UNIVERSITY OF OSLO

Advances in predictive modeling with tree-based machine learning

Applications to automated valuation models

Anders Dahl Hjort

February 28th, 2025

We use statistical models to make predictions

Training data $(x_1, y_1), \ldots, (x_N, y_N)$ → **Model** → Prediction $(x_{N+1}, ?)$

Historically, parametric models have been preferred

Training data $\rightarrow$

$$\boxed{\begin{aligned} y &= F_\theta(\boldsymbol{x}) + \varepsilon \\ \varepsilon &\sim \mathcal{N}(0, \sigma^2) \end{aligned}}$$

$\rightarrow$ Prediction

E.g., linear regression

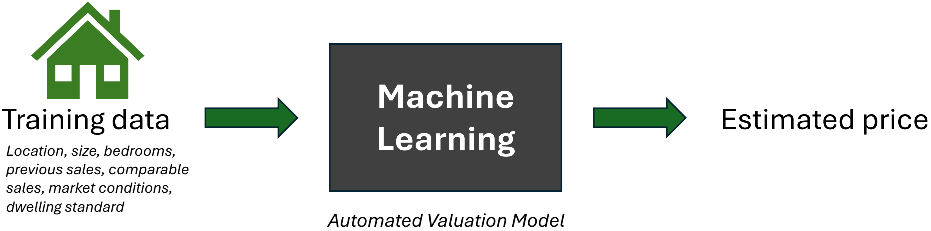
Machine learning models are increasingly being used

E.g., neural networks, random forests, gradient boosted trees

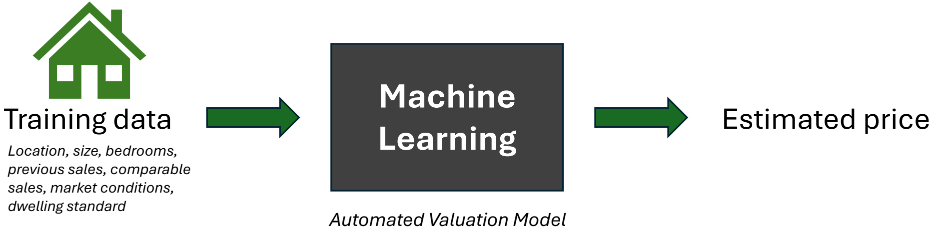
Machine learning comes with new challenges

How accurate is the model?
How should we interpret the model?
How certain are we about the predictions?

These models can be used for house price prediction



Industrial collaboration with Eiendomsverdi



Eiendomsverdi®

UNIVERSITY
OF OSLO

AVMs have several interesting use cases

Training data

Location, size, bedrooms, previous sales, comparable sales, market conditions, dwelling standard

→

Automated Valuation Model

Homeowner: "What is the value of my home?"

Bank: "What is the collateral value of these 100 000 mortgages?"

Real estate agent: "What should we list this home for?"

Central Bank: "How much did house prices rise or fall last year?"

Insurance: "What is a fair payout for a damaged home?"

Government: "What should people pay in property tax?"

AVMs are increasingly applied

Tech Policy .PRESS

AI in Property Valuation: The Most Consequential Algorithms You've Never Heard Of

ALEX ENGLER, SYLVIA BROWN / OCT 9, 2023

If we told you about an AI built on the latest foundation models that shapes multi-trillion-dollar markets and 'walks' through every home in the United States, would you say it was science fiction?

FINANCIAL TIMES Subscribe Sign In

Opinion **Artificial intelligence**

Welcome to a world where AI can value your home

If you thought that your neighbours were judgmental about the state of your front lawn, get ready for the bots

… but can also make mistakes


THE WALL STREET JOURNAL.
SIGN IN
SUBSCRIBE
BUSINESS
What Went Wrong With Zillow? A Real-Estate Algorithm Derailed Its Big Bet
The company had staked its future growth on its digital home-flipping business, but getting the algorithm right proved difficult


CURBED
AUTOMATION | NOV. 4, 2021
Why Did So Many Homeowners Sell to Zillow? Because It Overpaid.
By Kim Velsey

Zillow: Fun for browsing, but don’t trust the Zestimates. Photo: Tiffany Hagler-Geard/Bloomberg via Getty Images

This thesis

Training data

Location, size, bedrooms, previous sales, comparable sales, market conditions, dwelling standard

