BANCA D'ITALIA
EUROSISTEMA

How can big data improve the quality of tourism statistics?

The Bank of Italy's experience in compiling the "travel" item of the Balance of Payments

Andrea Carboni, Costanza Catalano, Claudio Doria
Department of Economics, Statistics and Research – Bank of Italy

11th IFC Biennal conference
Basel, 26/08/2022

Big data for tourism statistics. The Italian case

BANCA D'ITALIA
EUROSISTEMA

Tourism statistics: number, expenditures and nights spent of

- Foreigners travelers visiting Italy (the reporting country)
- Italian travelers visiting abroad

BoP standards: expenditures by counterpart countries, business vs. personal trips, border/seasonal workers, international transports…

Sources: sample survey at border points (since 1996)

Drawbacks: costly, time-demanding, subjected to external factors (e.g. the covid-19 pandemic)

Big data: timelier, cheaper, less impacted by external shocks

Experiments on:

- Mobile Phone data
- Electronic payment data
- Internet search queries (Google Trends)

Mobile Phone data

May represent an alternative data source to count travelers crossing the border points.

➡ Only **complementary** source, no info on expenditures

- Arrival of a foreign traveler: signaled by the connection of foreign SIM cards to the cells controlled by an Italian network operator;
- Departure of an Italian traveler abroad: disappearance of the signal of an Italian SIM card near the border.

Nationality of the company issuing the SIM card = proxy for the traveler's country of origin

Collaboration with one of the major Italian Mobile Network Operator:

- Algorithm for estimating travelers inflows and outflows
- Constant cooperation necessary to achieve BoP standards (ex. minimum docking time due to handover effect)
- Tests on two main Italian border points (Fiumicino airport and Tarvisio highway)

Mobile Phone data

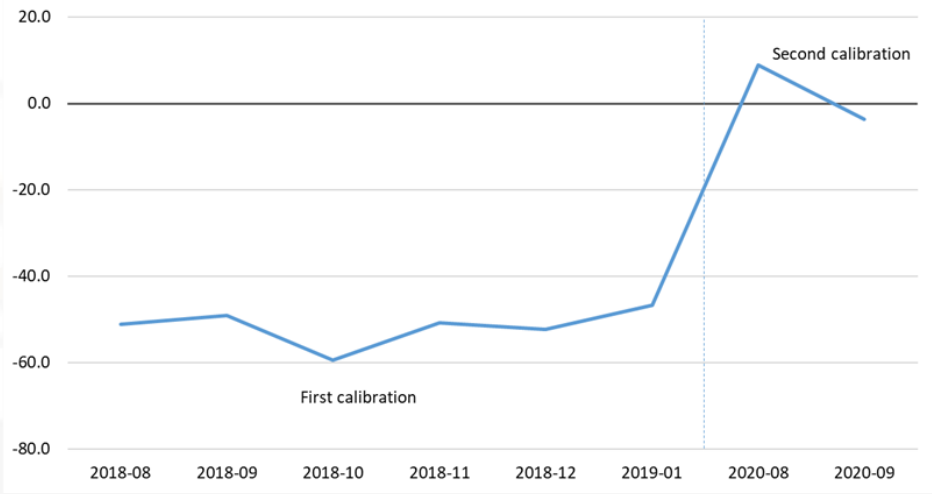
BANCA D'ITALIA
EUROSISTEMA

Fiumicino airport

	TOTAL			ITALIANS			FOREIGNERS		
	BI[(1)]	MPD[(2)]	MPD/BI%	BI[(1)]	MPD[(2)]	MPD/BI%	BI[(1)]	MPD[(2)]	MPD/BI%
Aug-18	1,717,076	1,802,051	4.9	640,288	621,419	-2.9	1,076,788	1,180,632	9.6
Sep-18	1,574,571	1,723,145	9.4	446,884	516,638	15.6	1,127,687	1,206,507	7.0
Oct-18	1,380,639	1,590,179	15.2	423,402	449,204	6.1	957,237	1,140,975	19.2
Nov-18	1,053,956	1,220,903	15.8	392,909	466,087	18.6	661,047	754,816	14.2
Dec-18	1,037,503	1,045,675	0.8	506,530	417,820	-17.5	530,973	627,855	18.2
Jan-19	831,120	1,113,629	34.0	344,529	457,947	32.9	486,591	655,682	34.8
Total	7,594,865	8,495,582	11.9	2,754,542	2,929,115	6.3	4,840,323	5,566,467	15.0

Tarvisio highway: huge discrepancies (order of 50%), needed a second test where the docking time was shortened.

