Combining mobile phone data and tax data to shed a new light on social segregation in urban areas

DGINS: 10th to 11th of October 2018

Sylvie Lagarde

Insee
Mesurer pour comprendre

Why combine mobile phone data and NSI data ?

To get the most of Mobile Phone Data (MPD) for Official Statistics it is necessary to combine them with NSI data at a fine level of granularity.

➢*Getting relevant data from MNOs*

- **Requires aggregated MPD** in a useful form for later statistics production
- Requires accurate spatial units for calibration when aggregating **individual MPD**
- Ensures **control over the methodology**

➢*Accessing data from MNOs*

- Brings data to the table in the **negociations**
- Demonstrates the **NSI added value**
- Benefits the MNOs in return from ***ad hoc* aggregates** shared with them

How to combine mobile phone data and NSI data ?

- ***On-going agreement between INSEE, Eurostat and Orange Labs on a specific dataset for research***

1) *Gives access to the MNO’s data on their in-house Big Data infrastructure*
2) *Puts some specificaly aggregated tax data (guaranteeing privacy) on that infrastructure*
3) *Computes individual indicators and aggregates on their premises*
4) *Exports only the aggregates to INSEE for final analysis*

► In a sense, we share the production process with the MNO by:

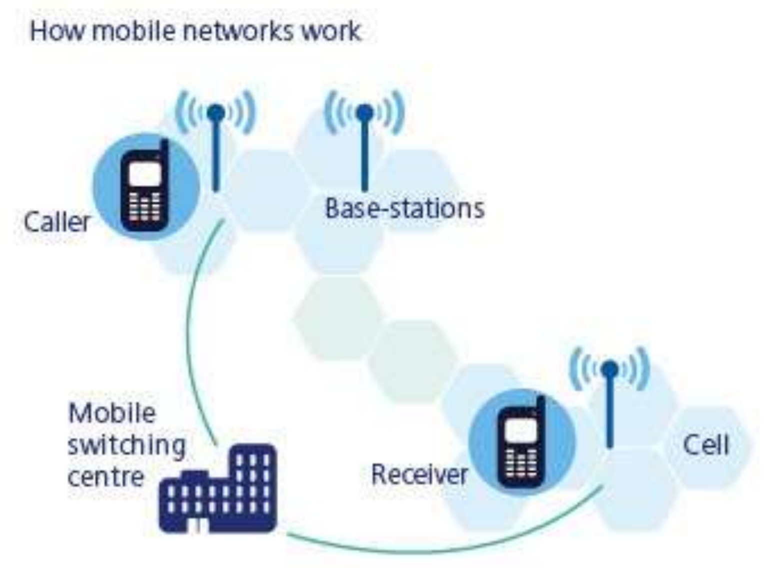
- going **beyond the use of external data** and accepting running part of the production from our partner’s premises.
- But we should keep **control over the methodology**

Insee Mesurer pour comprendre

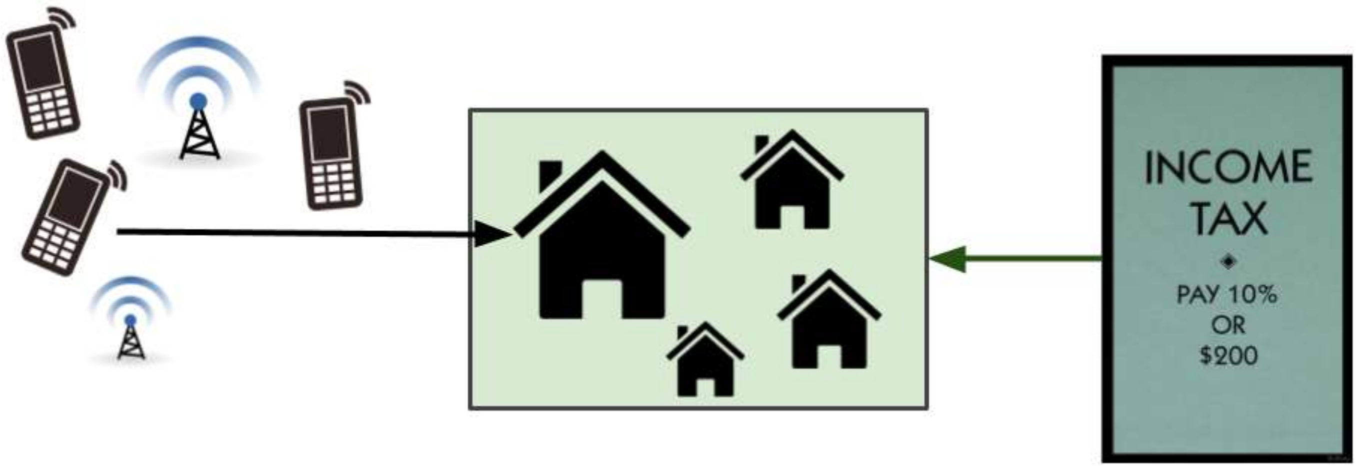
Example of a study on social segregation

A study that combines mobile phone data and tax data to describe social segregation in urban areas

- **Mobile Phone Data: 5 months of Call Details Records (active data) from 2007**
- 18 million users
- who is calling who ?
- through which antenna does the signal pass ?
- **Tax data :** geo-coded household incomes



How to estimate phone users' income ?



