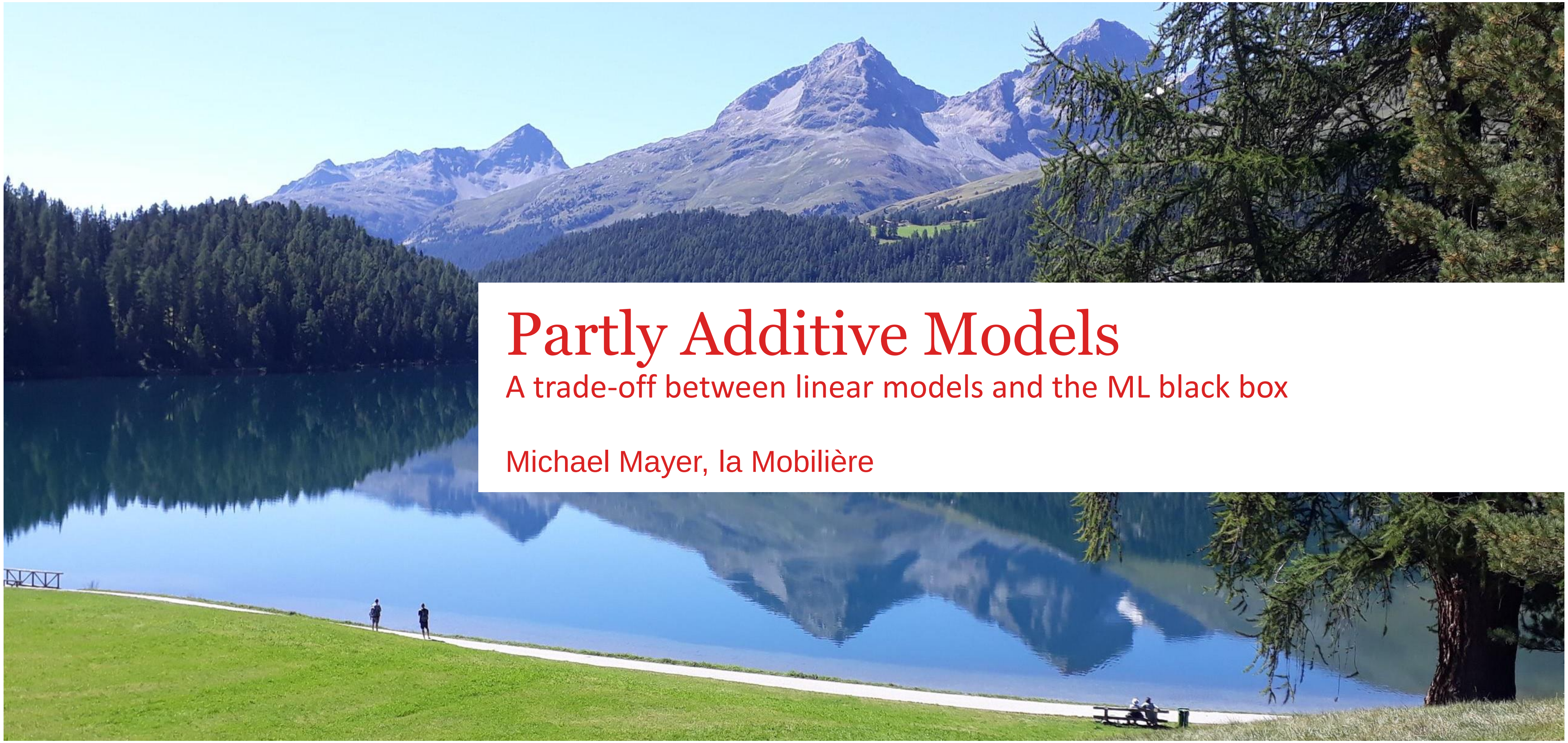


Partly Additive Models

A trade-off between linear models and the ML black box

Michael Mayer, la Mobilière

Use Case

Car Collisions with Large Animals

Annual figures for Switzerland

- Many large animals die on Swiss roads, among them ~8'000 deers
- Total vehicle damage: ~25 Mio CHF
- Covered by partial damage coverage of motor insurance ("Teilkasko")

Source: https://de.wikipedia.org/wiki/Wildunfall

Annual figures for Swiss Mobiliar

- Ca. 2'000 animal collision claims
- Claim frequency ~0.4%
- ~4 claims per 1000 car years

Question to randomly picked pricing actuary

?

How does claim frequency depend on

- traditional features and
- postal code level information?

