Transfer Learning for Tourism Flash Estimates with Mobile Network Operator Data

Cristina Faricelli, Alessandro Piovani and Tiziana Tuoto - ISTAT

Overview

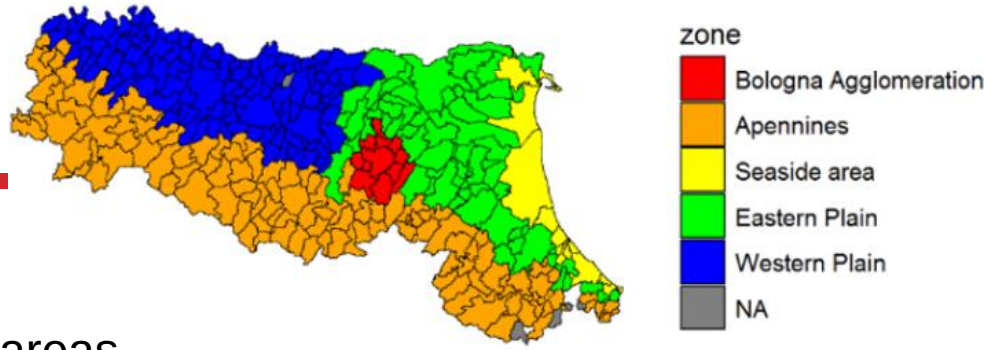
IDEA

use MNO data as a timely source to support or **anticipate official statistics** when traditional sources are delayed or incomplete.

MNO USAGE

In this work MNO data are used as **feature** values to enhance **model predictions** to anticipate official data.

Case study



Emilia-Romagna region

- Heterogeneous territory: plain, hills/mountain, seaside and urban areas
- **330 municipalities**, Bologna capital city

Survey data about Nights spent by tourists:

- **Monthly time series** at **municipality level** from **January 2021** to **October 2023**
- Available with **2 months delay** in general

MNO data for Nights spent by tourists:

- **Monthly time series** at **municipality level** from **August 2022** to **October 2023**
- Available with **delay ≤ 1 month for every municipality**

Overnight stays are inferred from device-presence patterns using data from a single operator, processed with **multi-MNO algorithms** to approximate population-wide behavior (TSS Multi-MNO project. 2025).

Auxiliary data at municipality level:

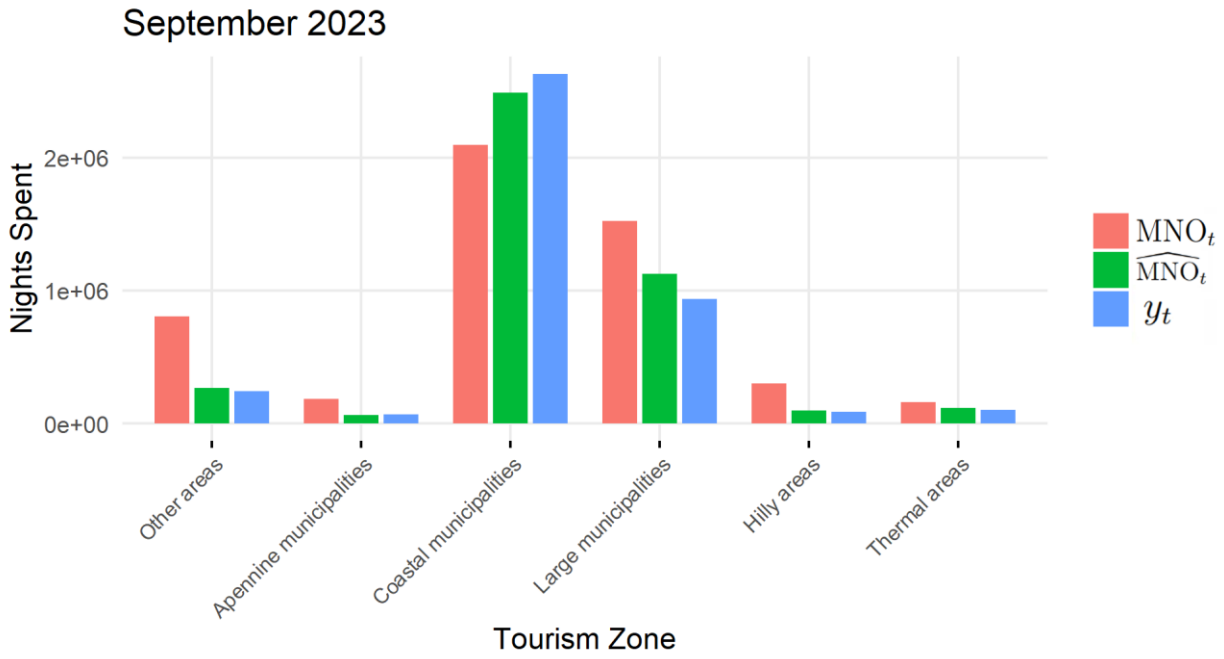
Capacity, Tourism zone, calendar features

Istat

MNO ADJUSTMENT

- The **MNO data** processed with the **MULTI-MNO** project algorithm showed the presence of a **bias**.

