Revealing the dynamic relations between traffic and crowding using big data from mobile phone network

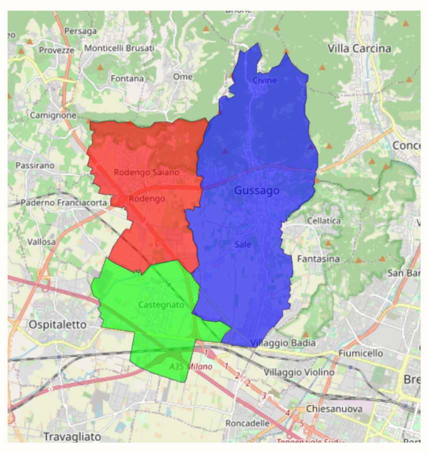
Selene Perazzini[1] Rodolfo Metulini[2] Maurizio Carpita[1]

[1]DMS Statlab, Department of Economics and Management, University of Brescia, Contrada Santa Chiara, 50, Brescia

[2]Department of Economics, University of Bergamo, Via Caniana, 2, Bergamo

UNIVERSITY OF BRESCIA
DMS StatLab
Data Methods and Systems Statistical Laboratory
DEPARTMENT OF ECONOMICS AND MANAGEMENT

Framework & goal



- The potential of **mobile phone data** in **traffic monitoring** is explored by analyzing their ability in estimating flows of people between small areas.
- We estimate traffic flows between **pairs of three "Aree di Censimento" (ACEs)** in the Province of Brescia: Rodengo Saiano (red), Castegnato (green), and Gussago (blue).
- A Vector AutoRegressive model with eXogenous variables (VARX) and complex seasonality captured by Dynamic Harmonic Regression (DHR) is used to model **traffic flows**.
- To improve the estimates, **crowding** and **traffic intensity** indicators defined on mobile phone data have been included in the model as exogenous regressors.

Data

Three types of mobile phone data about users subscribed to TIM have been used.

1. Traffic flows (y_{ijt})

Origin-Destination (OD) data are used to represent the flows of people y_{ijt}. Our OD database reports the number of SIM cards moving from ACE i to ACE j during the hour t between September 2020 and August 2021. Unusual traffic flows characterize holidays and have been replaced with the corresponding values observed seven days before (the previous same day of the week) [1].

2. Crowding indicator (CRO_{id})

CRO_{id} represents the average number of individuals in ACE i during the d-th hour of the day of the week (e.g., Monday at hour 00-01 AM). It has been defined using ***Mobile Phone Density*** (MPD) data for November 2020, which report the average number of mobile phone SIM cards in a 150×150 m^2 cell of a pixel grid during a 15-minute interval. The grid's cells have been assigned to the corresponding ACEs and the number of users in an ACE has been computed by aggregation. Then, values have been averaged among hours and days of the week.

3. Traffic intensity indicator (TRA_{id})

TRA_{id} has been defined using ***Minimization of Drive Test*** (MDT) data. The MDT database collects the number of signals (e.g., phone calls, text messages, ...) transmitted over the 3G/4G mobile network from/to terminal devices with GPS enabled in 15-minute intervals in a grid of pixels measuring 10 meters per side during 5 days of November 2021 (Wed 10th, Fri 19th, Sat 20th, Sun 21st, and Mon 22nd). The database has been restricted to pixels corresponding to streets using a polygon-based street map. For each ACE i and time interval t', we counted the number of street pixels from which at least one signal originated during a 15-minute interval.

Acknowledgments

This contribution has been developed for the European Union (EU) and Italian Ministry for Universities and Research (MUR), National Recovery and Resilience Plan (NRRP), within the project "Sustainable Mobility Center (MOST)" 2022-2026, CUP D83C22000690001, Spoke N 7 "CCAM, Connected networks and Smart Infrastructures".

References

[1] Metulini, R., and Carpita, M.: Modeling and forecasting traffic flows with mobile phone big data in flooding risk areas to support a data-driven decision making. Annals of Operations Research. 1–26, online first (2023)

[2] Perazzini, S., Metulini, R., Carpita, M. Statistical indicators based on mobile phone and street maps data for risk management in small urban areas. Submitted to journal.

[3] Perazzini, S., Metulini, R., Carpita, M. Integration of flows and signals data from mobile phone network for statistical analyses of traffic in a flooding risk area. Submitted to journal.