Automated Valuation Model

How accurate is the model? ***Paper 1***

How should we interpret the model? ***Paper 2***

How certain are we about the predictions? ***Paper 3 & 4***

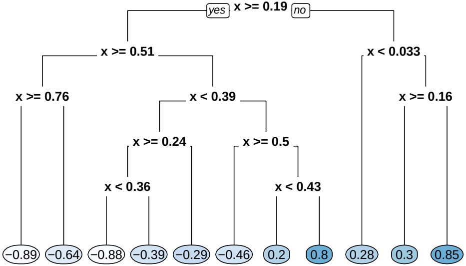
Estimated price

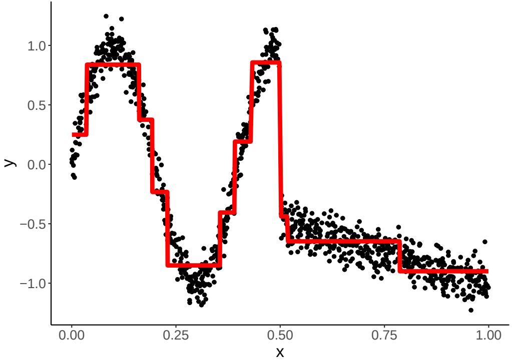
Paper I:
House price prediction with gradient boosted trees under different loss functions

Authors: Hjort, Anders; Pensar, Johan; Scheel, Ida; Sommervoll, Dag Einar

Published in: *Journal of Property Research 39 (4), 338-364, 2022*

A regression tree partitions the feature space





Gradient boosted trees (GBT)

- Aggregate multiple decision trees to form an ensemble,

$$\hat{f}(\boldsymbol{x}) = \sum_{m=1}^{M} \alpha_m \cdot \hat{h}_m(\boldsymbol{x}; q_m)$$

  where $\hat{h}_m$ is a decision tree with tree structure q_m, e.g., tree depth. The learning rate α_m regularizes the learning process
- Trees trained sequentially on residuals (gradients of loss function) from previous tree $\implies$ lowers bias
- ✔ Flexible, computationally efficient, high accuracy
- ✘ Hard to interpret, challenging to quantify uncertainty

Paper 1

Motivation:

- Compare tree-based ML methods to traditionally used pricing models
- Address the role of the loss function and its relation to the Within ±20% performance measure

Paper 1

Motivation:

- Compare tree-based ML methods to traditionally used pricing models
- Address the role of the loss function and its relation to the Within ±20% performance measure

Loss function:

- Squared Error loss: $L_{\text{SE}} = (y - \hat{y})^2$
- Squared Percentage Error loss: $L_{\text{SPE}}(y, \hat{y}) = \left(\frac{y-\hat{y}}{y}\right)^2$
- Hybrid loss: $L_{\text{hybrid}}(y, \hat{y}) = \frac{(y-\hat{y})^2}{y^{\alpha}}$

Paper 1

Motivation:

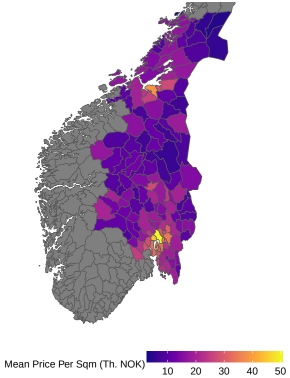
- Compare tree-based ML methods to traditionally used pricing models
- Address the role of the loss function and its relation to the Within ±20% performance measure

Loss function:

- Squared Error loss: $L_{\text{SE}} = (y - \hat{y})^2$
- Squared Percentage Error loss: $L_{\text{SPE}}(y, \hat{y}) = \left(\frac{y-\hat{y}}{y}\right)^2$
- Hybrid loss: $L_{\text{hybrid}}(y, \hat{y}) = \frac{(y-\hat{y})^2}{y^{\alpha}}$

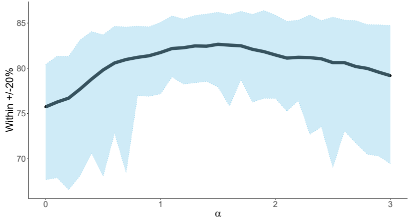
Data:

- Transactions from Oslo, Viken, Trøndelag, and Innlandet, 2013-2015 ($N = 126\,719$)
- Features: Location, estate type, size (m^2), number of bedrooms, neighborhood characteristics, *etc.*
- Arm's length transactions only



