What are the similarities and differences in deriving migration and tourism statistics from mobile positioning data?

1 National Statistics Office of Georgia

2 BPS Statistics Indonesia

Shorena Tsiklauri[1]
Mariana Jalagonia[1] Titi
Kanti Lestari[2]

Abstract

Migration is one of the main processes that affect the size and distribution of population. From the demographic point of view, it is also harder to estimate than the component of the natural increase.

The concept of international migration as a change of the country of residence is relatively easy to understand, but difficult to capture, because human mobility is not a discrete process, they used to permanently change their place of usual residence, particularly when taking into account rapid socio-economic developments in the contemporary world and it becomes challenging for several reasons: migration is mainly depend on the improve the quality of one's life and migrants keep changing their place of usual residence. The migrants do not always cross boundaries legally, even they are moving legally, they do not always report their intentions of moving.

Tourism refers to the activity of visitors. A visitor is a traveler taking a trip to a main destination outside his/her usual environment, for less than a year, for any main purpose (business, leisure or other personal purpose) other than to be employed by a resident entity in the country or place visited. These trips taken by visitors qualify as tourism trips.

In the modern society tourism has significant impact on the human life and is considered as one of the largest industries in the world. Recently, tourism has become one priority or national priority in some countries, moreover three SDGs indicators were also related to tourism (Goal 8 and 12).

Mobile phone data as one of the big data sources offer a great promise for its use in human mobility measurement. Mobile positioning is often considered to be a novel and exciting source of information for investigating the spatial dynamics of human society. There is an expectation that mobile phone data could fill data gaps and make no one left behind. However, new data sources cannot be substitution of the traditional ones.

The paper aims to provide the use of a new and innovative data source i.e. mobile positioning data to estimate official migration and tourism statistics. In this paper, mobile positioning data is used as complement of other data sources such as administrative data as well as replacing other data sources (i.e. survey data). Moreover, the paper also shows the similarities and differences of the use of mobile phone data for official migration and tourism statistics.

Keywords: *Big Data, MPD, migration, tourism*

References:

UN Committee of Experts on Big Data and Data Science for Official Statistics
website: https://unstats.un.org/bigdata/index.cshtml

What are the similarities and differences in deriving migration and tourism statistics from mobile positioning data?

Shorena Tsiklauri[1]

Mariana Jalagonia[1]

Titi Kanti Lestari[2]

1 National Statistics Office of Georgia
2 BPS Statistics Indonesia

Keywords: *migration, tourism, official statistics, MPD, Big Data*

Introduction

Migration is one of the main processes that affect the size and distribution of population. From the demographic point of view, it is also harder to estimate than the component of the natural increase.

Definition of **internal migration** usually varies across the countries, i.e. relies on the definition of the usual residence and data source(s). The UNECE for the census purposes defines the internal migrants as "persons who are usually resident in a particular geographical area and who have previously been resident in another geographical area in the country of enumeration"[1]. According to UN recommendation „**international mobility** includes all movements that cross international borders within a given year. **International migration** is a subset of international mobility and includes all movements resulting in a change in the country of residence within a given year"[2].

There are several reasons for migration. The most widely given is based on the concept of income differentials, based on neoclassical economic theory. Additionally, an important factor today is the social factors such migrating to join family or friends. At the same time, political situations or environmental threats also can be push factors for population movement.

Tourism refers to the activity of visitors. **A visitor** is a traveler taking a trip to a main destination outside his/her usual environment, for less than a year, for any main purpose (business, leisure or other personal purpose) other than to be employed by a resident entity in the country or place visited. These trips taken by visitors qualify as tourism trips.

Recently, tourism has become one priority or national priority in some countries, moreover three SDGs indicators were also related to tourism (Goal 8 and 12).

The concept of international migration as a change of the country of residence is relatively easy to understand, but difficult to capture, because human mobility is not a discrete process, they used to permanently change their place of usual residence, particularly when taking into account rapid socio-economic developments in the contemporary world and it becomes challenging for several reasons: migration is mainly depend on the improve the quality of one's life and migrants keep changing their place of usual residence. The migrants do not always cross boundaries legally, even they are moving legally, they do not always report their intentions of moving.