- In most municipalities, there is a **seasonal** behavior and therefore the bias is corrected as follow:



$$\widehat{\mathrm{MNO}}_t = \mathrm{MNO}_t \cdot \frac{y_{t-12}}{\mathrm{MNO}_{t-12}}$$

Where

$\widehat{\mathrm{MNO}}_t$	denotes the bias-corrected estimate derived from MNO data in month *t*
MNO_t	denotes the estimate derived from MNO data in month *t*
MNO_{t-12}	denotes the estimate derived from Mobile Network Operator (MNO) data in month *t-12*
y_{t-12}	denotes the number of nights spent according to the official survey in month *t-12*

About 10% of the municipalities don't show a seasonal behavior and the bias is corrected following the method proposed by the Spanish Institute in the deliverable 3.2 of the MNO-MINDS project (see references).

Istat

Approaches and Model

Approaches used to integrate MNO, survey and auxiliary data:

- **Traditional Modelling (TM)** - baseline model for comparison
- **Augmented Learning (AL)**
- **Quasi-Transfer Learning (QTL)**

Common Elements

- All approaches rely on a **Random Forest Regressor**.

Key Differences

- **Feature selection** - different sets of predictors
- **Training strategy**: **TM:** trained on the complete historical dataset
  AL & QTL: trained **only on the data of the target month**

Istat

Approaches description

Augmented Learning (AL)

- **Idea:** Build the training set by replacing missing target values with surrogate targets.

  In our case, the substitution mechanism is: **missing *y* replaced with y_{t-12}** (previous year values)
- **Purpose:** Performing nowcasting using proxy data.

Transfer Learning (TL)

- **Idea:** Build a model in the past (source period *b*) when survey labels are available.
  - Use that past model to generate predictions for the **current month.**
  - Feed those predictions as an extra feature into the current predictive model.
- **Purpose:** Transfers significant and stable information from history to the present using *past-model predictions used as an additional feature* (μ_b).

(See *Zhang and Haug,2024* for other substitution mechanisms, transfer schemes and relations)

Istat

Sets of predictors

A Random Forest regressor is built on the following set of features:

TM $\left\{ \begin{matrix} \mathrm{MNO}_t & y_{t-1} & y_{t-12} \\ Capacity, Tourism\ zone, & Calendar_features & \end{matrix} \right\}$

AL $\{\ y_{t-1} \quad \widehat{\mathrm{MNO}_t}\ \}$

With y_{t-12} used as a proxy (surrogate value) for y_t

TL $\{\ y_{t-1}\ \widehat{\mathrm{MNO}_t}\ \mu_b\}$

With y_{t-12} used as a proxy (surrogate value) for y_t. The predictor μ_b comes form the transfer learning method as previous described.

Istat

Results (1)

Models results for September 2023:

	AL	QTL	TM
Observed total (y)		4 111 172	
Predicted total ($\hat{y}$)	3 969 531	3 974 471	3 891 314
Absolute Error (AE)	481 771	383 500	439 893
Mean Absolute Error (MAE)	1 667	1 332	1 388
Mean Absolute Percentage Error (MAPE)	38.4%	37.0%	61.7%
Weighted Absolute Percentage Error (WAPE)	11.7%	9.3%	10.8%

Mean Absolute Error: $MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, i = municipality$

