UiO : Department of Mathematics
University of Oslo

Clustered Conformal Prediction for the Housing Market

The 13th Symposium on Conformal and Probabilistic Prediction with Applications, 2024

Anders Hjort

September 11, 2024

Collaborators

Jonathan P. Williams
North Carolina State University

Johan Pensar
University of Oslo

Eiendomsverdi
Norwegian fintech company

Motivation

- Estimating the current value of a home is essential for homeowners, banks, real estate agents, insurance companies, investors, government, etc.
- Increasing use of automated valuation models (AVMs) instead of manual appraisal
- Extremely noisy prediction problem $\implies$ need to quantify prediction uncertainty
- State-of-the-art: Tree-based models combined with temporal and spatial smoothing

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

ALEX ENGLER, SYLVIA BROWN / OCT 9, 2023

Shutterstock

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?

Well, let us introduce you to Automated Valuation Models, or AVMs, invented a century ago.

Article by researchers at at Brookings Institution and Georgetown University, published in *Tech Policy* on 9th of October 2023.

Wall Street Journal article from 17th of November 2021.

Quantifying Uncertainty in AVMs

- CP applied to the housing market previously:
  - Bellotti 2017: Adjust for temporal drift (London, UK)
  - Lim and Bellotti 2021: Design novel non-conformity scores for AVMs (Ames, US)
  - Hjort et al. 2024, preprint: Spatially-weighted CP (Oslo, Norway)
  - Bastos and Paquette 2024, preprint: Conformalized QR outperforms QR (San Francisco, US)

Quantifying Uncertainty in AVMs

- CP applied to the housing market previously:
  - Bellotti 2017: Adjust for temporal drift (London, UK)
  - Lim and Bellotti 2021: Design novel non-conformity scores for AVMs (Ames, US)
  - Hjort et al. 2024, preprint: Spatially-weighted CP (Oslo, Norway)
  - Bastos and Paquette 2024, preprint: Conformalized QR outperforms QR (San Francisco, US)
- Our target: Approximately conditional coverage across municipalities
- We study $N = 84\,975$ transactions from $K = 286$ different municipalities in Norway

Conformal prediction

Indcutive conformal prediction approach:

- Split data set at random into training, calibration, test set
- Train a regression model $\hat{f} : \mathcal{X} \mapsto \mathcal{Y}$ on training set
- Calculate scores $s_i = \Psi(X_i, Y_i; \hat{f})$ on calibration set
- On test set:

$$C_{1-\alpha}(X_{N+1}) = \{y \in \mathcal{Y} : \quad \Psi(X_{N+1}, y; \hat{f}) \leq \hat{q}_{1-\alpha}\}$$

where $\hat{q}_{1-\alpha}$ is an empirical quantile of $s_1, ..., s_{N_{\text{calib}}}$.

Challenges

- Naïve CP: Calculate $\hat{q}_{1-\alpha}$ once
  ✓ Marginal validity guarantees
  ✗ In practice high coverage gap in some regions

Challenges

- Naïve CP: Calculate $\hat{q}_{1-\alpha}$ once
  ✓ Marginal validity guarantees
  ✗ In practice high coverage gap in some regions
- Mondrian CP: Separate $\hat{q}_{1-\alpha}$ for each region
  ✓ Theoretical guarantees on coverage per region
  ✗ Struggle in classes with few observations

Challenges

- Naïve CP: Calculate $\hat{q}_{1-\alpha}$ once
  ✓ Marginal validity guarantees
  ✗ In practice high coverage gap in some regions
- Mondrian CP: Separate $\hat{q}_{1-\alpha}$ for each region
  ✓ Theoretical guarantees on coverage per region
  ✗ Struggle in classes with few observations
- Spatially weighted CP: Separate $\hat{q}_{1-\alpha}$ for each observation
  ✓ Some theoretical guarantees (Mao et al. 2023), and empirical success (Hjort et al. 2024, preprint)
  ✗ Fails if data is sparse

Challenges

- Naïve CP: Calculate $\hat{q}_{1-\alpha}$ once
  ✓ Marginal validity guarantees
  ✗ In practice high coverage gap in some regions
- Mondrian CP: Separate $\hat{q}_{1-\alpha}$ for each region
  ✓ Theoretical guarantees on coverage per region
  ✗ Struggle in classes with few observations
- Spatially weighted CP: Separate $\hat{q}_{1-\alpha}$ for each observation
  ✓ Some theoretical guarantees (Mao et al. 2023), and empirical success (Hjort et al. 2024, preprint)
  ✗ Fails if data is sparse
- Clustered CP: Cluster together similar regions, calculate $\hat{q}_{1-\alpha}$ per cluster
  ✓ Works well in classification (Ding et al. 2023)
  ✗ Small bias in coverage guarantees if clustering is poor