1)Map the antenna-coverage areas corresponding to the chosen grid of analysis (500m^2 cells)
see paper for details

2)From calls and SMSs sent during nightime hours (September 2007) : estimate the place of residence

3)From geo-coded tax data : compute the median income of that cell

4)►an example of combining MNO and NSI data, based on geographical attributes from both sides

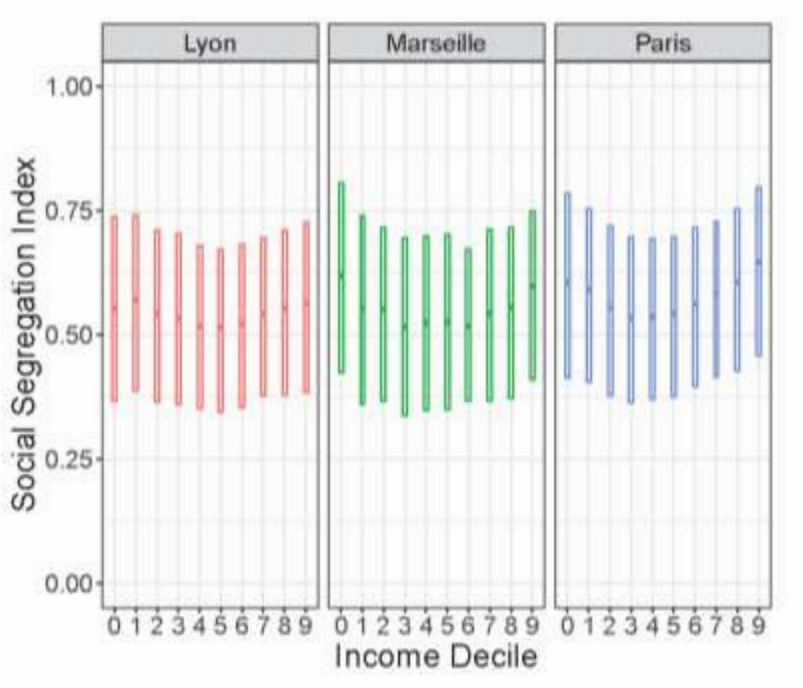
Segregation indicators

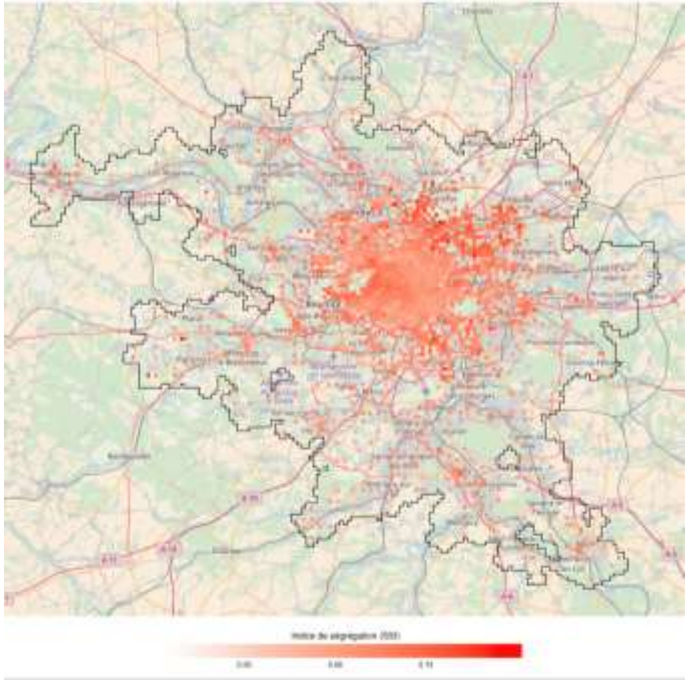
Social segregation in the sense of people tending to have more contacts – *via the phone* - with similar others - *in terms of income level of the place of residence.*

- Phone users in an urban area are ranked according to their income estimate.
- From that ranking we compute a “social similarity index” between pairs of users.
- The individual social segregation index is the average of the social similarity indexes between a user and all its contacts weighted by the frequency of their phone calls.

Results

- Individual social segregation indexes are aggregated by income decile or cell of residence





- Social segregation is present in all 3 urban areas and higher at the extremes of the income distribution (or within rich or poor neighborhoods)

Conclusion, limits & future plans

Conclusion

- Original measure of segregation
- Complementary to traditional ones on residential segregation and coherent with them

Limits

- A specific mode of social contacts (phone calls and SMSs)
- A bias in having data from only one single MNO and from an old dataset
- Home detection estimation is not always plausible – could achieve better with signaling data (passive data)
- Income assignation relies on home detection and is uniform within a cell

Future plans

- Next: simulating an individual income
- Another dimension of segregation: being in the same place at the same time

Collaborating closely to combine data sources

Take home message

- Combining data from two sources based on geography enhances the usual statistical production by :
  - describing a social phenomenon unobservable with only one of the two data sources
- But, it requires a close collaboration with the MNOs

Keep in touch

insee.fr

Sylvie LAGARDE