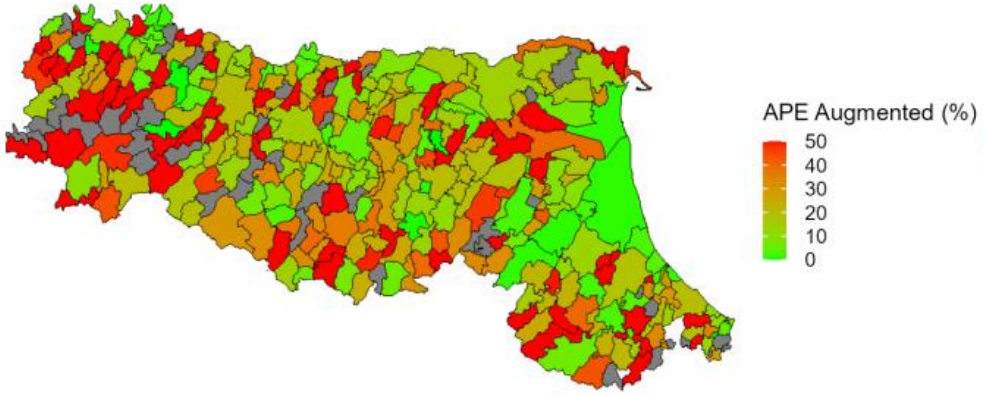
Mean Absolute Percentage Error: $MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \times 100, i = municipality$

Weigthed Absolute Percentage Error: $WAPE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n} y_i} \times 100, i = municipality$

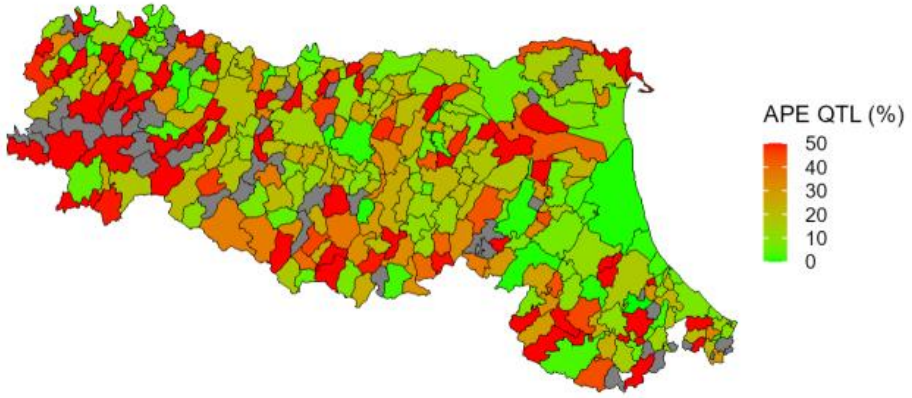
Istat

Results (2)

Absolute Percentage Error – Augmented Model

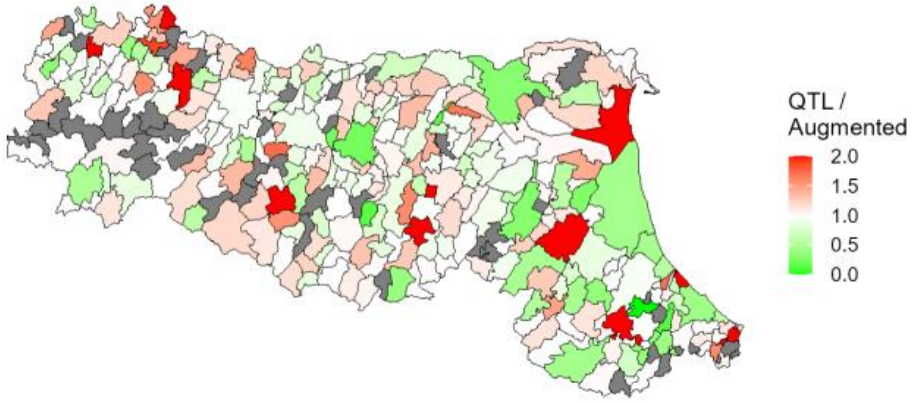


Absolute Percentage Error – QTL Model



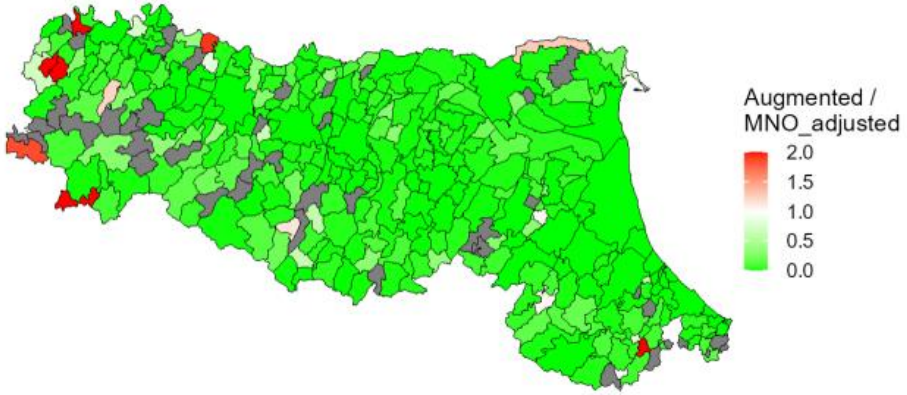
APE ratio – QTL vs Augmented

Green = QTL better



APE ratio – Augmented vs MNO_adjusted

Green = Augmented better



Istat

Uncertainty Estimation

- To quantify uncertainty, we used **conformal prediction intervals.** These intervals rely **only** on the **model residuals**, without assuming normality or any specific error distribution.