Mobile phone data as one of the big data sources offer a great promise for its use in human mobility measurement. Mobile positioning is often considered to be a novel and exciting source of information for investigating the spatial dynamics of human society. There is an expectation that mobile phone data could fill data gaps and make no one left behind. However, new data sources cannot be substitution of the traditional ones.

---

[1] UN, Principles and Recommendations for Population and Housing Censuses, 2017
[2] Conceptual frameworks and Concepts and Definitions on International Migration - https://unstats.un.org/unsd/statcom/52nd-session/documents/2021-11-MigrationStats-E.pdf

This paper aims to provide the use of a new and innovative data source i.e. mobile positioning data to estimate official migration and tourism statistics. In this paper, mobile positioning data is used as complement of other data sources such as administrative data as well as replacing other data sources (i.e. survey data). Moreover, this paper also shows the similarities and differences of the use of mobile phone data for official migration and tourism statistics.

Traditional sources vs new data sources

Traditional censuses and population registers have been considered as the solid sources for assess the long-term processes of migration. For the short-term and everyday mobility, more flexible methods such as various registers and indirect databases[3], satellite-based methods[4], or modern sensing technologies[5] are needed, because classical data sources are designed to describe the resident population. Even if census or register data contain data about the existence of another home, these sources do not offer extensive information regarding the amount of time spent there and year-round activities.

The use of mobile positioning data has the potential to improve several aspects of migration statistics, such as timeliness (in some cases up to near real-time), access to statistical information previously not available (new indicators), calibration opportunities for existing data, space and timely resolution/granularity (i.e. the scale or level of detail in a set of data) and accuracy[6].

The advantage of MPD derived information is its automatic (independent) collection and relatively low costs. However, mobile positioning data do not cover the entire population, and therefore, it is not possible to extract all possible migration events. In addition, it should be considered that MPD includes noise and erroneous records and special care should be taken to clean the data of errors and noise, because they might affect the final result[7].

The mobile positioning data cannot be a replacement, but rather a supplementary source for the official migration statistics[8]. For tourism statistics, it can both complementary and replacing other data sources.

Methods for population mobility using MPD

Migration is not easy to measure. Migration can be classified by several factors:

- Administrative border crossing (international or internal migration, tourism);
- direction (inward, outward, origin-destination);
- purpose (labour, education, social, economic, ecological, political conflicts, violence etc.);
- duration of stay.

According to UN recommendations[9] an **international migrant** is defined as a person who has changed his or her country of residence and established new residence in the country within a given year. International migrant can be either 'immigrant' or 'emigrant' and include those with national or foreign citizenships or stateless persons.

Nevertheless, the definitions are mainly applicable for traditional data sources. Applying new technologies for the human mobility analysis based on the classical definitions seems unrealistic and causes much uncertainty. In order to identify a migrant, detailed algorithm of the data processing and methodology has to be developed. It has to be noted that compliance with the UN principles and recommendations, as well as considering the best practices is important.

---

[3] Raymer et al., 2007
[4] Chen et al., 2006
[5] Kwan, 2000; Eagle and Pentland, 2005; Shoval, 2007
[6] Ahas, R., J. Armoogum, S. Esko, M. Ilves, E. Karus, J. L. Madre, O. Nurmi et al. "Feasibility study on the use of mobile positioning data for tourism statistics. EuroStat." (2014): 235-251.
[7] Draft: Practitioners' Guide on Harnessing Data Innovation for Migration Policy (to be clarify after publishing)
[8] Rango, M., and M. Vespe. "Big Data and Alternative Data Sources on Migration: From Case-Studies to Policy Support." Summary Report, European Commission Joint Research Centre (2017) - https://bluehub.jrc.ec.europa.eu/bigdata4migration/workshop-outcome
[9] Conceptual frameworks and Concepts and Definitions on International Migration - https://unstats.un.org/unsd/statcom/52nd-session/documents/2021-11-MigrationStats-E.pdf

Main challenges in methodology of migration using MPD are connected to identifying the place of residence (home), and the identification of the time period when a certain place is home and when it is changed.