E(Answer)

- Poisson regression model for the expected claim frequency
- Using historic data on policies and claims

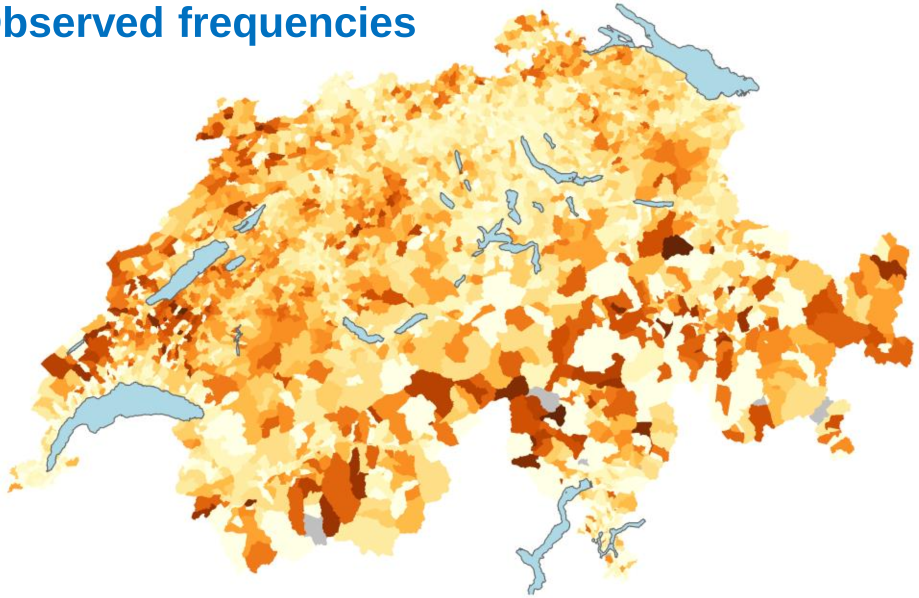
!

Data

	Y: Response	X: Traditional features					R: 14 regional features			
policy	**claim frequency**	**year**	**driver age**	**car age**	**car price**	**...**	**postal code**	**forest %**	**pop. density**	**…**
A	0	2017	40	4	40,000		1201	3%	36,000	
A	0	2018	41	5	40,000		1201	3%	36,000	
B	0	2017	28	1	18,000		3012	34%	3,600	
B	0	2018	29	2	18,000		3012	34%	3,600	
B	1	2019	30	3	18,000		3012	34%	3,600	

*Fictive data rows for illustration

Observed frequencies



- Data over multiple years
- Millions of rows
- Train/test split grouped on policy

Modeling Techniques

Linear Model or ML Black Box?

Claim frequency model with log link

$$\log\big(E(Y)\big) = f(\text{traditional features } X, \text{regional features } R)$$

	Model structure	Performance	Interpretation	Regional effect
GLM	$\beta X + \gamma R$	😐	😍	$\hat{\gamma} R$
ML black box	$f(X, R)$	🙂	😒	Depends on X
Partly additive ML model (??)	$\sum f_j(X_j) + g(R)$	?	?	$\hat{g}(R)$
GLM with proxy	$\beta X + \delta \hat{g}(R)$	?	?	$\hat{\delta}\hat{g}(R)$

My dream

- All models minimize Poisson deviance
- GLM does not use postal code (→ GLMM?)
- ML black box = boosted trees model (LightGBM tuned with grouped 5-fold CV)

Partly Additive Model, but how??

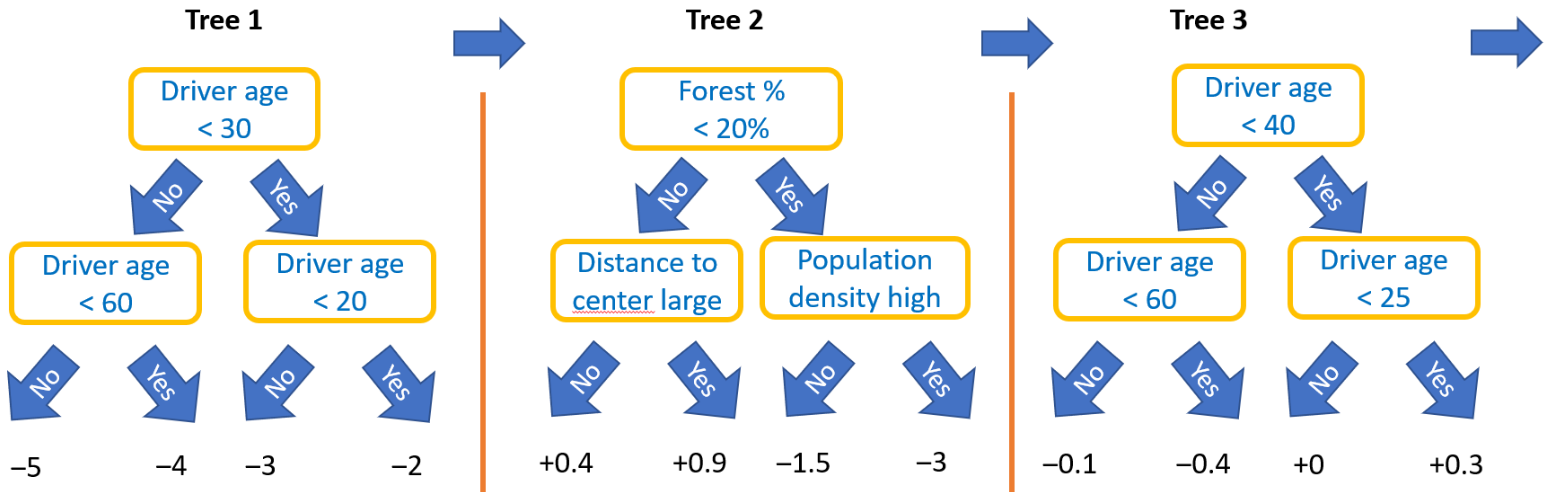
- Boosted trees with interaction constraints
  Simon Lee, Sheldon Lin, and Katrien Antonio (2015)
- Deep learning
- Component-wise boosting
  Peter Bühlmann and Torsten Hothorn (2007)