Clustered CP

Algorithm:

- Use fraction $\gamma \in (0, 1)$ of calibration data for clustering
- Cluster the ECDFs $\hat{F}_1, ..., \hat{F}_K$ into $M < K$ clusters, minimizing within-cluster variance
- Let $\hat{q}_{1-\alpha}^{(m)}$ be the $(1-\alpha)$th quantile of scores in cluster m
- Calibrate cluster-wise: for every observation in any class k in cluster m we use $\hat{q}_{1-\alpha}^{(m)}$ to create the prediction interval

Clustered CP

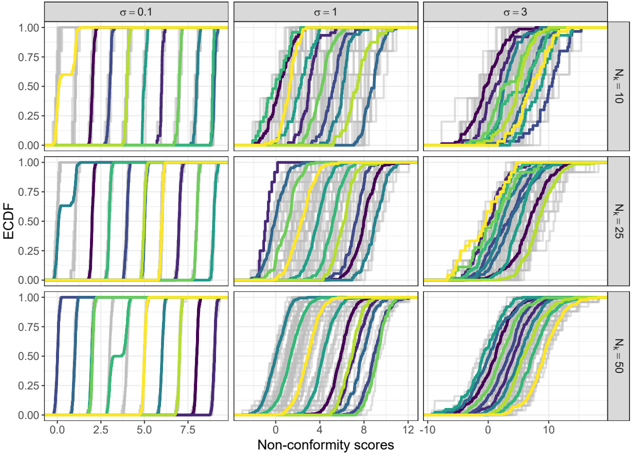
Algorithm:

- Use fraction $\gamma \in (0,1)$ of calibration data for clustering
- Cluster the ECDFs $\hat{F}_1, ..., \hat{F}_K$ into $M < K$ clusters, minimizing within-cluster variance
- Let $\hat{q}_{1-\alpha}^{(m)}$ be the $(1-\alpha)$th quantile of scores in cluster m
- Calibrate cluster-wise: for every observation in any class k in cluster m we use $\hat{q}_{1-\alpha}^{(m)}$ to create the prediction interval

Theoretical properties: Let ε_m be the maximum Kolmogorov-Smirnov distance between two classes in cluster m. Then,

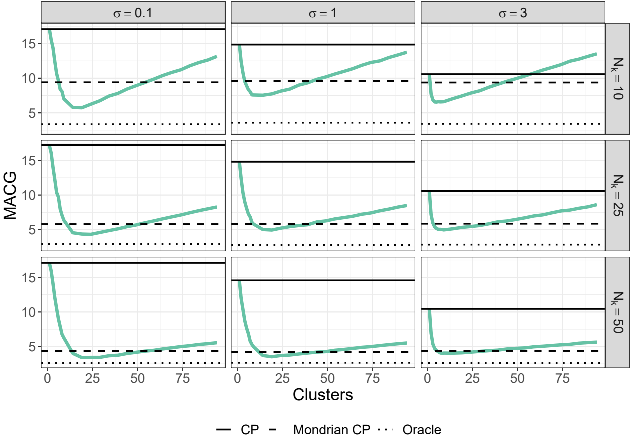
$$P\Big(Y_{N+1} \in C(X_{N+1}) | \text{class } \ k\Big) \geq 1 - \alpha - \varepsilon_m, \quad \forall k \in m.$$

Clustered CP: Synthetic data



ECDFs of $K = 100$ classes (in grey) and $M = 10$ clusters (in colors). Non-conformity scores in class k is drawn from from $\mathcal{N}(\mu_k, \sigma^2)$, with $\mu_k \sim U(0, 1, ..., 10)$.

Clustered CP: Synthetic data



Mean Absolute Coverage Gap (MACG) as a function of the number of clusters.

The data set

We study $N = 84\,975$ from the Norwegian housing market in 2015. Transactions come from $K = 286$ different municipalities; $N_k < 100$ for more than 167 municipalities and $N_k > 1\,000$ for 16 municipalities.

