BANCA D'ITALIA
EUROSISTEMA

Statistics

Methods and Sources:
Special Topics

28 January 2026

For further information: statistiche@bancaditalia.it
www.bancaditalia.it/statistiche/index.html

Mobile phone data and the "Travel" item in the Balance of Payments: the Bank of Italy's experience

by Andrea Carboni, Claudio Doria and Valeria Valletta[1]

Key Messages

- Since the end of 2020, the Bank of Italy has used mobile phone data (MPD), alongside administrative sources and manual counting operations, to estimate the number of international travellers.
- Compared to traditional techniques, MPD offers continuous temporal coverage, broader geographical reach, greater timeliness, and reduced costs relative to manual counting operations.
- Effective integration of MPD into tourism statistics requires ongoing collaboration with mobile network operators to ensure the adoption of procedures and methodologies that filter raw data so as to align with official statistical definitions.
- Comparative analyses with administrative data from border authorities, considered highly reliable, suggest that MPD provides high-quality estimates of international tourist flows.

Abstract

Despite the potential of Mobile Phone Data (MPD) as a resource for international tourism statistics, few countries use them in official statistics due to the difficulty in translating tourism definitions into telephony terminology. Since 2020, the Bank of Italy has integrated MPD into the estimation process of the "Travel" item of the Balance of Payments, resulting in more comprehensive information on border crossings and lower costs compared to traditional survey methodologies. This paper describes the advantages and challenges related to the use of MPD and their performance through a comparison with administrative data sources.

[1] External statistics division – Statistical analysis directorate.

1. Introduction

Technological advancements, the spread of the internet and the proliferation of digital products and services have generated new data flows with heterogeneous characteristics. These data – which involve large volumes of information, a high update frequency, and are complex to use – are commonly referred to as "big data".

One type of big data is mobile phone data (MPD, also called mobile positioning data), which provide valuable information about individuals' location and movements, representing a potentially significant resource for the study of social mobility and, more specifically, of tourism.

Tourism is a complex phenomenon, with a significant impact on social, environmental, and economic dimensions, being an important share of the global gross domestic product (Leng et al., 2016). Analysing and understanding tourism is crucial for effectively managing its impact on destinations and communities, and it enables informed decision-making regarding: a) the urban space design (Xu et al., 2023), i.e., the sustainable development of transportation and infrastructure; b) the enhancement and preservation of natural and cultural heritage; c) the promotion of local culture and events, creating job opportunities and encouraging the development of tourism services to support the local economy (Saluveer et al., 2020).

Until a few years ago, tourism statistics mainly relied on expensive and demanding surveys with large sampling errors. More recently, research in this field has begun to leverage MPD, recognizing the advantages of this data source in estimating traveller flows, including lower costs, timeliness, continuous data collection and broad geographic coverage. In general, MPD enable full coverage of the surveyed territory, and they reach a large share of the population due to the widespread use of mobile phones; however, their reach is strictly dependent on the market share of the Mobile Network Operator (MNO) providing the data and the extension of its network. Despite growing international attention (Demunter, 2017; UN, 2019; UN-ESCAP, 2022; Eurostat, 2014, 2023), leveraging MPD for official tourism statistics by individual countries remains very limited. On the one hand, the reason for this limited use can be attributed to the strict privacy regulations that MNOs must adhere to in the vast majority of countries; on the other hand, translating the definitions spelled out by international statistical standards for official tourism statistics into telephony terminology is objectively complex (Grassini & Dugheri, 2021; UN, 2019).

Against this background, the Bank of Italy, in the constant search for new sources and methodologies to improve the statistics it publishes, has begun to integrate MPD into the estimation process of the "Travel" item of the Balance of Payments (BoP) following an experimentation phase started in 2018.

This study describes this experience, highlighting the main characteristics of the data used and how they are being used.

Section 2 reviews the existing literature on mobile phone data, with a particular focus on studies related to the estimation of international travel. Section 3 describes the survey used by the Bank of Italy since 1996 for compiling the "Travel" item of the BoP. Section 4 describes how MPD have been gradually integrated into the existing estimation process and the expected benefits. Section 5 outlines the current use of MPD, emphasizing their key characteristics, primary benefits and significant challenges. In Section 6 some quality issues of MPD are assessed through a comparison with survey and administrative data sources. Finally, in Section 7 sets out the conclusions.