Will use this approach (LightGBM tuned with grouped 5-fold CV)

Boosted Trees with Interaction Constraints

- Feature groups define variables allowed to interact
- Simple example: $\{\{\text{Driver age}\}, \{\text{Forest \%}, \text{Distance to center}, \text{Population density}\}\}$
- Build trees such that each branch uses features from one feature group only

$$f(\text{Driver age}) + g(\text{Forest \%}, \text{Distance to center}, \text{Population density})$$

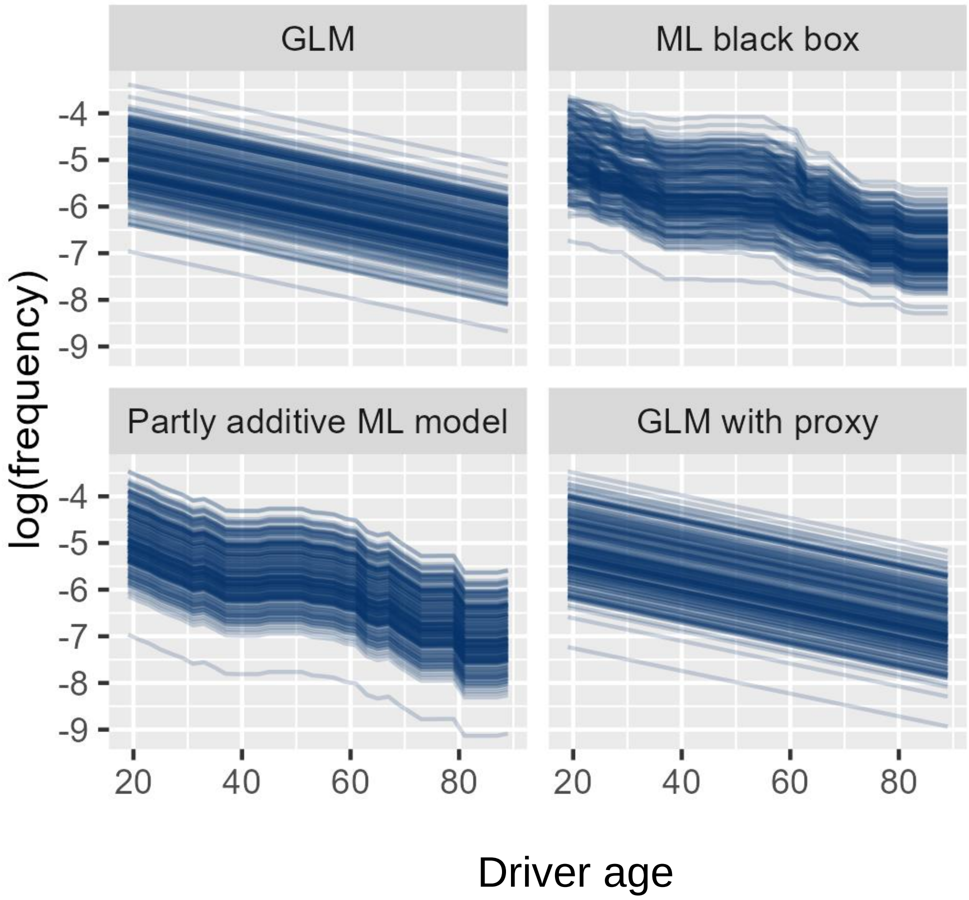


Implemented in
- XGBoost (2018),
- LightGBM (2020),
- Scikit-Learn (2022?)