Results and Implications

	RMSPE	Within ±20%
Linear regression	32.7 (0.5)	60.2 (0.3)
NN regression	28.4 (0.9)	70.9 (0.2)
Random forest	18.8 (0.4)	87.1 (0.2)
XGBoost (SE loss)	16.2 (0.5)	89.4 (0.1)
XGBoost (SPE loss)	**15.4 (0.4)**	**90.0 (0.2)**
XGBoost (Combined)	**15.3 (0.4)**	**90.4 (0.1)**



Main findings:

- Tree-based models significantly outperform linear regression and NN regression
- SPE loss further improves XGBoost, but a weighted average of SE loss and SPE loss is even better
- The loss functions are optimized for different price segments
- Appendix: A hybrid loss function $L_{\text{hybrid}}(y, \hat{y}) = \frac{(y-\hat{y})^2}{y^\alpha}$ shows promising results with $\alpha \approx 1.5$.

Paper II:
Locally interpretable tree boosting: An application to house price prediction

Authors: Hjort, Anders; Scheel, Ida; Sommervoll, Dag Einar; Pensar, Johan

Published in: *Decision Support Systems 178, 2024*

Paper 2: Background on GAMs

- **Data**: $N = 14\,382$ transactions from 15 city districts in Oslo (2018)

Paper 2: Background on GAMs

- **Data**: $N = 14\,382$ transactions from 15 city districts in Oslo (2018)
- A gradient boosted tree (GBT) with tree stumps an be viewed as a generalized additive model (GAM),

$$\hat{f}_{\text{GAM}}(\boldsymbol{x}) = f_0 + f_1(x_1) + \ldots + f_d(x_d)$$

- No interactions makes model fully interpretable, but does not capture city district specific effects

Paper 2: Background on GAMs

- **Data**: $N = 14\,382$ transactions from 15 city districts in Oslo (2018)
- A gradient boosted tree (GBT) with tree stumps an be viewed as a generalized additive model (GAM),

$$\hat{f}_{\text{GAM}}(\boldsymbol{x}) = f_0 + f_1(x_1) + \ldots + f_d(x_d)$$

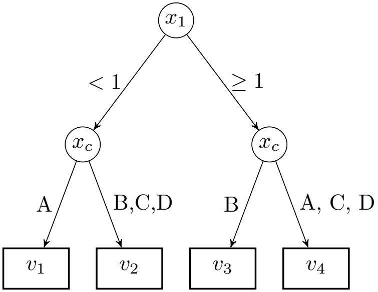
- No interactions makes model fully interpretable, but does not capture city district specific effects
- Training a local GAM per city district does not facilitate borrowing of strength between models

Paper 2: Locally Interpretable Tree Boosting

- Proposed solution:

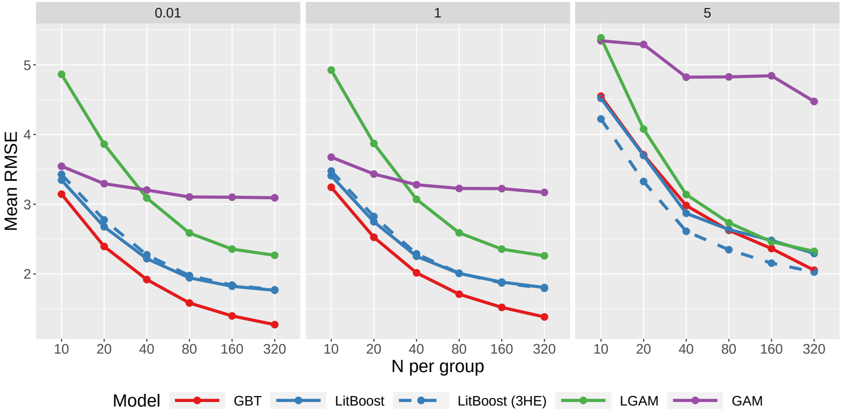
  $$\hat{f}_{\text{LitBoost}}(\boldsymbol{x}) = f_0 + f_1(x_1, x_c) + \ldots + f_d(x_d, x_c)$$

  for a categorical x_c (e.g., city district).
- The interpretability benefits of an ordinary GAM, yet almost similar predictive accuracy as a GBT



