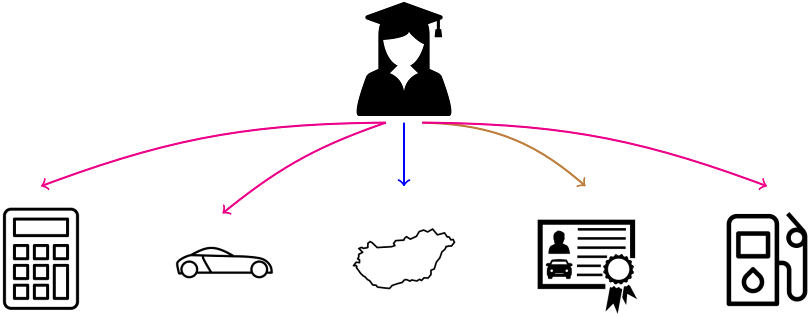
Sparsity with multi-type Lasso regularized GLMs

Sander Devriendt (email: sander.devriendt@kuleuven.be)
Joint work with K. Antonio, T. Reynkens, E. Frees, R. Verbelen

eRum 2018, Budapest
May 15, 2018



Claim frequency and claim severity

as function of

nominal / numeric $\sim$ ordinal / spatial

features

- Generalized Linear Models (GLMs) for frequency ($\sim$ Poisson) and severity ($\sim$ Gamma).
- How to:
  - (1) select variables or features?
  - (2) cluster (or bin or fuse) levels within a variable?

    age groups / postal code clusters / clusters of car models
- Procedure should be data driven, scalable to large (big) data.
- End product is interpretable, within actuarial comfort zone.

Research questions rephrased

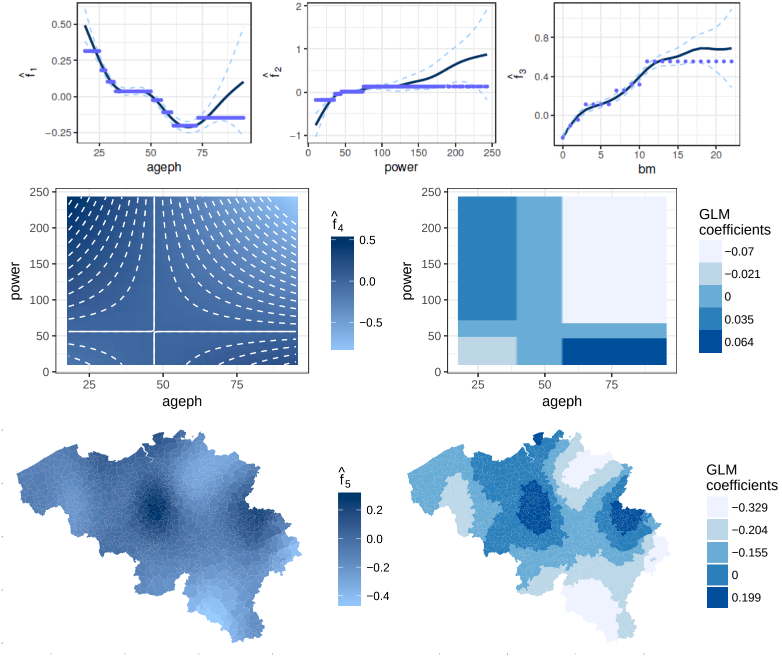
- Generalized Linear Models (GLMs) for frequency ($\sim$ Poisson) and severity ($\sim$ Gamma).
- How to:
  (1) avoid overfitting with too many variables or levels?
  (2) avoid underfitting with a priori binning/selection?

Henckaerts, Antonio et al., 2018 (Scandinavian Actuarial Journal)

Stepwise procedure

1. Do an exhaustive search through variables to find best GAM model.
2. Use well-chosen clustering algorithm to bin 2D spatial effect.
3. Use evolutionary trees to bin 1D continuous effects and interactions.
4. Fit GLM with bins and clusters obtained in previous steps.