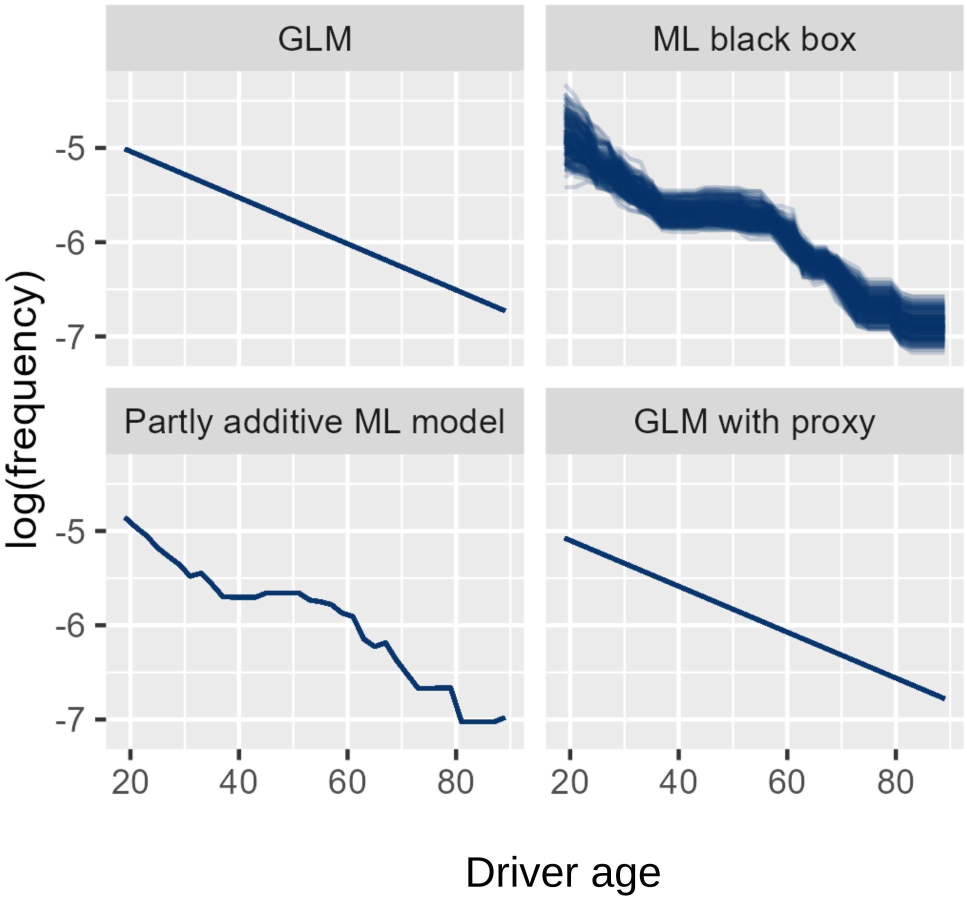
Results

Effect of Traditional Feature: Driver Age

Individual conditional expectation (ICE) plots



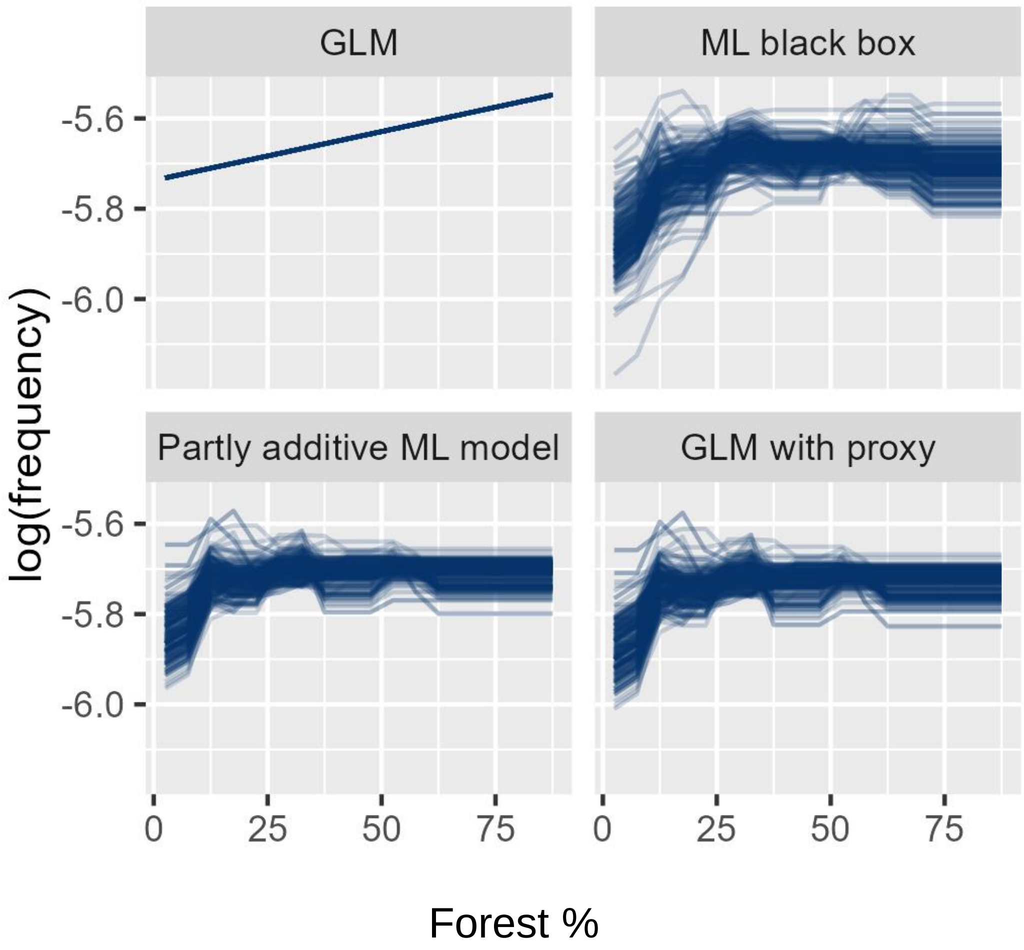
Centered ICE plots