- Internal validation procedure: for each municipality, the model was refitted excluding that unit, and the resulting prediction error was used to approximate out-of-sample residual variability.

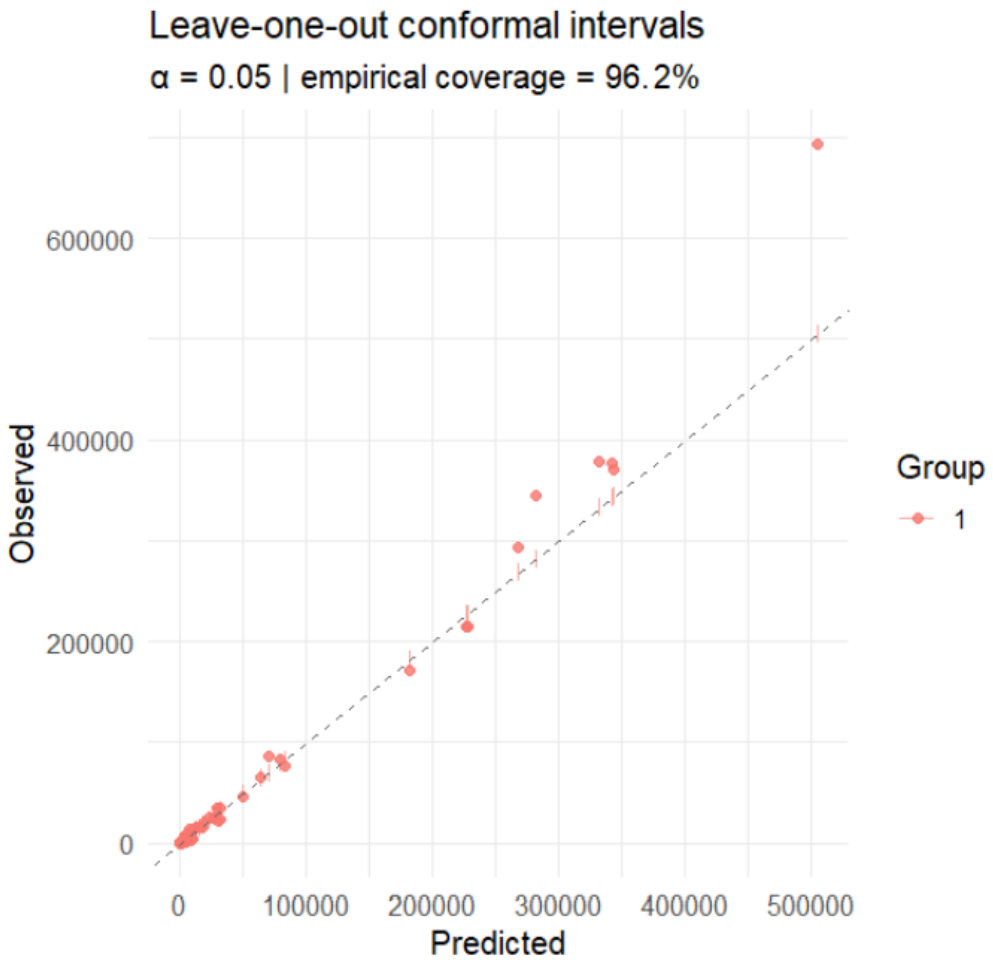
- We applied a clustering step based on the predicted values to obtain adaptive intervals:
  - wider for highly touristic municipalities,
  - narrower for small ones.

