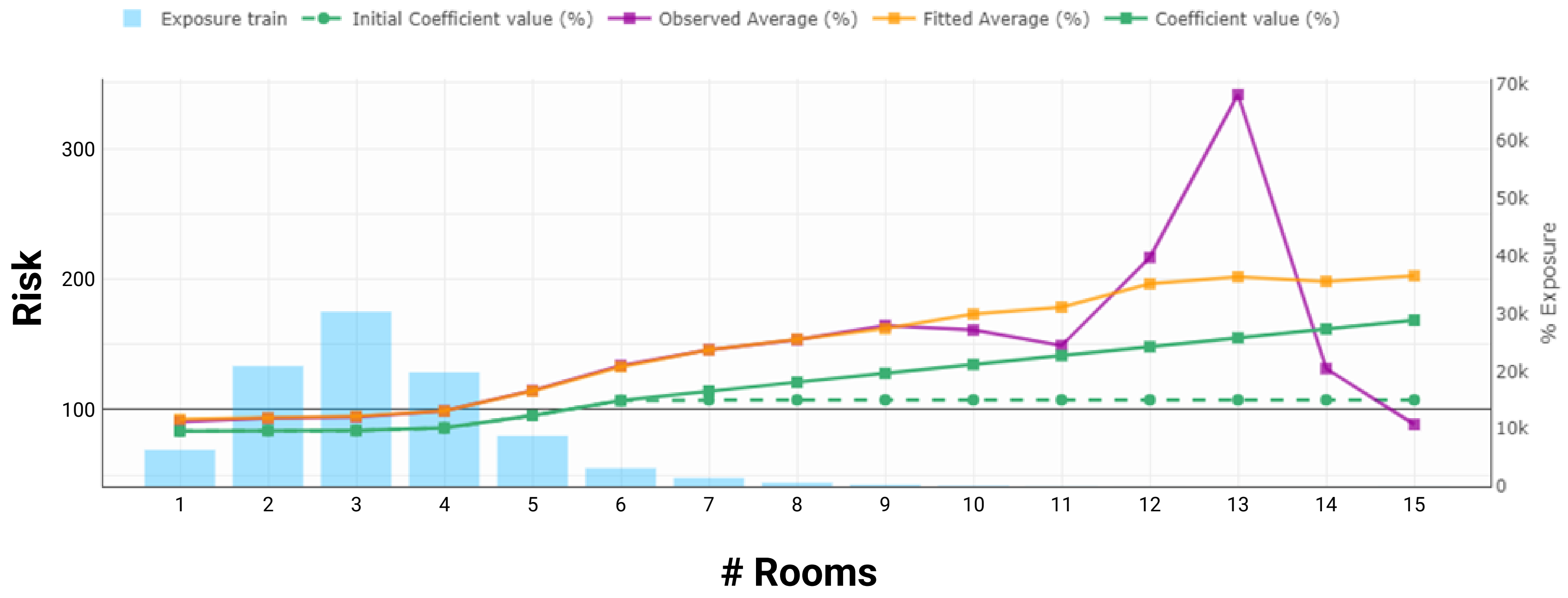
Spotting the issues is nice..



*Screenshot from Akur8's interface

3. Interact with the models

… solving the issues is better !



*Screenshot from Akur8's interface

3. Interact with the models

A 3-step process

It is possible to directly leverage a model right out of the fit process.

This would be similar to a classic data-science approach.

However, handling transparent models opens the possibility of interacting with them, integrating expert knowledge in the modeling.

So the process is (on purpose) mixing elements of:

- Machine-Learning: automated fit, purely **data-driven** model creation, acting on **global parameters** to control overfitting.
- Direct interaction with the models: control of all the **effects** captured in the fitting model, analysis and potentially edition of the **effects** to ensure a good extrapolation of the model.

GLOBAL PARAMETERS

- Smoothness level
- Parsimony level

DATA

- Explanatory Variables
- Target Variable

FIT

MODEL PARAMETERS

- Effect functions values
  Fitted from a purely data-driven process

ADAPT

FINAL MODEL PARAMETERS

- Effect functions values
  Adapted based on expertise, to ensure safe extrapolation on low-data segments

Conclusion

Mixing Data-Science automation and Actuarial Expertise

ML & Back-test performance

- Allows automated models creation
- Based on statistical criteria
- Easy to measure & reproduce
- Data-driven
- Pushes toward complexity over understanding

+

Actuarial expertise and transparency

- Minimizing the back-test error is not enough
- Performance can't be measures before deployments (and sometimes not even after)
- Direct interactions with the model itself is key to include all the operational constraints.

→

Understanding and capability to interact with a model is key; model's simplicity has value.

Models must allow the inclusion of expertise, safety and provide extrapolation capabilities.

Transparent modeling can and should be combined with machine-learning techniques.

Transparency is not "under-sophistication" or "primitiveness" but realism and efficiency.

Thank you!

Guillaume Béraud-Sudreau

guillaume.beraud@akur8.com

AKUR8