Model

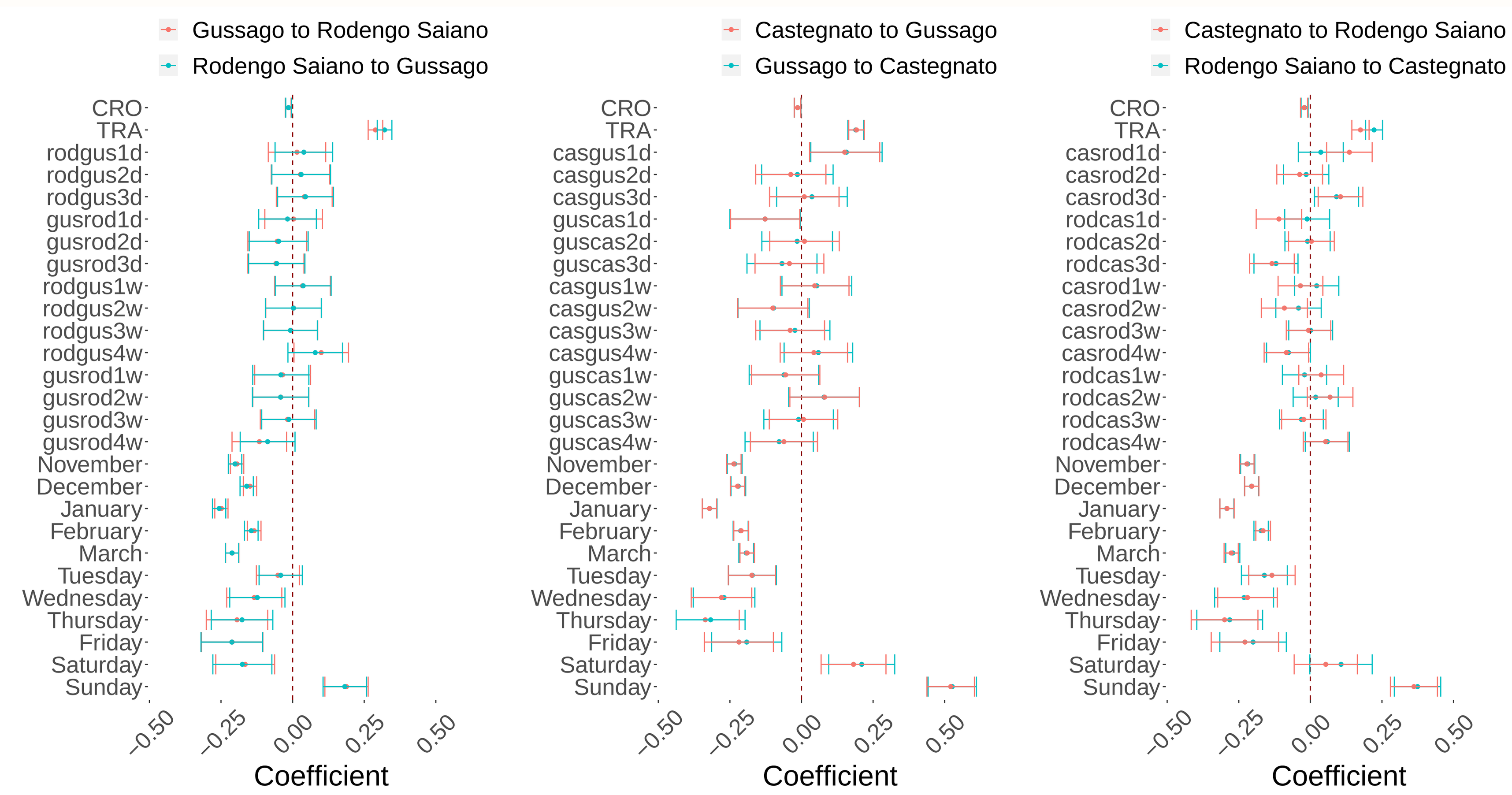
To model traffic flows between pairs of ACEs, we refer to the VARX model defined by [1]:

$$\mathbf{y}_t = \boldsymbol{\nu} + \sum_h \mathbf{A}_h \mathbf{y}_{t-h} + \mathbf{B}\mathbf{x}_t + \mathbf{C}\mathbf{z}_t + \boldsymbol{\epsilon}_t$$

- ν is a 2×1 vector of constants;
- $\mathbf{y}_t$ is a 2×1 vector containing the flows between ACEs i and j, namely y_{ijt} and y_{jit};
- ϵ_t is a 2×1 vector of the error terms at time t;
- h is a set of lag terms. $h = (24, 48, 72, 168, 336, 504, 672)$.
- $\mathbf{x}_t$ is a vector of exogenous variables containing $\overline{CRO}_{ijd}$, $\overline{TRA}_{ijd}$, 6 daily dummies, and 5 monthly dummies.
- $\overline{CRO}_{ijd} = (CRO_{id} + CRO_{jd})/2$ and $\overline{TRA}_{ijd} = (TRA_{id} + TRA_{jd})/2$
- $\mathbf{z}_t$ is a $l \times 1$ vector of exogenous variables capturing seasonality.
- $\mathbf{A}_h$, $\mathbf{B}$, and $\mathbf{C}$ are matrices of coefficients to be estimated;
- The elements of the vector $\mathbf{C}\mathbf{z}_t$ are modeled using a DHR through which we account for both daily and weekly seasonality.

Results

Figure 1: Estimated coefficients (points) and associated confidence intervals (lines). Lagged variables are indicated by the three initial letters of the ACEs of origin and destination and two digits representing the lag (1d corresponds to $h = 24$, 2d to $h = 48$, 3d to $h = 72$, 1w to $h = 168$, 2w to $h = 336$, 3w to $h = 504$, 4w to $h = 672$).



- We find that the monthly and daily dummies are among the main determinants of traffic flows.
- $\overline{TRA}$ and $\overline{CRO}$ in general appear significant although $\overline{CRO}$ plays a minor negative effect.
- Each pair of ACEs appears affected by different lags h of $\mathbf{y}_{t-h}$. Moreover, many lagged variables appear not significant.
- We find evidence that **including** $\overline{CRO}$ **and** $\overline{TRA}$ **improves the fitting** of the model.

	AIC		**BIC**	
Model	with $\overline{CRO}$, $\overline{TRA}$	no $\overline{CRO}$, $\overline{TRA}$	with $\overline{CRO}$, $\overline{TRA}$	no $\overline{CRO}$, $\overline{TRA}$
Castegnato - Gussago	-8.464	- 8.421	-8.306	- 8.269
Rodengo - Gussago	-8.443	-8.301	-8.285	-8.150
Rodengo - Castegnato	-7.797	-7.736	-7.640	-7.584

Conclusions

Three sources of mobile phone data have been used to represent different social phenomena related to people's movements. Traffic flows have been estimated between three small areas on the basis of crowding and traffic intensity indicators. Moreover, our model captures complex seasonality through lagged variables and weekly and monthly dummies. As a future development, the model forecasting ability will be investigated.