2. Literature overview

2.1. Definition of MPD

MNOs collect a large amount of data in their systems for billing, monitoring, management, and network planning purposes.

In the literature, two types of MPD are primarily used: active and passive mobile positioning data (Ahas et al., 2008). In the former case, data are generated through a dedicated app, installed on the

phone; as they allow real-time tracking of the phone's location, a user permission is required for sending the data to the MNO. They are rarely used due to their limited volume.

Passive mobile positioning data, on the other hand, are collected automatically by MNOs, without the need for specific user consent and are the most commonly used in tourism research (Saluveer et al., 2020). They are further divided into event-driven data (such as Call Detail Records, CDRs) and network-driven data (such as signalling or probing data). The former are recorded by the MNO with detailed temporal and positional information whenever the mobile device user performs a trigger event, such as making a call or sending an SMS or MMS. The latter are collected periodically without a trigger event occurring.

Among the data collected by MNOs, CDRs have been the most commonly used in research, thanks to their greater availability, relatively low cost, and high scalability (Chen et al., 2018; Grassini & Dugheri, 2021; Rahmadian et al., 2023). However, the time frequency of CDRs is uncertain, generally low, and further decreasing, considering that the use of instant messaging apps continues to reduce the number of SMS, MMS and phone calls. Additionally, the time frequency of CDRs is closely dependent on the users' mobile device usage behaviour and, consequently, on their economic and sociodemographic characteristics and habits (Saluveer et al., 2020). Nonetheless, they are easily accessible and sometimes sufficient for research purposes.

Network-driven data are generally more efficient, detecting a higher number of SIM cards compared to CDRs, with a much higher time frequency, typically in the order of seconds or minutes; consequently, they produce a large amount of data that require specific technologies for their management (Grassini & Dugheri, 2021). This kind of data may be affected by statistical noise due to the presence of certain types of international travellers, such as those flying over the country, crossing seas, or unintentionally connecting to the network while being near territorial borders. This noise can generally be corrected, with good results, through appropriate techniques and algorithms (Lestari et al. 2018).

2.2. The use of MPD in research on tourism

Multiple studies in the literature show the significant role of MPD in various fields, including healthcare, socioeconomics, disaster response, urban management and more (UN, 2019).

In the field of travel, several analyses acknowledge the potential of such data in studying both domestic and international tourism. Ahas and Mark (2005) were among the first to recognize both the potential of MPD and the associated issues, primarily related to privacy. They introduced the Social Positioning Method which investigates social flows over time and space through phone localization and the social characteristics of owners.

Subsequently, interest in MPD grew along with the number of studies using them. In particular, there is considerable interest in studying travel behaviours and identifying travel patterns (Phithakkitnukoon et al., 2015; Chen et al., 2018; Xu et al., 2021; Park et al., 2020), assessing the impact of various factors on tourism such as climate change (Kubo et al., 2020), and defining new indicators (Leng et al., 2016).

In Ahas et al. (2007), different travel patterns are identified based on the season, as well as differences in the predominant nationalities between summer and winter tourism areas.

In Ahas et al. (2008) a detailed description of MPD is provided and the main points of attention are described, such as access issues due to privacy regulations, data collection methods, spatial resolution, network density, statistical noise due to traffic near borders, and various representativeness issues, such as different mobile phone usage among different population subgroups (especially applicable to CDRs).

Raun et al. (2016) develop a methodology for measuring visitor flows to specific destinations using space-time tracking data. The paper proposes that a tourism destination has five dimensions – spatial, temporal, compositional, social and dynamic. Results show that smaller destination areas within a country can be differentiated by the geographical, temporal and compositional parameters of the visits.

The issue of MPD quality, which can normally be verified by comparison with more reliable traditional sources, is of central importance. In several works, a comparison of MPD on incoming

flows with official statistics on accommodation is carried out (Ahas et al. 2007, 2008; Heerschap et al., 2014; Saluveer et al., 2020). MPD show a high correlation and superiority for areas with few structures and transit flows. For outgoing flows from Estonia Saluveer et al., 2020 report a low availability of data from other countries.

MPD also work efficiently in analysing domestic tourism. Indeed, Indonesia's national statistical institute (Badan Pusat Statistik - BPS) has leveraged MPD since 2018 for estimating domestic tourism and commuting.