The MNOs do not cover the whole population and there is a need of some statistical methods and skills. With extrapolation methods the absolute numbers could be presented. But if there is no reliable reference data to calibrate MPD results, relative numbers regarding migration are more reliable (e.g. the number of subscribers with changed home location compared to the total number of home locations in an area). The extrapolation to the general population is problematic, as the number of mobile devices or SIM cards is never equal to the number of people[10].

Data processing algorithm

In order to create an algorithm for data processing, first of all, it is crucial to introduce the notion of "the usual place of residence" for the person and "Nighttime population", which cannot be achieved based on the traditional data sources.

Due to the opportunities available to replace movements and live in multiple residences[11], the dynamics of the act of migration might no longer be quite so clear. When the person visits several locations, one needs to identify "home" based on the duration, e.g. the location where the majority time is spent. This means that the duration of stay defines the meaning and importance of different locations. Subsequently, it allows to distinguish the change of residence from other moves.

Another factor which would help to identify the meaning of location is the start and end times of trips. These indicators vary according to the visit type and this makes it possible to view the trips as related to either work or vacation. In case of trips related to work, a clear working-week pattern is discernible, the trips start most often on a Sunday or a Monday and end on a Thursday or a Friday. During weekend the people usually are visiting their second home or move for tourism reason.

Tracing activities (incoming/outgoing calls and SMS, internet connections) of the person with the time linkage would subsequently help us to determine "home" (where the person spends most of the night time and weekends) and "usual environment" in general (for example, working place, where the person spends most of the day time during the weekdays, as many of cases show). For domestic subscribers, **the usual environment** can be defined with the precision of country of reference, region, municipality or some other geographical areas or administrative units. The level of detail used depends on the data available and on the producers' needs (which differs from the official definition). Eventually, usual place of residence can be identified as the place where the person has spent most of the nights rather than relying on the data based on the frequency of communications.

As regards the outgoing calls, the following parameters have to be considered:

       a) the exact time of the call activity;

       b) the random ID number for the phone (not related to phone or SIM card number);

       c) the cell ID with the geographical coordinates of the antenna.

    –   *Change of residence/usual environment*

To define a change of residence or usual environment most likely home location has to be assigned. Some scholars[12] are offering the method with following assumptions:

    (1) using the locations from which a person has made calls, that is, the places that the person has physically visited, and based upon the intensity of calls and when the calls are made, it is possible to assign meaning to each location;

---

[10] Draft: Practitioners' Guide on Harnessing Data Innovation for Migration Policy (to be clarify after publishing)

[11] McHugh et al. 1995, Renkow and Hoover 2000, Axisa et al. 2012, Huber and Nowotny 2013

[12] Kamenjuk, Aasa, & Sellin, 2017

(2) a cluster of home anchor points that can be derived from the anchor point model for at least 7 months over the course of a 13-month period is defined as a stable home area (SHA) which itself can be interpreted as a usual place of residence;

(3) and change in a SHA is defined as a change of residence.

Anchor points are locations where people regularly stay. Regular cells are referred to as anchors when the model gives meaning (home, work-time, multifunctional) to them. Frequent use of certain mobile phones in specific locations gives the possibility to learn about important or meaningful places for this phone (person). **Meaningful places or meaningful locations**[13] are defined as regularly visited places that have meaning for individuals. Technically, they are similar to personal anchor points; home and work anchors are the most common among them. Anchor points are important variables in describing humans' behavior in time and space. Anchor point modeling is one of the possibilities for making useable the anonymous data of passive mobile positioning, GPS tracking, etc.