R packages: `mgcv`, `classInt`, `evtree`, `rpart`


0.50
0.25
0.00
−0.25
25
50
75
ageph
$\hat{f}_1$
2
1
0
−1
0
50
100
150
200
250
power
$\hat{f}_2$
0.8
0.4
0.0
0
5
10
15
20
bm
$\hat{f}_3$
250
200
150
100
50
0
power
25
50
75
ageph
$\hat{f}_4$
0.5
0.0
−0.5
GLM
coefficients
−0.07
−0.021
0
0.035
0.064
$\hat{f}_5$
0.2
0.0
−0.2
−0.4
GLM
coefficients
−0.329
−0.204
−0.155
0
0.199

Sparsity with multi-type Lasso regularized GLMs

Devriendt, Antonio, Reynkens, Frees, Verbelen, 2018 (in progress)

$\boxed{\text{Standard GLM}}$

fit data as good as possible,
no constraint on parameters.

$\updownarrow$

$\boxed{\text{Regularized GLM}}$

tradeoff between fit and interpretability/sparsity/stability,
constraint on parameters.

- Less is more: (Hastie, Tibshirani & Wainwright, 2015)

  a sparse model is easier to estimate and interpret than a dense model.

- Regularize (with budget constraint t, or regularization parameter λ):

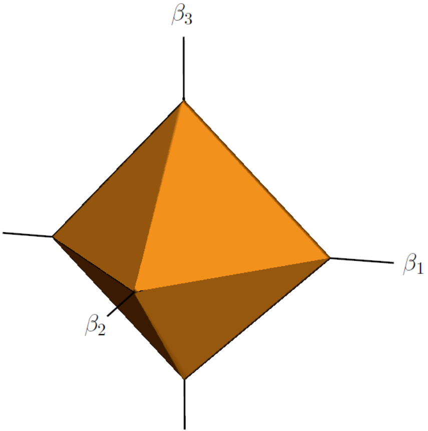
$$\min_{\beta_0, \boldsymbol{\beta}} \{-\mathcal{L}(\beta_0, \boldsymbol{\beta})\} \text{ subject to } \|\boldsymbol{\beta}\|_1 \leq t,$$

  or equivalenty

$$\min_{\beta_0, \boldsymbol{\beta}} \left\{ -\mathcal{L}(\beta_0, \boldsymbol{\beta}) + \lambda \cdot \sum_{j=1}^{p} |\beta_j| \right\}.$$

  Shrinks coefficients and even sets some to zero.

Regularization = limited budget for $\beta_1, \beta_2, \beta_3$.

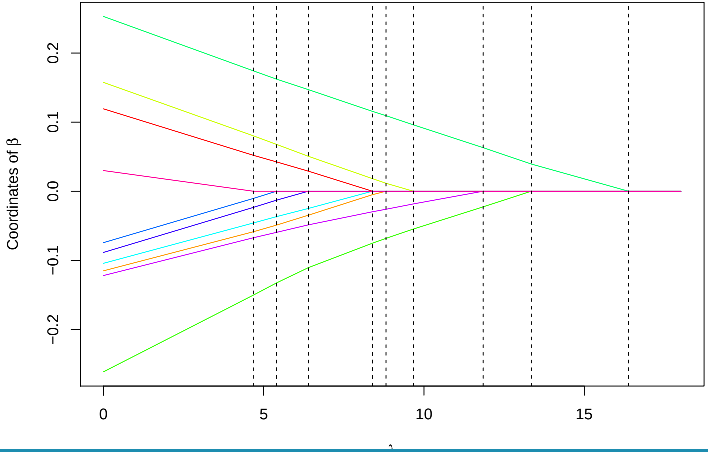


'Statistical Learning with Sparsity' - Hastie et al. (2015)