Additivity → complete description of effect

Effect of Single Regional Feature: Forest %

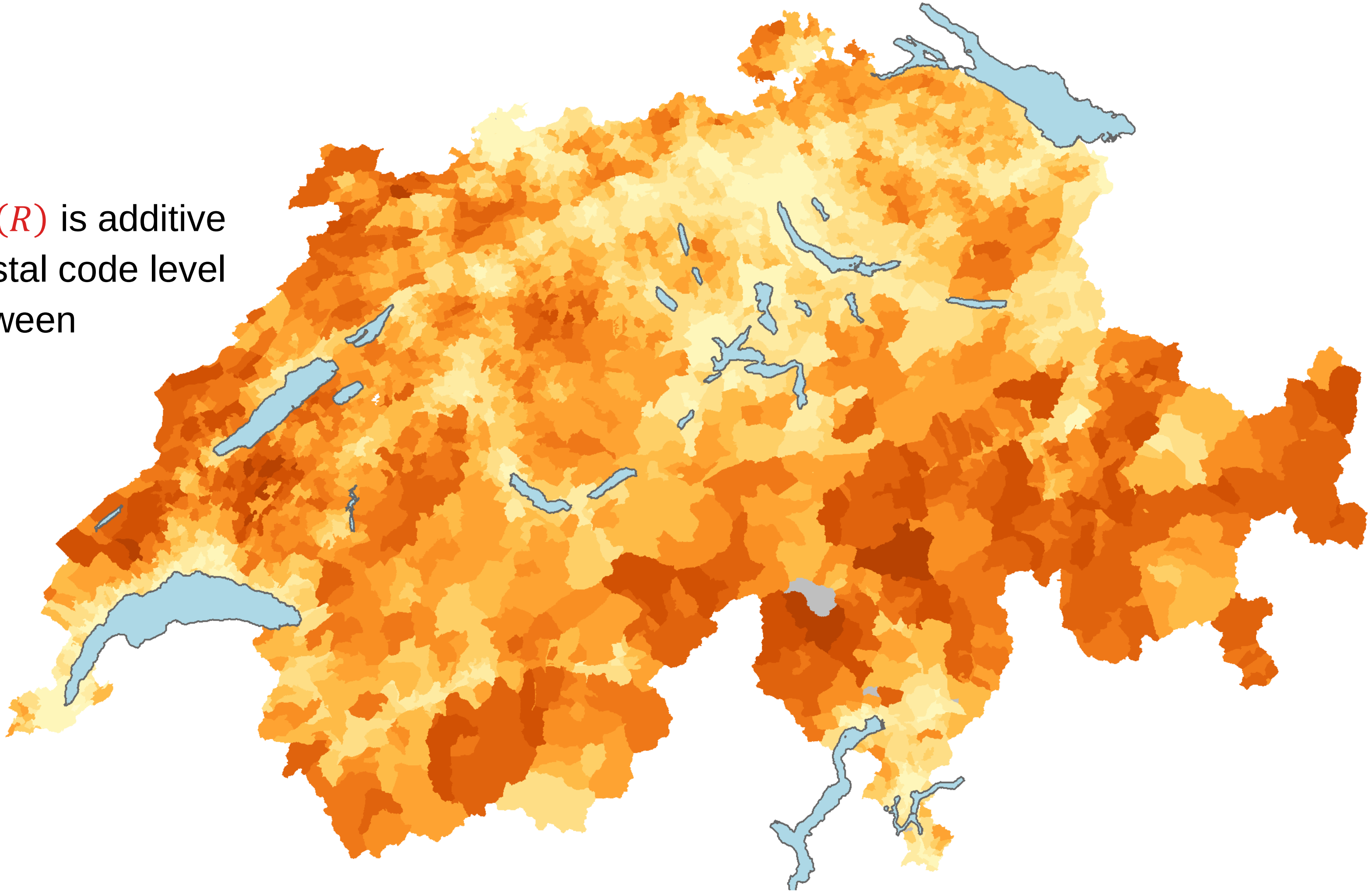
Centered ICE plots



- Effect depends on other regional factors
- Somewhat interpretable
- Helps to check model