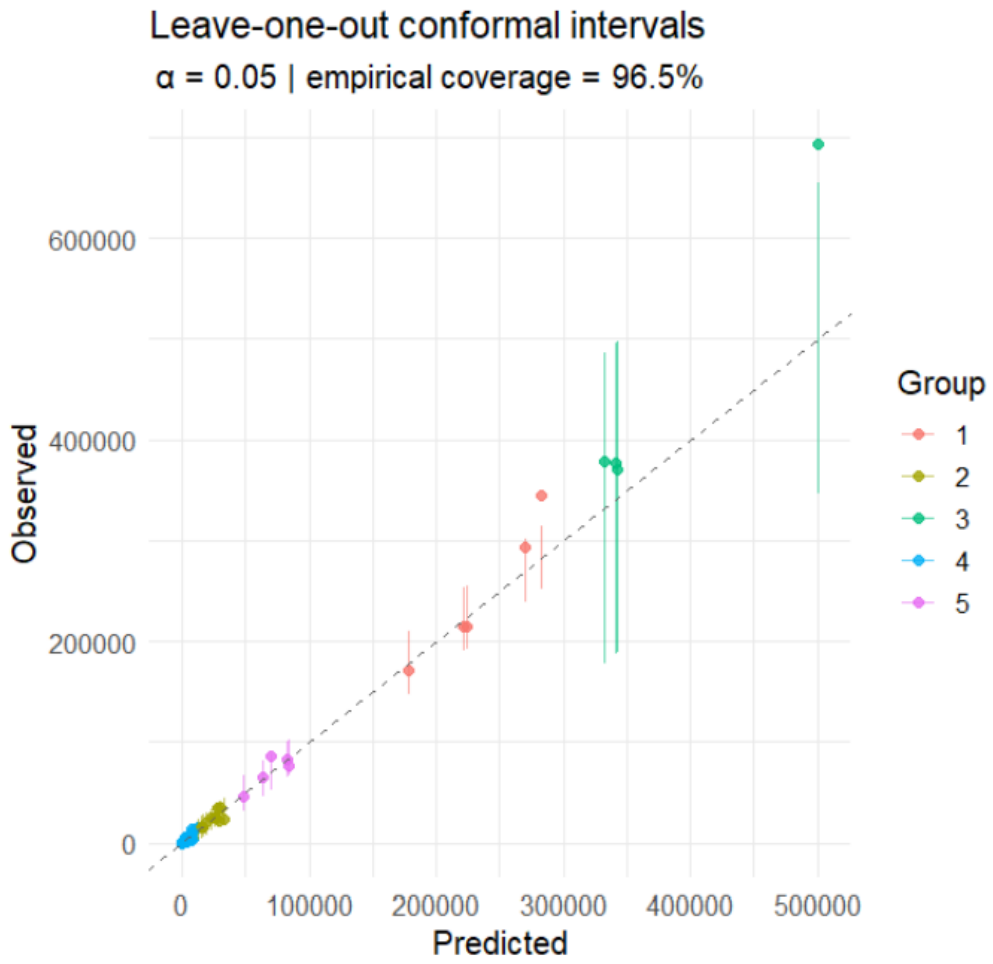
- The empirical coverage obtained with this internal validation approach was about 97%, which is consistent with the nominal 95% level.