The interest in MPD is part of the ongoing initiatives at BPS to develop big data governance (Rahmadian, 2023). Noviyanti et al. (2020) propose a framework to support the use of big data in the production of official statistics.

Finally, MPD enable the construction of origin-destination matrices with advantages relative to those derived from conventional data sources. (Caceres et al., 2020; Bonnel et al., 2015).

2.3. MPD for official tourism statistics

The use of MPD for compiling official statistics has begun only recently. Notable examples include contributions from France (Cousin & Hillarieu, 2020) and Netherlands (Heerschap et al., 2014).

Outside Italy, MPD have been officially integrated into the estimation of international tourism flows only in Estonia and Indonesia.

In particular, the Estonian Central Bank, Eesti Pank, has been using CDRs in combination with payment data for the compilation of the "Travel" item of the BoP since 2008 (Eesti Pank, 2019). The methodology used by Estonia is described and adopted by Saluveer et al. (2020), who, for the first time, use MPD to derive statistics on Estonian inbound and outbound international tourism flows (such as the number of travellers and nights spent). The authors emphasize the importance of developing and harmonizing regulations and laws for data access.

BPS-Statistics has been using MPD (signaling data) for compiling Indonesian official statistics on the number of arrivals of international travellers since 2016. Previously, administrative data on immigration and a border survey conducted one month a year were used. However, both sources faced significant issues of under-coverage, due to the vastness of the Indonesian territory, the variety of border points, and the scarcity of control points. MPD have allowed to achieve better coverage and timeliness compared to traditional sources (Lestari, 2018).

On a more general note, a limitation common to the different types of MPD is that they do not contain qualitative information, such as travel characteristics (i.e., main reason of the trip, type of accommodation, etc.) and, to some extent, traveller attributes. Since some of these characteristics are mandatory details prescribed by the international guidelines for the BoP compilation,[2] MPD need to be integrated with data from traditional surveys in order to obtain more comprehensive tourism statistics. Most importantly, it should be remarked that, while telephone data can provide information on the size of tourism flows, they cannot provide any information on the expenditure associated with these trips, which is the fundamental variable for BoP statistics.

3. The Bank of Italy's Survey on International Tourism

The Survey on International Tourism has been conducted by the Bank of Italy since 1996, with the main purpose of collecting information useful for the compilation of the "Travel" and "Transport - passengers" items of the BoP. Simultaneously, it also collects additional information on tourism flows, of interest for researchers, industry operators, and local authorities[3].

In the survey a stratified sampling plan is used, which takes into account the following variables: direction (inward vs outward), mode of transport, border crossing, type of day (weekday/holiday), time slot, day of the week.

[2] As outlined in the Balance of Payments and International Investment Position Manual Sixth Edition (BPM6) of the IMF.
[3] The survey does not collect information on domestic tourism.

The survey takes place at international border points (ports, airports, railway and road crossings) and relies on two main operations: interviews and counts. A subset of counts are "qualified" because they also collect information on the country of residence of the travellers.[4] Until 2019, about one hundred thousand interviews and one million counting operations were carried out annually at more than 60 border points.

The purpose of the qualified counts is to estimate the size of the reference population, namely the total number of incoming non-resident international travellers and outgoing resident travellers, by country of residence or destination, respectively. Counts are supplemented, where available, with administrative data (e.g. information provided by the airport authorities about the number of travellers entering or leaving Italy) provided by the authorities responsible for border points. These data provide the "exact" population size of the reference universe but do not provide any information on the residence of travellers.

The interviews, conducted using the Computer Assisted Personal Interview (CAPI) method, serve to collect data on expenditure and other relevant aspects for compiling BoP items, such as the counterparty country and the purpose of the trip. Secondly, they gather additional information useful for a broader understanding of the tourism phenomenon. The interview takes place at the end of the journey (when leaving Italy for non-residents and when entering Italy for residents), minimizing the traveller's difficulty in recalling expenditure and trip characteristics, as well as ensuring that the expenditure is not hypothetical or planned but already incurred by the respondent. Each interview is administered to a single individual per group of travellers (i.e., those who share travel expenses, such as a family), who also responds on behalf of the other members.

The information acquired with the questionnaires is then grossed up to the reference population as determined from the counting operations, administrative data sources and, as of 2020, MPD.