Combined Regional Effect of Partly Additive ML Model

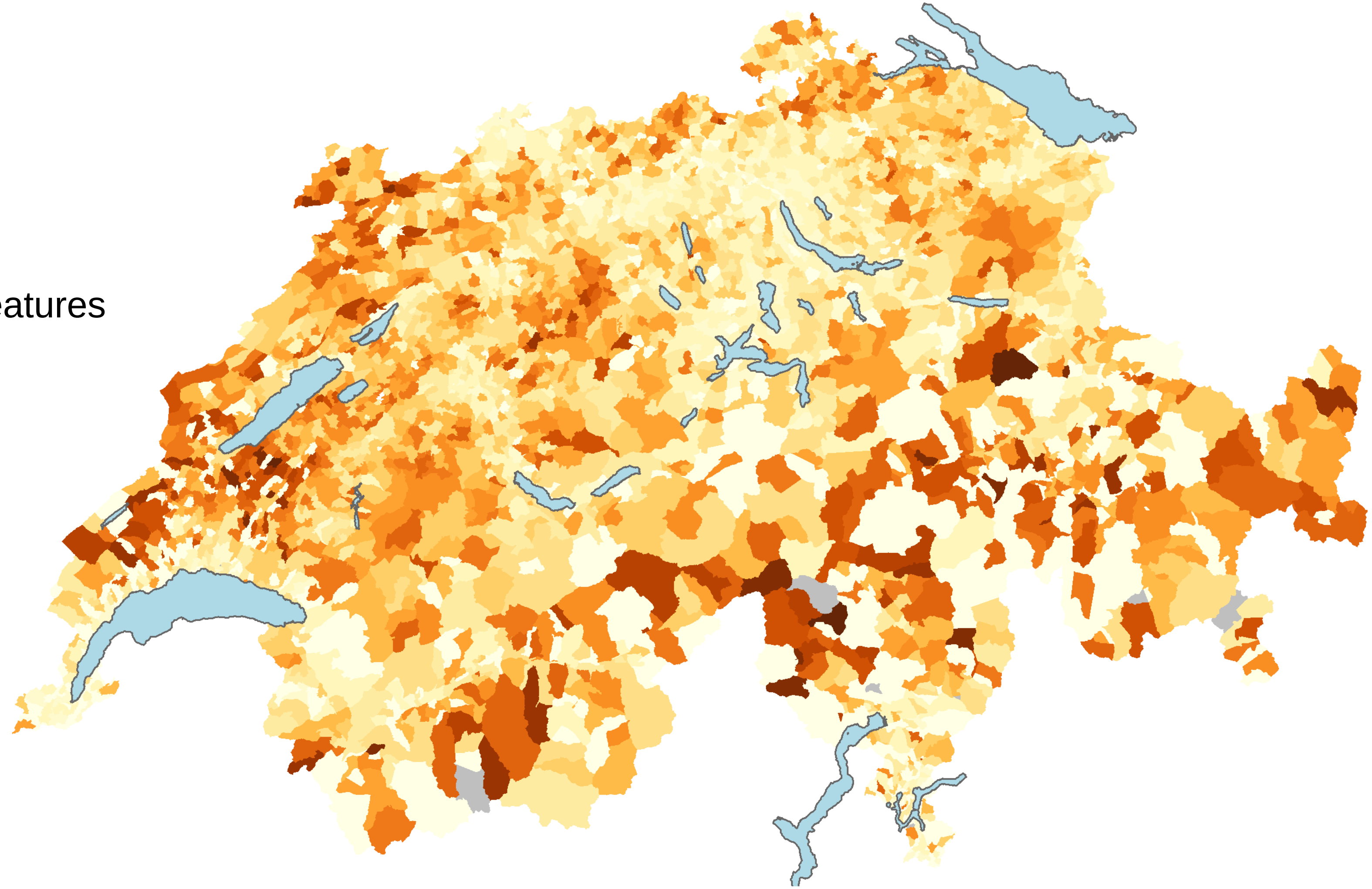
- $\log(E(Y)) = \sum f_j(X_j) + g(R)$
- Combined regional effect $\hat{g}(R)$ is additive
- Complete description at postal code level
- Thanks to 1:1 mapping between R and postal code