Nowadays: The Bank of Italy has partially replaced the counting procedures in the frontier survey by MPD



Electronic payment data

Database: aggregated by date, nationality of bank emitting the card, type of purchase (10 categories), country of the POS/website

Foreign card & Italian POS/website ➡ Foreigners expenditure in Italy
Italian card & foreign POS/website ➡ Italian expenditure abroad

Main drawbacks in using payment data for BoP statistics:

- The nationality of the card is a proxy of the traveler's residence
- Confidentiality issues allow only aggregated data
- No info on the reason of the trip (business/personal)
- Difficult to correctly classify the Digital Platform transactions:

Expedia

airbnb

Booking.com

  - Payment of a stay in Paris by an Italian on Airbnb is recorded as from Italy to Ireland and not to France;
  - Payment of a stay in Rome by a French on Airbnb is recorded as from France to Ireland, thus not appearing in the database;
  - Payment of a stay in Rome by an Italian on Airbnb is recorded as from Italy to Ireland, despite it is a domestic trip.

Electronic payment data

BANCA D'ITALIA
EUROSISTEMA

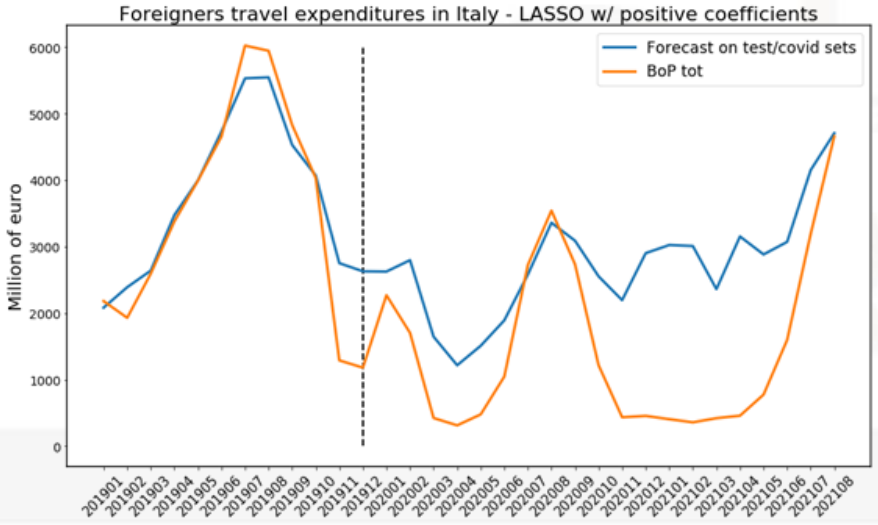
Market-share unknown ➡ Impossible the grossing-up of raw data

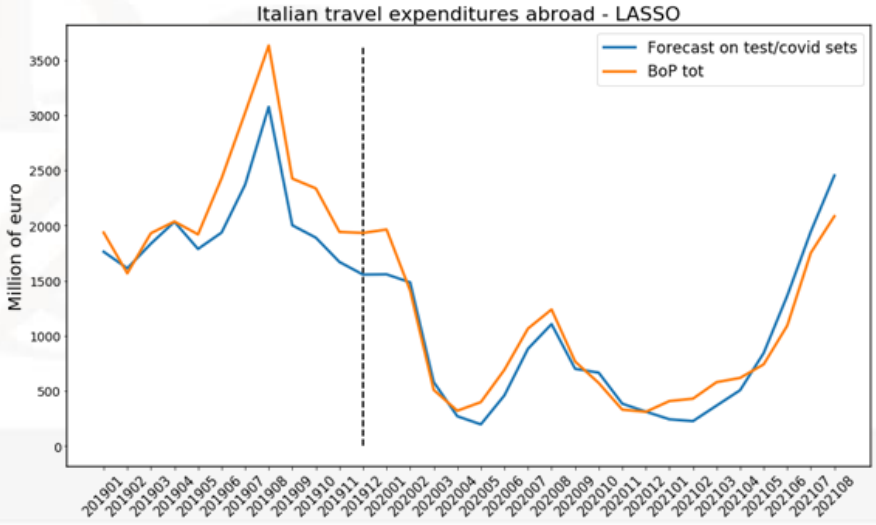
Tested some forecast models:

Ridge, Lasso, regression trees, boosted regression trees

Training set: years 2015-2017, Validation set: year 2018, Test set: year 2019, Supplem. test set: covid years.

Best performance in terms of MSE on the test set:





Internet search data: Google Trends

BANCA D'ITALIA
EUROSISTEMA

Google Trends index: reports the popularity of a given query in a given time period, country and category. It spans from 0 to 100.

Can the GT index be used to improve the provisional estimates on the number of travelers?