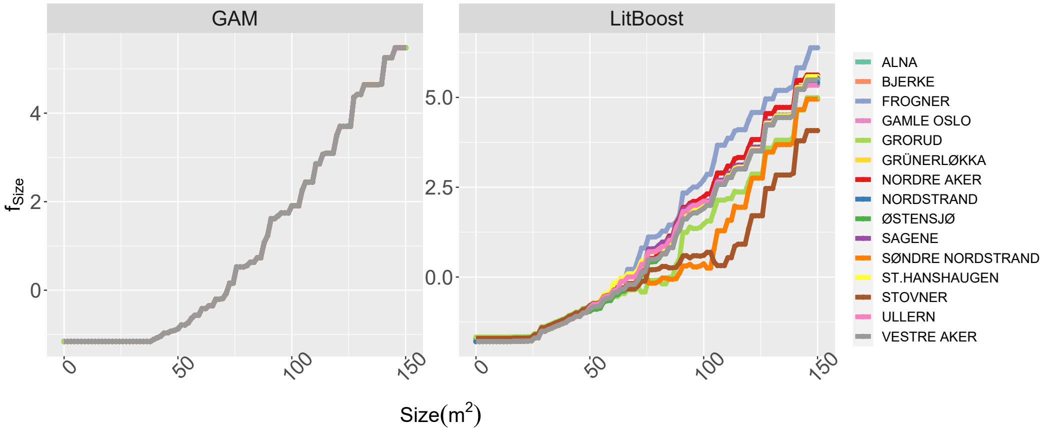
UNIVERSITY
OF OSLO

Accuracy is close to GBT



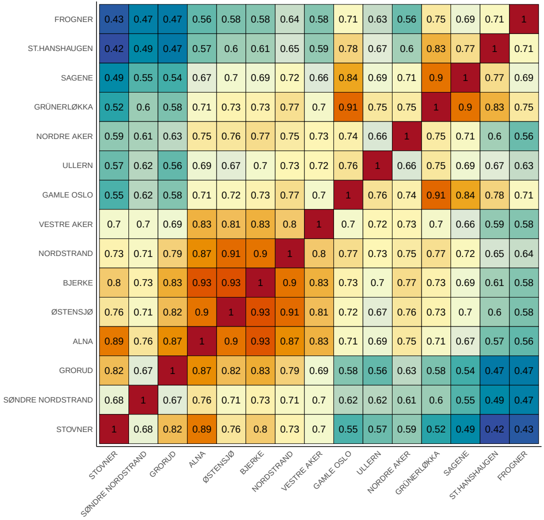
UNIVERSITY OF OSLO

Paper 2: The shape functions



UNIVERSITY
OF OSLO

Paper 2: Groupwise proximity measure



Paper III:
Uncertainty quantification in automated valuation models with spatially weighted conformal prediction

Authors: Hjort, Anders; Hermansen, Gudmund Horn; Pensar, Johan; Williams, Jonathan P.

Manuscript, arXiv preprint arXiv:2312.06531, 2024.

Conformal prediction

- From point prediction to prediction intervals (PI)
- Introduced in Vovk et al. 2005, have become popular in ML communities in recent years
- Distribution-free, minimal assumptions on underlying model, marginal finite-sample guarantees:

$$P\{y_{N+1} \in \mathrm{PI}(\boldsymbol{x}_{N+1})\} \geq 1 - \alpha$$

for any $\alpha \in (0, 1)$, given *exchangeable* data.

Inductive conformal prediction

What are like values of y_{N+1} given $\boldsymbol{x}_{N+1}$?

- Assume that we have a pre-trained model $\hat{f}$ and a calibration set

Inductive conformal prediction

What are like values of y_{N+1} given $\boldsymbol{x}_{N+1}$?

- Assume that we have a pre-trained model $\hat{f}$ and a calibration set
- Choose a non-conformity score, e.g., $s_i = |y_i - \hat{f}(\boldsymbol{x}_i)|$

Inductive conformal prediction

What are like values of y_{N+1} given $\boldsymbol{x}_{N+1}$?

- Assume that we have a pre-trained model $\hat{f}$ and a calibration set
- Choose a non-conformity score, e.g., $s_i = |y_i - \hat{f}(\boldsymbol{x}_i)|$
- Let $s_1, \ldots, s_N$ be the calibration scores

Inductive conformal prediction

What are like values of y_{N+1} given $\boldsymbol{x}_{N+1}$?

- Assume that we have a pre-trained model $\hat{f}$ and a calibration set
- Choose a non-conformity score, e.g., $s_i = |y_i - \hat{f}(\boldsymbol{x}_i)|$
- Let $s_1, \ldots, s_N$ be the calibration scores
- Let $\hat{q}_{1-\alpha}$ be the $(1-\alpha)$th percentile of calibration scores

Inductive conformal prediction

What are like values of y_{N+1} given $\boldsymbol{x}_{N+1}$?