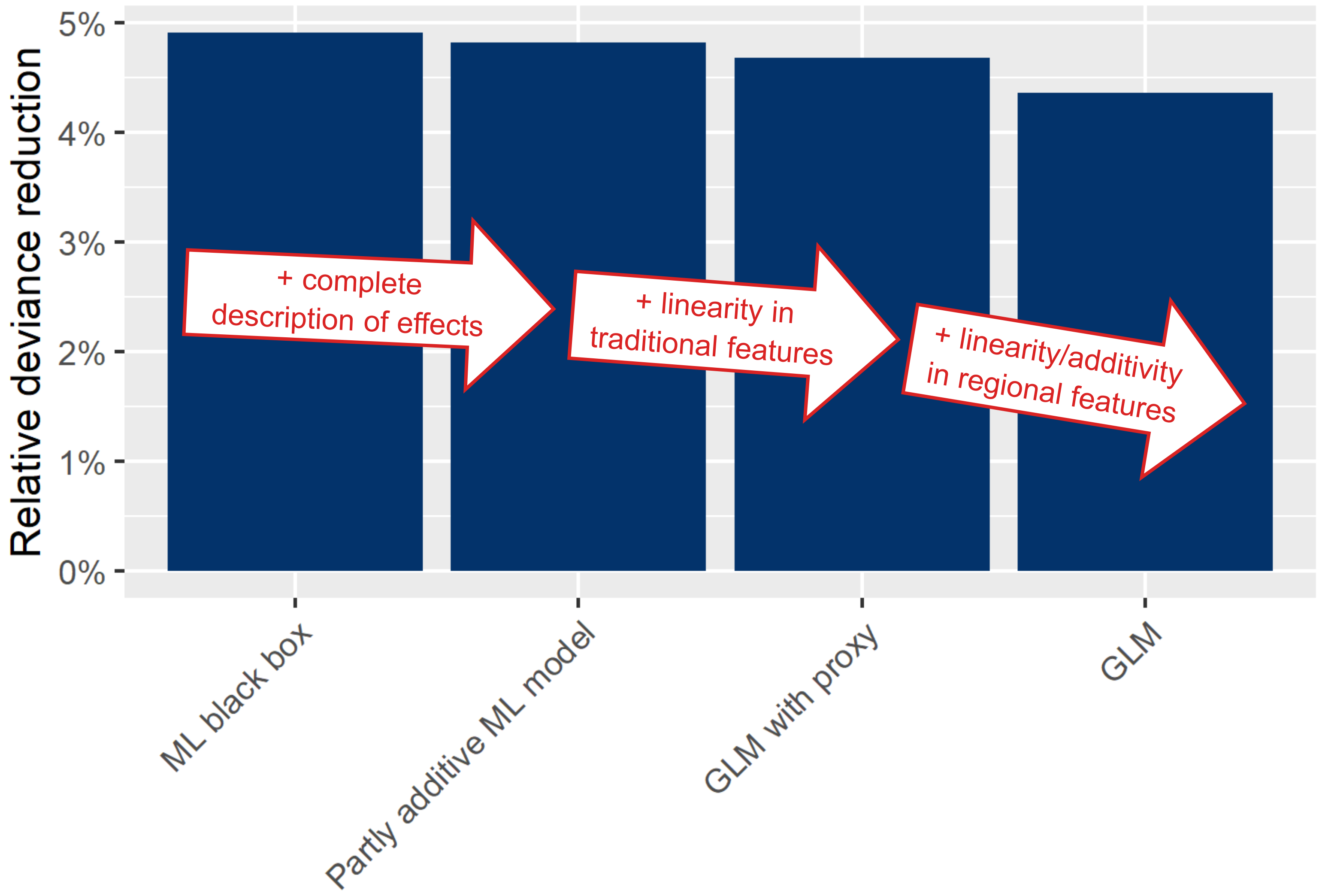
Compare with Observed Frequencies

Effect is

1. smoother and
2. adjusted for traditional features



Model Performance on 20% Test Data



Key Takeaways?

Excellent Trade-off between Performance and Interpretability

	Model structure	Performance	Interpretation	Regional effect
GLM	$\beta X + \gamma R$	😐	😍	$\hat{\gamma} R$
ML black box	$f(X, R)$	🙂	😒	Depends on X
Partly additive ML model	$\sum f_j(X_j) + g(R)$	🙂	🙂	$\hat{g}(R)$
GLM with proxy	$\beta X + \delta \hat{g}(R)$	🙂	😍	$\hat{\delta} \hat{g}(R)$

Other use cases of partly additive ML models?