Lasso plot

Package `glmnet`

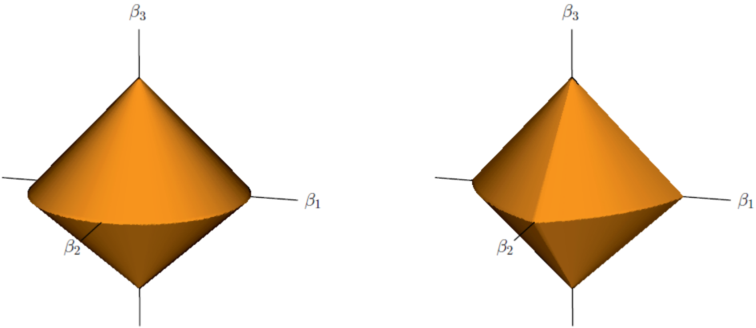
overfitting $\longleftarrow$ λ $\longrightarrow$ underfitting



- Adjust lasso regularization to the type of variable:
  - Determine type (nominal / numeric $\sim$ ordinal / spatial);
  - Allocate logical penalty.
- Thus, for J variables, each with regularization term $P_j(.)$, we want to optimize:

$$-\mathcal{L}\left(\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J\right) + \lambda \cdot \sum_{j=1}^{J} P_j\left(\boldsymbol{\beta}_j\right).$$

Different variable type $\rightarrow$ different penalty budget.

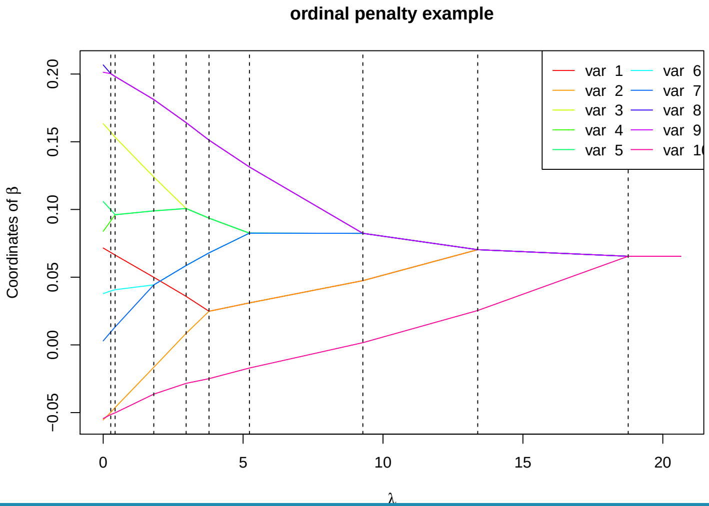


'Statistical Learning with Sparsity' - Hastie et al. (2015)

Fused Lasso

Package `genlasso`

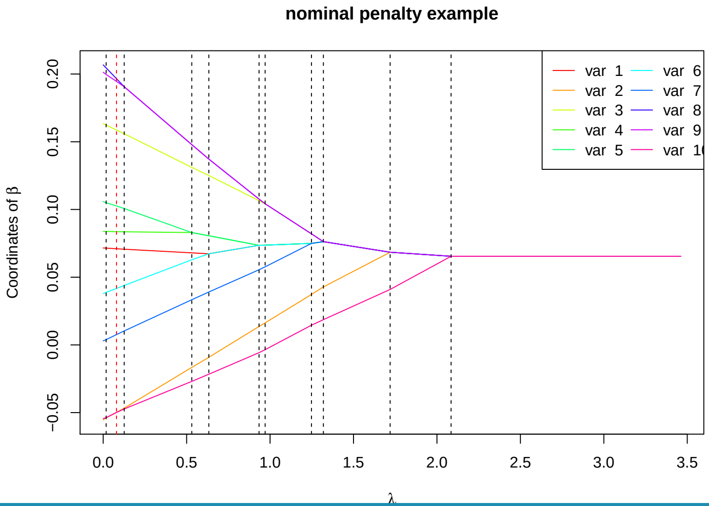
overfitting $\longleftarrow$ λ $\longrightarrow$ underfitting



