Comparison of data of mobile operators and Rosstat on the number and distribution of the population in the Moscow Metropolitan region

Roman Babkin, Postgraduate, Moscow State University
POSSIBLE SOURCES OF INFORMATION ABOUT THE POPULATION AND ITS DYNAMICS

- **Traditional statistics**
- **Departmental statistics**
- **Data from opinion polls**

「Big Data」

- Census of the population
  - data on the current account of the population

- **«Big Data»**
  - mobile operator data
  - data about Bank card transactions
  - social network data
  - satellite image data
  - mobile app data
  - etc.
Strengths of both data types

Rosstat data:
- high coverage of all population groups
- high detailed data and its structure (gender, age, education, etc.)
- ability to compare between different territories of the country
- long period of data collection, possibility of analysis based on long time series

Mobile phone data:
- high correspondence of reality in time and space
- calibration capability for existing data
- no reference to an administrative division
- applicability to different spatial scales and time slices
- comparability and integration with other spatial information sources
- high penetration rate among all population groups
Weaknesses of both data types

Rosstat data:

- double counting of the population in the census, when a person is first counted at the place of permanent registration, and then live census at the place of real residence
- poor accounting of the real place of residence, especially for pensioners, students, and military personnel
- under-reporting of migrants and waste workers

Mobile phone data:

- non-personalized nature and related problems of accounting for mobile phones registered to other persons (for example, relatives), or people who do not have mobile phones
- under-accounting of SIM-cards of other operators and disabled devices
- technical interference: overlapping signals from neighboring repeaters, causing subscribers to "jump" from cell to cell, not related to real movements in space
- high data cost
- short period of information collection
Two data types do not match

Types of districts

- The overestimation of the population
- The underestimates of the population

With an approximate match population in both types of data

- Residential municipalities of Moscow
- Cities of the Moscow oblast
- Closed city of the Moscow oblast
- New Moscow
- Most municipalities in the Moscow oblast
- Center and subcenter of Moscow
- Peripheral areas of the Moscow oblast (in the Eastern sector)
The main features of the area, affecting the relationship between the two data sets:

- Accommodation in the second (dacha) housing of part of the citizens registered here.
- Disadvantages of population censuses that serve as "reference points" for subsequent current accounting.
- A large number of residents registered at their place of residence (often changing their location, which is not accounted for by official statistics).
The underestimates of the population

The main features of the area, affecting the relationship between the two data sets:

- Active housing construction.
- Purchase of housing in the Moscow region or in New Moscow without re-registration.
- A significant number of unregistered residents (including illegal migrants).
- Permanent residence of some Muscovites in their dacha.
- The spread of "semi-commuting".
With an approximate match population in both types of data

The main features of the area, affecting the relationship between the two data sets:

For the center and subcenter of Moscow:
- Significant scale of night employment (due to the large number of offices).
- A large number of hotels.

For the Eastern periphery of the Moscow region:
- Low involvement of the population in pendulum migrations, the spread of "semi-commuting"
- Less than for other sectors of the region, the spread of suburban suburban development
Conclusions

- The emergence of a new source of population data – data from mobile operators that is comparable to Rosstat in terms of coverage – makes it relevant to analyze these two data sets.

- The use of mobile data is appropriate as an alternative source of information for demographic statistics.

- Official statistics significantly underestimate the population of the Moscow region and slightly overestimate the population of Moscow, and in general underestimates 2.2 million people living in both regions.

- In addition, official statistics, due to their inertia, are unable to provide reliable information about the population of municipalities where active housing construction is underway (New Moscow, some municipalities of the Moscow region and Old Moscow outside the Moscow ring road), taking into account 2-4 times fewer residents than actually live in these territories.

- Rosstat data does not take into account the seasonal population and residents living in second homes or dacha, as well as unregistered migrants. We can say that "invisible" categories of the population in official statistics are shown in the data of mobile operators, and for "visible" groups, it is possible to determine their actual residence.
Thanks for your attention