Variable	Unit	Mean	St. Dev.	Min	Max	Type
Sale Price	NOK (mill.)	3.07	1.72	0.02	28.7	Numerical
Size	m^2	100	54	0	819	Numerical
Gross Size	m^2	112.42	67.48	0	1131	Numerical
Longitude	degrees	9.82	2.90	4.79	30.47	Numerical
Latitude	degrees	60.71	2.37	57.99	70.72	Numerical
Altitude	m	101.69	136.49	0	1151	Numerical
Bedrooms	-	2.56	1.20	0	15	Numerical
Municipality	-	-	-	-	-	Categorical

Experimental setup

- Random split into training (25%), calibration (50%) and test (25%)
- Three non-conformity scores:

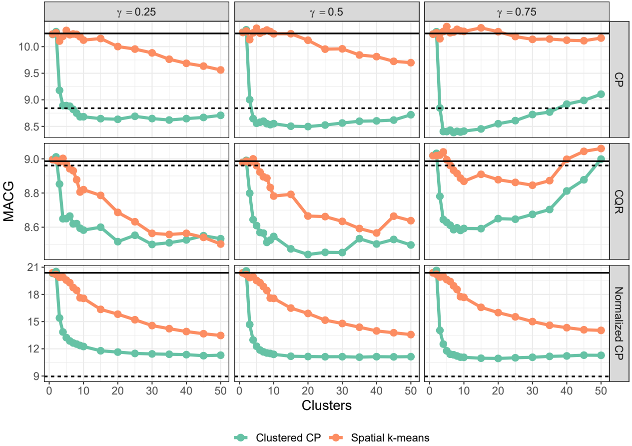
$$\Psi_{CP}(X_i, Y_i) = |Y_i - \hat{f}(X_i)| \quad \text{(CP)}$$

$$\Psi_{Norm.CP}(X_i, Y_i) = |Y_i - \hat{f}(X_i)|/\hat{f}(X_i) \quad \text{(Normalized CP)}$$

$$\Psi_{CQR}(X_i, Y_i) = \max\{\hat{Q}_{\alpha/2}(X_i) - Y_i, Y_i - \hat{Q}_{1-\alpha/2}(X_i)\} \quad \text{(CQR)}$$

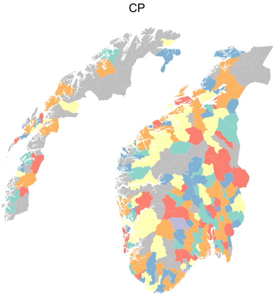
- We use a random forest to train $\hat{f}$, and quantile regression forest (Meinshausen 2006) for CQR
- Clustering:
  - Experiment with cluster fractions $\gamma \in (0.25, 0.5, 0.75)$.
  - Discretize each ECDF, i.e., $\hat{F}_k \approx [q^k_{10}, q^k_{20}, \ldots, q^k_{90}]$. Solve by M-means clustering in $\mathbb{R}^9$.
  - If $N_k < 10$: Assign to NULL cluster, calibrate globally.

Results



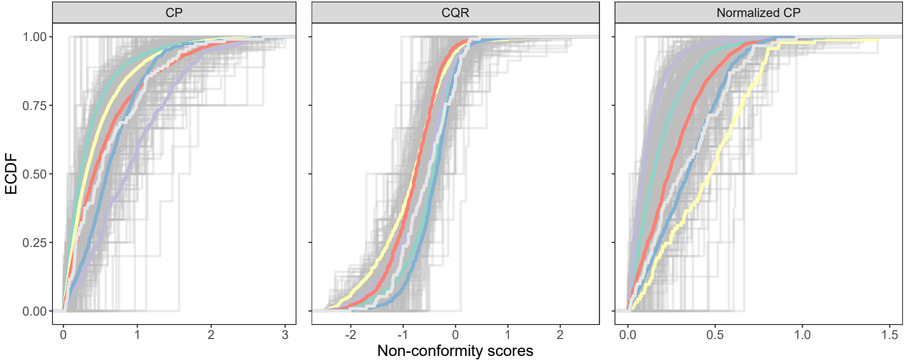
Straight line: Global calibration with $\gamma = 0$ (CP). Dotted: Mondrian CP with $\gamma = 0$. Note that the range of MACG is different for the different non-conformity scores.

Results



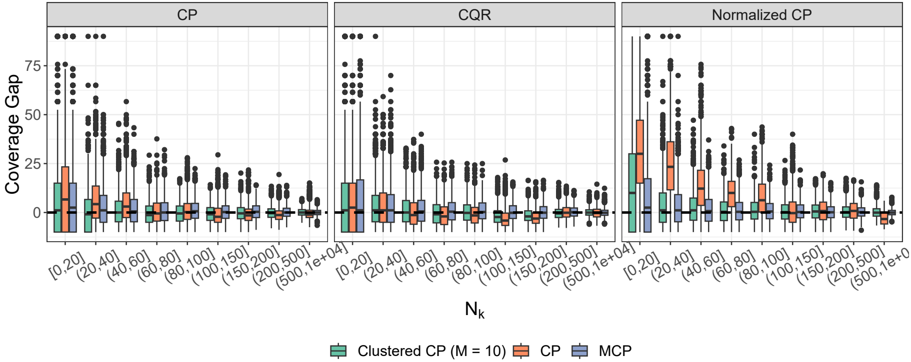
An example of the identified clusters with the Clustered CP methodology for $M = 6$ clusters. The grey municipalities either have no observations or are part of the NULL cluster.