- Assume that we have a pre-trained model $\hat{f}$ and a calibration set
- Choose a non-conformity score, e.g., $s_i = |y_i - \hat{f}(\boldsymbol{x}_i)|$
- Let $s_1, \ldots, s_N$ be the calibration scores
- Let $\hat{q}_{1-\alpha}$ be the $(1-\alpha)$th percentile of calibration scores
- Create prediction set for y_{N+1}:

$$C_{1-\alpha}(\boldsymbol{x}_{N+1}) = \left\{ y \in \mathbb{R} : \quad s_{N+1}(\boldsymbol{x}_{N+1}, y) \leq \hat{q}_{1-\alpha} \right\}$$

Paper 3: Motivation

- Uncertainty quantification in AVMs is essential to most users

Paper 3: Motivation

- Uncertainty quantification in AVMs is essential to most users
- **Goal**: Investigate methods to achieve a low coverage gap also in geographic subsets

Paper 3: Motivation

- Uncertainty quantification in AVMs is essential to most users
- **Goal**: Investigate methods to achieve a low coverage gap also in geographic subsets
  - Two possible solutions from the literature:

$$C_{1-\alpha}(\boldsymbol{x}_{N+1}) = \left\{y \in \mathbb{R} : \quad s_{N+1}(\boldsymbol{x}_{N+1}, y) \leq \hat{q}_{1-\alpha}\right\}$$

Choose a better non-conformity score, e.g., $s_i = |y_i - \hat{f}(\boldsymbol{x}_i)|/\hat{\sigma}(\boldsymbol{x}_i)$

Paper 3: Motivation

- Uncertainty quantification in AVMs is essential to end users
- **Goal**: Investigate methods to a low coverage gap in subsets of space
    - Two possible solutions from the literature:

$$C_{1-\alpha}(\boldsymbol{x}_{N+1}) = \left\{ y \in \mathbb{R} : \quad s_{N+1}(\boldsymbol{x}_{N+1}, y) \leq \hat{q}_{1-\alpha} \right\}$$

Choose a better non-conformity score, e.g., $s_i = |y_i - \hat{f}(\boldsymbol{x}_i)|/\hat{\sigma}(\boldsymbol{x}_i)$
or calibrate differently, e.g., spatially weighted CP or Mondrian CP

Paper 3: Motivation

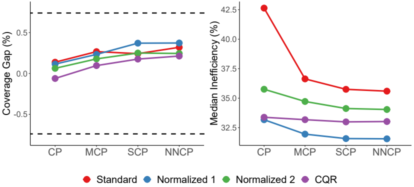
- Uncertainty quantification in AVMs is essential to end users
- **Goal**: Investigate methods to a low coverage gap in subsets of space
    - Two possible solutions from the literature:

$$C_{1-\alpha}(\boldsymbol{x}_{N+1}) = \left\{y \in \mathbb{R} : \quad s_{N+1}(\boldsymbol{x}_{N+1}, y) \leq \hat{q}_{1-\alpha}\right\}$$

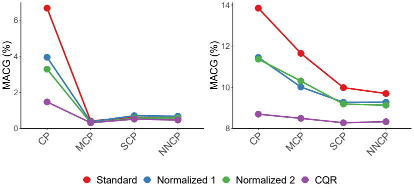
  Choose a better non-conformity score, e.g., $s_i = |y_i - \hat{f}(\boldsymbol{x}_i)|/\hat{\sigma}(\boldsymbol{x}_i)$
  or calibrate differently, e.g., spatially weighted CP or Mondrian CP

- **Data**: Transactions from Oslo 2016-2017 ($N = 26\,362$)
- **Setup**: Four non-conformity scores $\times$ four weighting methods

Results and implications



Marginal results: Coverage and interval size on average.

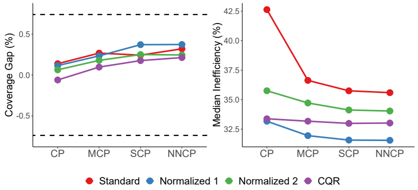


Conditional results: Mean Absolute Coverage Gap per city district.

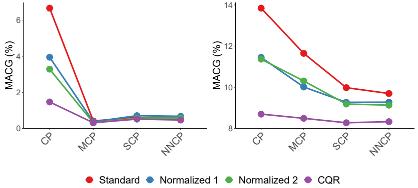
Results and implications

Main findings:

- CP is a promising technique for the AVM application
- Conformalized Quantile Regression (Romano et al. 2019) is overall the preferred choice for the application
- **Should we focus on designing a better non-conformity score or better weighting method?** Weighting is a "duct tape solution" that fixes poor non-conformity scores, but a good non-conformity score does not need weighting



Marginal results: Coverage and interval size on average.