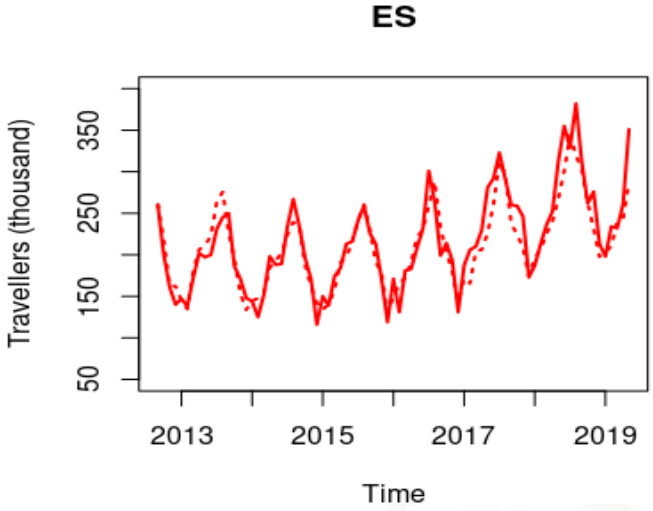
The model: seasonal AR(1)

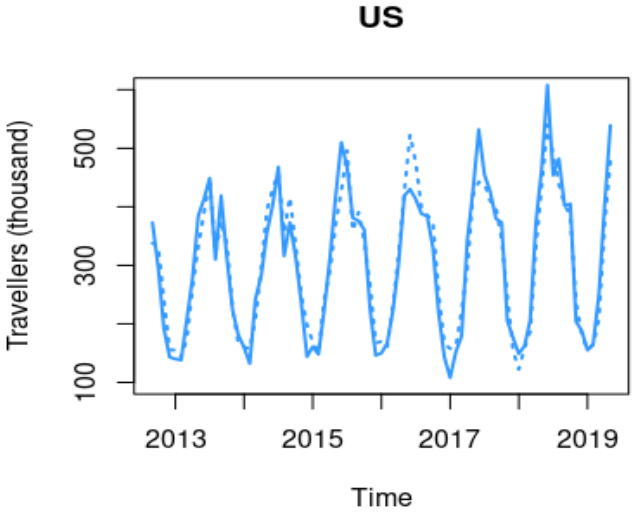
$$N_{c,t} = \phi_0 + \phi_1 N_{c,t-1} + \phi_{12} N_{c,t-12} + \beta GT_{c,t-l} + \varepsilon_{c,t}$$

- queries including the word 'Italy', category= 'Travel'
- $N_{c,t}$ = number of travelers from country c in month t
- one-step ahead forecast with expanding windows approach
- lag for GT index chosen by minimizing the out-of-sample MSE
- five countries: France, Germany, Spain, UK and USA

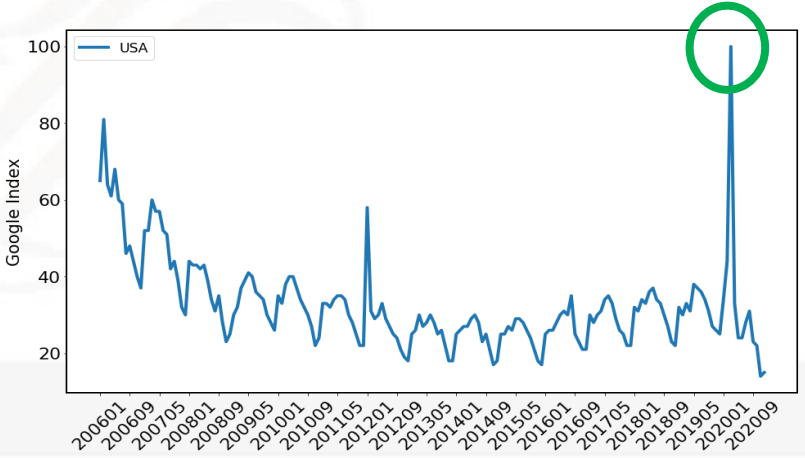
Results: In all cases the GT index increased the performance of the predictive model, except for France where β was not statistically different from zero.

Internet search data: Google Trends





Limits: peak of search queries in March 2020 while Italy was blocking the tourist inflow.
In presence of extraordinary events the Google classification seems to be less effective with high risk of outliers.



Take-away

All the data sources needed adjustments in order to define metrics that are coherent with the BoP standards.

Mobile phone data:

- the most suitable ones to be integrated with the frontier survey in the estimate of the number of international travelers
- Bank of Italy uses MPD for tourism statistics since the end of 2020

Electronic payment data:

- useful to achieve a preliminary timelier estimation of the total expenditure of the "travel" item
- For now, the mentioned relevant issues make it usable only for checking purposes

Google trends:

- useful as explanatory variable for estimating the number of international travelers
- possible noise of the index could be misleading. Use of other/more words as search queries could generate more accurate results.

Thank you for your attention!

…Questions?