Results



The ECDF of the identified clusters with Clustered CP for $M = 6$, overlaying the individual ECDFs for each municipality.

Results



Coverage gap for different bins of N_k for MCP, CP, and Clustered CP with $M = 10$. The results are for confidence level $\alpha = 0.1$ with a fraction $\gamma = 0.5$ set aside for clustering in Clustered CP.

Discussion

- Clustered CP is a pragmatic version of Mondrian CP where similar classes are pooled together
- Induces a small coverage gap ε_m in theory which is reduced if the clustering is good
- Clustering based on ECDFs outperforms clustering based on spatial distance
- Open questions:
  - How to decide the optimal number of clusters a priori?
  - How to handle the imbalanced classes?
  - Adjusting the CP intervals for temporal drift in the housing market

References I

- Bastos, JA and J Paquette (2024). *On the uncertainty of real estate price predictions*. Preprint downloaded from https://rem.rc.iseg.ulisboa.pt/wps/pdf/REM_WP_0314_2024.pdf.
- Bellotti, Anthony (2017). 'Reliable region predictions for automated valuation models'. In: *Annals of Mathematics and Artificial Intelligence* 81.1, pp. 71–84.
- Ding, Tiffany, Anastasios Nikolas Angelopoulos, Stephen Bates, Michael Jordan and Ryan Tibshirani (2023). 'Class-Conditional Conformal Prediction with Many Classes'. In: *Thirty-seventh Conference on Neural Information Processing Systems*.
- Hjort, Anders, Gudmund Horn Hermansen, Johan Pensar and Jonathan P. Williams (2024). *Uncertainty quantification in automated valuation models with locally weighted conformal prediction*. arXiv: 2312.06531.

References II

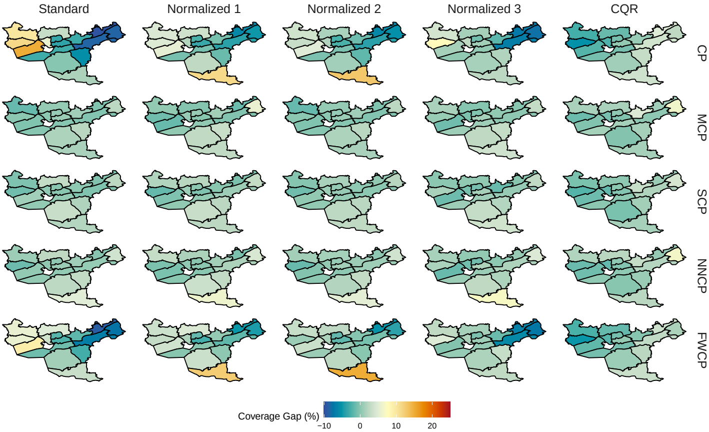
- Lim, Zhe and Anthony Bellotti (2021). 'Normalized nonconformity measures for automated valuation models'. In: *Expert Systems with Applications* 180, pp. 115–165.
- Mao, Huiying, Ryan Martin and Brian J. Reich (2023). 'Valid Model-Free Spatial Prediction'. In: *Journal of the American Statistical Association* 119, pp. 1–11.
- Meinshausen, Nicolai (2006). 'Quantile Regression Forests'. In: *Journal of Machine Learning Research* 7.35, pp. 983–999.

Appendix: Synthetic data, details

Draw data from $K = 100$ different classes. Each class is drawn from a normal $\mathcal{N}(\mu_k, \sigma^2)$. Importantly: Some of the groups are drawn with similar μ_k!

$$
\begin{aligned}
G &\sim U(1, ..., K) \\
\mu_k &\sim U(1, 2, ..., \sqrt{K}) \\
S|G = k &\sim \mathcal{N}(\mu_k, \sigma^2).
\end{aligned}
$$

Appendix: Results from Hjort et al. 2024



The map shows the performance for different non-conformity measures (horizontally) and weighting methods (vertically) on a data set of $N = 26\,362$ observations from Oslo (2016-2017).

UiO : Department of Mathematics
University of Oslo

Anders Hjort

Clustered Conformal Prediction for the Housing Market

The 13th Symposium on Conformal and Probabilistic Prediction with Applications, 2024