For calculating anchor points and finding the home and work-time locations, as well as secondary anchors for every ID the model was developed and are including the following steps: [14]

1. determine points of regular cells and separates them from random cells;
2. remove from the database persons with too high or too low number of calls. If the number of calls made is too low, it is not possible to calculate anchor points. The reason for there being too many calls is an organized call procedure (service center etc.) or a technical device using a GSM network;
3. define home and work-time anchor points. Home and work-time anchor points are determined using regular cells, based on the average start time of calls (the average of all calls made during a 24 h day) and the standard deviation of call commencement times;
4. consider the neighboring relationship in the case of 2 home or 2 work-time anchor points;
5. assess the proportion of days spent at an anchor point;
6. determine the missing home or work-time anchor point by the addition of a third point;
7. classify an anchor point as the missing home or work anchor point;
8. format every day and secondary anchor points.

– *Internal Migration*

Internal migration or domestic migration is human migration within a country. Internal migrants are broadly defined as persons who change their place of residence within the country in a given year[15]. The main reasons for internal migration are education, economic improvement, natural disaster, civil disturbance, family, amenity or other socio-economic or political issues. A general trend of movement from rural to urban areas, in a process described as urbanization, has also produced a form of internal migration[16].

The notion of the "usual resident" when analyzing MPD is often associated with the 50%+1 principle. Based on the MPD non-roamers data as a starting point, it is possible to observe the person during 12 month (or more) and consider him/her as the usual resident of the municipality, where more than 50% of communications (incoming/outgoing calls and SMS, internet connections) have been traced in the first half of the 12-month period; the same person can be considered as an internal migrant if the same occurred in the second half of the observation period in another municipality. However, this criterion is not enough for identifying the migrant. Together with the number of connections/communications, it is essential to specify the time also. It can be done in the following manner:

Based on the IPDR/CDR data, one could distinguish the following categories of the population:

- Nighttime population – activities traced from 07:00 p.m. to 06:59 a.m.;

- Daytime population – activities traced from 07:00 a.m. to 06:59 p.m.

[13] Nurmi and Koolwaij, 2006
[14] E. Saluveer, O. Järv
[15] https://www.unece.org/fileadmin/DAM/stats/publications/2015/ECECES41_EN.pdf
[16] World Migration Report 2020. https://www.un-ilibrary.org/migration/world-migration-report-2020_b1710e30-en

After observing the population during the 12-month period, namely analyzing movements to/from the place of origin (municipality/region), checking the territorial unit border cross patterns, can help to identify the migrant for the specific time point.

> – *International Migration*

From the point of view of MPD, the concept of migration can be defined as international mobility of subscribers. Several factors need to be taken into account:

If a person lives and is active in several spaces, his or her phone number(s) will be registered in one or several of these countries. Usually, a mobile operator is chosen from the country that the person has the strongest connection to, where they spend more time than elsewhere and where the important people and social networks are located. It may also be important to be available for business, social or state service reasons. If a person decides to commit permanently to a destination and switches off the mobile phone number (SIM-card) of the previous homeland, this is a sign of being estranged from the previous homeland. It should be noted that roaming rates matter as well.

The problem which arises analyzing international migration is that local MNOs might not possess information on the person's movements abroad unless he/she is not using roaming. However, based on the border cross data (dates of inward/outward crosses) can help to infer missing information about the SIM card activity (person's). Based on that the duration of being abroad can be assessed.

Some assumption-based methods might also be applied. For example, the disappearance of foreign or appearance of new domestic SIM cards can suggest the arrival of migrants in specific areas of the country.

> – *Tourism*

In official tourism statistics, the important thing, as mentioned above were the concept of usual environment. People considered as visitors or tourists were if there are outside their usual environment. The home-work anchor model is actually the usual environment of a person.

- For domestic tourism, the anchor point model mentioned above is used to detecting home of the subscribers. Then this data can be used to identify the tourist, which is subscribers that have trips outside their usual environment. While the usual environment data can be used as commuting statistics (non-permanent migration) and if more than one year it can be called internal migration.

- For inbound and outbound tourism, the method will show the residency of the subscribers, the whether they are tourists, transnational migration and border workers.

References:

UN Committee of Experts on Big Data and Data Science for Official Statistics
website: https://unstats.un.org/bigdata/index.cshtml