Generalized Fused Lasso

Package `genlasso`

overfitting $\longleftarrow \quad \lambda \quad \longrightarrow$ underfitting



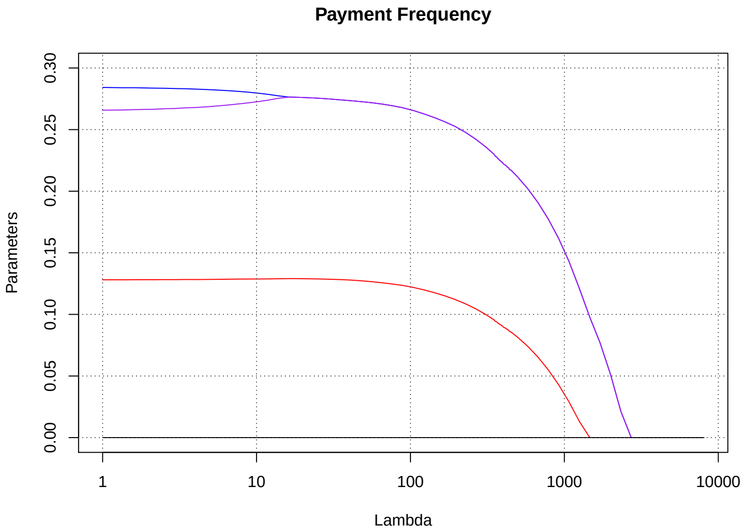
Unified GLM framework with multiple type of penalties

- Gertheiss & Tutz (2010) and Oelker & Gertheiss (2017):
  - GLMs with various penalties.
  - `R` package available: `gvcm.cat` (not maintained).
- Uses local quadratic approximations of penalties and PIRLS:
  - non-exact selection or fusion;
  - computationally intensive.

- Our contribution:
  - implements an efficient algorithm (with proximal operators);
    - code bottleneck in C++ (`Rcpp`)
    - efficient linear algebra (`RcppArmadillo`)
    - parallel computations (`parallel`)
  - scalable to big data (splits into smaller sub-problems);
  - flexible regularization
    - penalty takes type of variable into account;
    - works for all popular penalties;

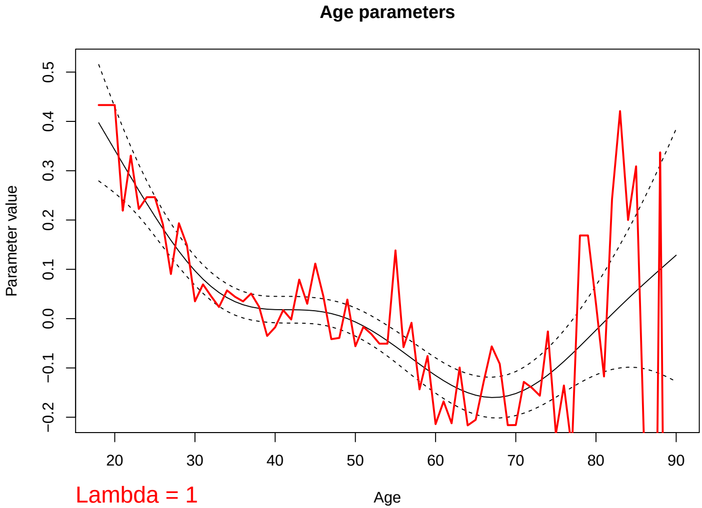
$\Rightarrow$ Package under construction.

- Frequency (and severity) information for $n = 163,234$ policyholders.
- 14 variables: binary, ordinal and nominal.
- Exposure modeled as offset.
- Fit Poisson GLM for frequency data with different penalties.
  - $N_i \sim \text{Poisson}(\mu_i)$
  - $\log(\mu_i) = \log(\text{exposure}_i) + \beta_0 + \sum_{j=1}^{14} X_j \boldsymbol{\beta}_j$
  - $\mathcal{O}(\boldsymbol{\beta}) = -\mathcal{L}\left(\beta_0, \boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{14}\right) + \lambda \cdot \sum_{j=1}^{14} P_j\left(\boldsymbol{\beta}_j\right)$



Payment Frequency
Parameters
0.00
0.05
0.10
0.15
0.20
0.25
0.30
1
10
100
1000
10000
Lambda