Conditional results: Mean Absolute Coverage Gap per city district.

Paper IV:
Clustered conformal prediction for the housing market

Authors: Hjort, Anders; Williams, Jonathan P.; Pensar, Johan
Published in: *Proceedings of Machine Learning Research, 230, 366-386, 2024*.

Paper 4: Motivation

- Spatially weighted CP (Paper 3) fails in sparse areas

Paper 4: Motivation

- Spatially weighted CP (Paper 3) fails in sparse areas
- Therefore: Discard geography! Form clusters of municipalities with similar distributions of non-conformity scores (Ding et al. 2023)

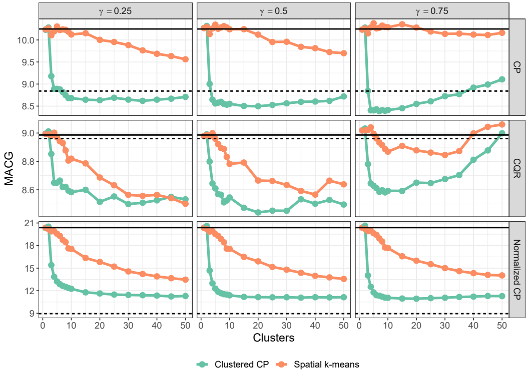
Paper 4: Motivation

- Spatially weighted CP (Paper 3) fails in sparse areas
- Therefore: Discard geography! Form clusters of municipalities with similar distributions of non-conformity scores (Ding et al. 2023)
- Create M clusters, calibrate cluster-wise with $\hat{q}^1_{1-\alpha}, \hat{q}^2_{1-\alpha}, \cdots, \hat{q}^M_{1-\alpha}$.
- **Housing market intuition**: If our model errs “similarly” in two municipalities, we merge them when calibrating the PIs

Paper 4: Setup

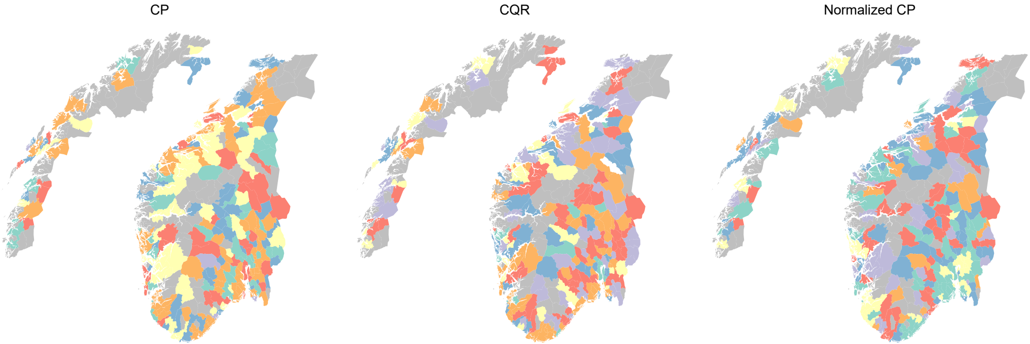
- **Data**: Entire Norway in 2015 ($N = 84\,975$)
- Only include municipalities with > 10 sales $\implies$ $K = 286$ municipalities
- Evaluate Mean Absolute Coverage Gap (MACG) across municipalities for varying number of clusters M

Paper 4: Results



Straight line: One cluster (global calibration). Dotted line: 286 clusters (Mondrian calibration).

Paper 4: Results



Main contributions

- Improve the accuracy, interpretability and uncertainty quantification in AVMs
- The Locally Interpretable Tree Boosting (LitBoost) model
- Contribute the literature on applied conformal prediction with spatial data

References I

- Ding, Tiffany, Anastasios Nikolas Angelopoulos, Stephen Bates, Michael Jordan and Ryan Tibshirani (2023). “Class-Conditional Conformal Prediction with Many Classes”. In: *Advances in Neural Information Processing Systems*. Vol. 36.
- Romano, Yaniv, Evan Patterson and Emmanuel Candes (2019). “Conformalized Quantile Regression”. In: *Advances in Neural Information Processing Systems*. Vol. 32.
- Vovk, Vladimir, Alex Gammerman and Glenn Shafer (2005). *Algorithmic Learning in a Random World*. Springer-Verlag.