Istat

Uncertainty Estimation

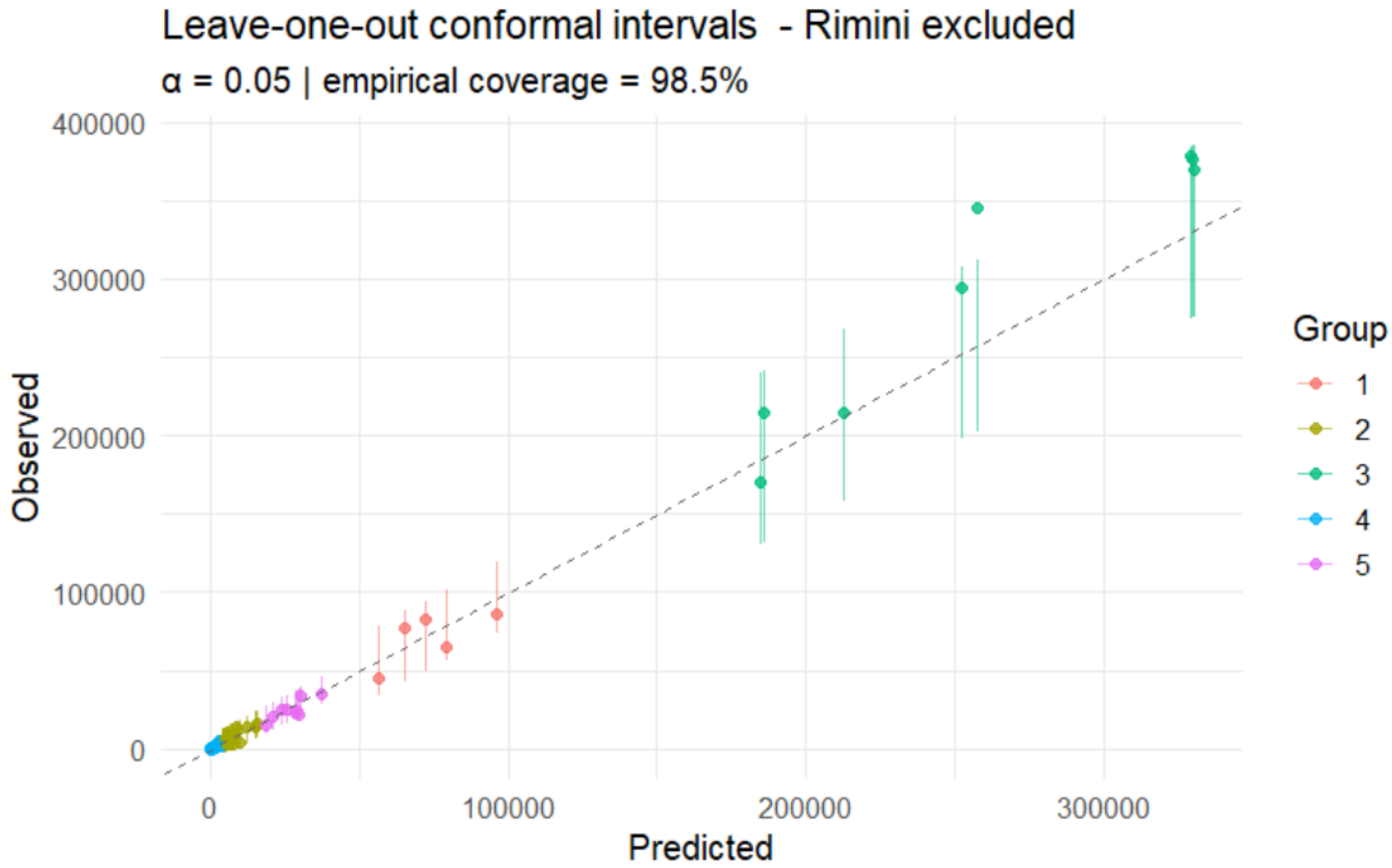
Conformal intervals for AL results





Istat

Uncertainty Estimation



Concluding Remarks

- Integrating **MNO data** with survey data proves **effective** in producing accurate estimates.
- Case studies show the **high informational value of MNO data**, especially when combined with traditional sources via **Augmented** and **Quasi-Transfer Learning**.
- Particularly **promising results for municipalities with low touristic intensity**, where longer historical MNO series can serve as valuable features even in more complex time-series models.
- **Conformal prediction intervals** using the conditional leave-one-out approach offer reliable uncertainty quantification, achieving **empirical coverage close to the nominal level**.

Istat

Future Works

- Comparison with traditional models built with longer time series (**SARIMA**)
- Considering **different approaches** for **different volume of tourism**
- Repeat the study in **other regions**
- Investigate **other uncertainty estimation methods** (e.g. retrospective validation Zhang et al, 2024)

References

ESSnet MNO-MINDS. 2025. "Report on methodologies for combining MNO and non-MNO data sources". *Report*. ESSnet MNO-MINDS. https://cros.ec.europa.eu/book-page/mno-minds-wp3.

Guan, L. 2022. "Localized conformal prediction: a generalized inference framework for conformal prediction". *Biometrika*, Volume 110, N. 1: 33–50. ISSN: 1464-3510. DOI: https://doi.org/10.1093/biomet/asac04010.1093/biomet/asac040. https://doi.org/10.1093/biomet/asac040.

TSS Multi-MNO project. 2025. "Final version of technical documentation for scenarios, requirements, use cases and methods, and high-level architecture". *Report*. Development, implementation and demonstration of a reference processing pipeline for the future production of official statistics based on multiple Mobile Network Operator (MNO) data. TSS Multi-MNO project. https://cros.ec.europa.eu/book-page/methodology-framework-high-level-architecture-requirements-use-cases-and-methods.

Vovk, V., A. Gammerman, and G. Shafer. 2005. *Algorithmic Learning in a Random World*. New York: Springer.

Zhang, L. and J. K. Haug. 2024. "Turnover flash estimation by purposive sampling and debit card transactions". *Journal of Official Statistics*, Volume 41, N. 1: 382–403.

Istat

THANK YOU!

Istat