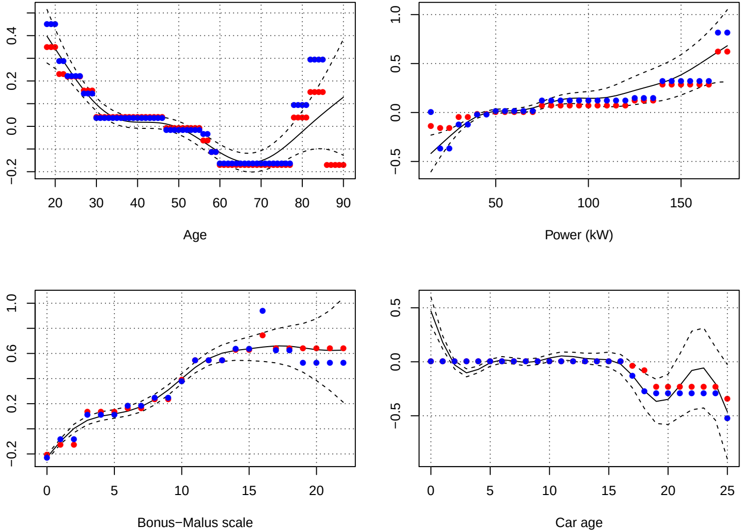
Case study: MTPL data



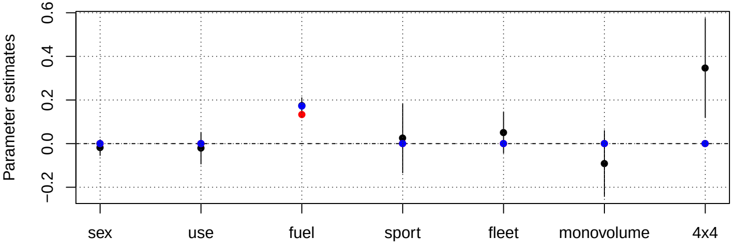
Lambda = 1

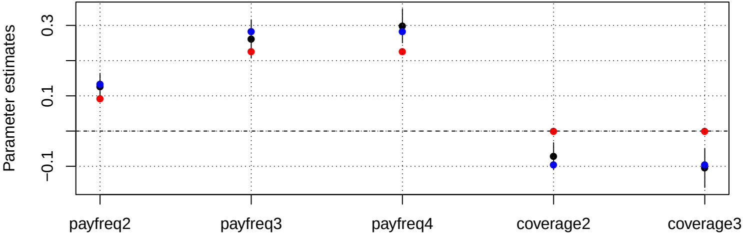
- Settings:
  - Incorporate adaptive (GLM) and standardization weights for better consistency and predictive performance.
  - Tune λ with out-of-sample MSE ($\hat{\lambda} = 380$)
- Re-estimate the final sparse GLM with standard GLM routines (from 164 to 38 params.).

MTPL claim frequency with multiple type of penalties



GAM fit, penalized GLM fit, GLM refit with new clusters.





GAM fit, penalized GLM fit, GLM refit with new clusters.

Wrap-up

- Less is more.
- Flexible regularization can help predictive modeling.
- R package combines general framework with efficient algorithm.
- Package and working paper to be finalized.

Thank you

Ageas Continental Europe

+ Tom Reynkens and colleagues

Henckaerts, R., Antonio, K., Clijsters, M. and Verbelen, R. (2018)
A data driven strategy for the construction of insurance tariff classes.
Scandinavian Actuarial Journal, published online.

Wood, S. (2006)
Generalized additive models: an introduction with R.
Chapman and Hall/CRC Press.

Gertheiss, J. and Tutz, G. (2010).
Sparse modeling of categorial explanatory variables.
The Annals of Applied Statistics, 4(4), 2150-2180.

Oelker, M. and Gertheiss, J. (2017).
A uniform framework for the combination of penalties in generalized structured models.
Advances in Data Analysis and Classification, 11(1),97-120.

Parikh, N. and Boyd, S. (2013).
Proximal algorithms.
Foundations and Trends in Optimization, 1(3):123-231.

Hastie, T., Tibshirani, R. and Wainwright, M. (2015)
Statistical learning with sparsity: the Lasso and generalizations.
Chapman and Hall/CRC Press.