- $\sum f_j(X_j) + g(R) + h(\text{Car features})$
- $h(\text{Time}) + g(\text{other features})$
- $h(\text{Gender}) + g(\text{other features})$

Resources

- Lee, S., Lin, S., and Antonio, K. (2015). Delta Boosting Machine and its Application in Actuarial Modeling. Institute of Actuaries of Australia

- Bühlmann, P., and Hothorn, T. (2007). Boosting Algorithms: Regularization, Prediction and Model Fitting. Statistical Science 22

- Mayer, M., Bourassa, S.C., Hoesli, M., and Scognamiglio, D. (2022). Machine Learning Applications to Land and Structure Valuation. Journal of Risk and Financial Management

- R code of similar case study
  https://github.com/mayer79/star_boosting

Questions?

Backup: Partly Additive Model with Deep Learning

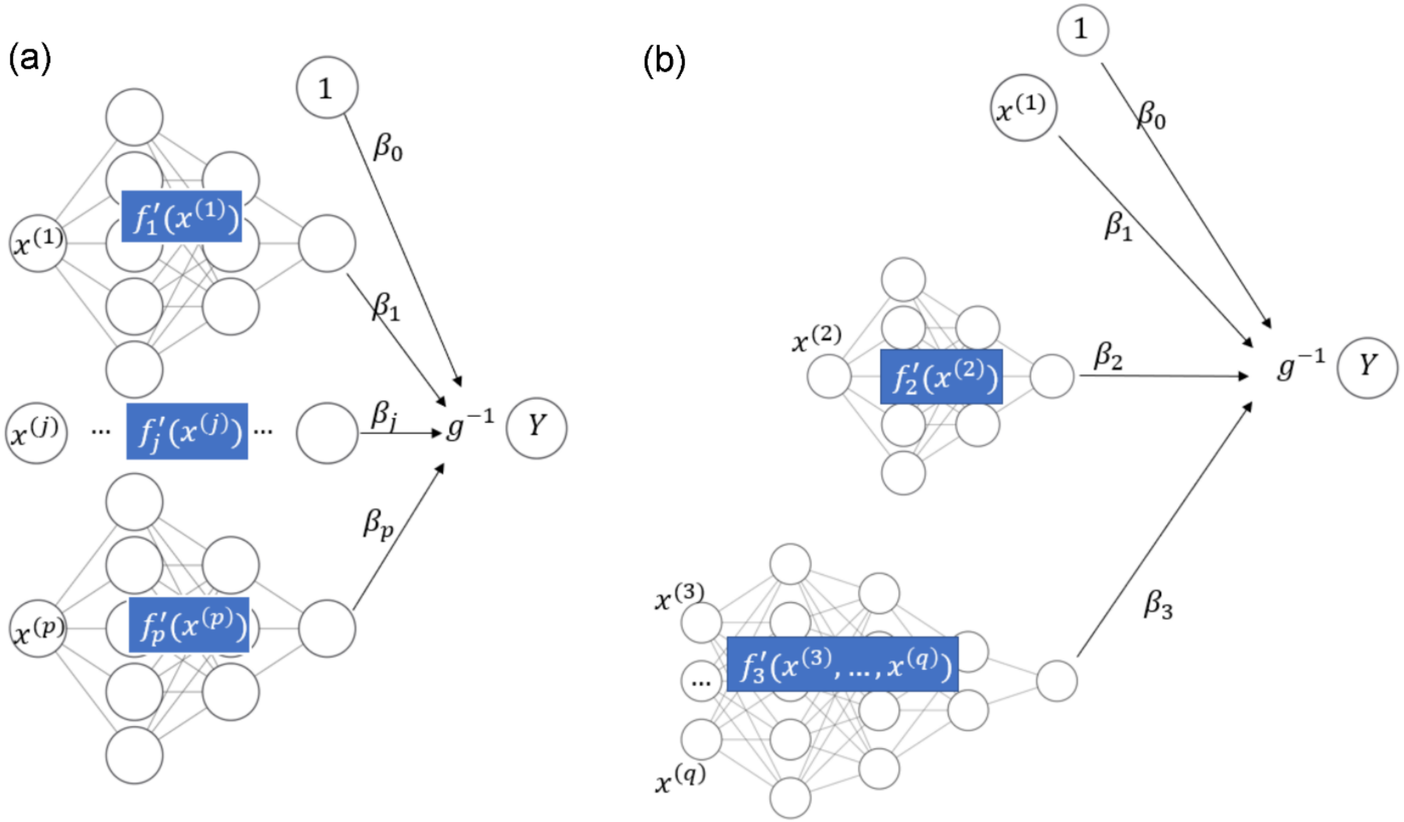


Figure 1 from Mayer, M.; Bourassa, S.C.; Hoesli, M.; Scognamiglio, D. (2022)
Machine Learning Applications to Land and Structure Valuation, Journal of Risk and Financial Management