Indeed, information gathered through MPD since 2020 (see Sections 4 and 5 for the details) has only complemented counting operations, reducing them by about two-thirds, while the number of interviews – and the information they aim to collect – has remained unchanged.

4. The path toward the use of Mobile Phone Data

The integration, of MPD in the process of estimating the "travel" item took place in several stages, starting with the general idea that they could be useful since a SIM card issued in Italy disappearing from the network near a border may indicate a resident travelling abroad and conversely a SIM card issued by a foreign operator appearing near a border may indicate a non-resident entering Italy.

With this idea in mind, a pilot study was conducted in 2018 (Carboni et al., 2023) that saw the collaboration of the Bank of Italy with one of the main Italian telephone operators. Subsequently, during the period when the COVID-19 emergency placed Italy in an almost complete lockdown, the absence of traditional information sources required the use of alternative ones, including MPD, although limited to non-residents detected in Italy (Carboni et al. 2020).

Finally, MPD were officially integrated into the Bank of Italy's tourism statistics at the end of 2020.

4.1. The pilot study and the COVID-19 emergency

The pilot study was initiated in 2018, in order to evaluate the feasibility of permanently integrating MPD into the estimation process for the "travel" item. This stage involved the collaboration between the Bank of Italy and Vodafone, one of the leading Italian mobile network operators.

The pilot study focused on three border points, two airports (Roma Fiumicino and Milano Malpensa) and a highway frontier point (Tarvisio), and concerned flow data, i.e., the number of travellers leaving Italy (non-residents) or returning (residents) through these border points. MPD were compared with time series derived from surveys and, specifically for the two airport checkpoints, with administrative data.

[4] The information on the destination is instead gathered through the interviews.

The results highlighted the informative potential of MPD and their ability to improve the estimation of population size for border points where administrative data are not available. Moreover, for all types of border points, MPD were able to improve the classification by counterpart country of the reference population. Additionally, the data demonstrated high timeliness and the ability to reduce costs by minimizing manual counts, which are relatively expensive.

On the other hand, the challenge to translate the definitions of official tourism statistics into telephony terminology became evident, as well as the consequent need for constant interaction between statisticians and MPD providers and the need to develop specific algorithms depending on the type of border point.

During the period from March to June 2020, the survey was suspended, in compliance with government measures aimed at containing and managing the COVID-19 health emergency and due to the practical impossibility of conducting counts and interviews at borders. International tourist flows for this period were therefore estimated based on the application of y-on-y variations 2020/2019 derived from alternative data sources, such as MPD for foreigners in Italy and payment card data for Italians abroad, to the survey data referred to the corresponding period of 2019.

The number of foreign travellers in Italy was estimated using mobile phone data provided by the same operator involved in the pilot study. The need to quickly obtain data led to the use of stock data on the number of non-residents present in the national territory at a given point in time, rather than flow data at border crossings. Indeed, flow statistics require tracking the behaviour of a SIM card over time, whereas stock statistics are based on data collected at a specific point in time. Specifically, information on the number of foreign SIM cards detected in the Italian territory at least once a month was used, broken down by nationality of the issuing operator, as a proxy for the traveller's residence.

The quality of the estimates thus obtained was subsequently verified through comparisons with other available information sources, such as administrative data, other official statistics, and mirror data.[5] These comparisons confirmed the reliability of the method used.

4.2. Adoption of MPD and expected benefits

The experience gained with MPD during the experimentation phase and the subsequent emergency phase led to the final integration of this source into official tourism estimation procedures as of the end of 2020. The integration occurred gradually: at first, MPD were only used as a reference benchmark, then as a substitute for a part of the counting operations.

The main benefits expected from the adoption of MPDs can be summarized in the following points:

- Airport crossing points: all international airports have administrative data and, consequently, the size of the reference universe can be considered as known. However, as stated above, these administrative data do not provide any information on the residence of travellers, but only about the number of travellers who travelled on international flights.
- The use of MPD should lead to a better qualification of the universe, especially for large airports characterized by the presence of numerous gates, some of which may be very distant from each other, making monitoring through counting operations very complex.
- Road passes: In Italy there are dozens of road passes connecting it with neighbouring countries often located in mountainous areas, far from major population centres and, therefore, difficult to reach. Inevitably, unless a significant increase in the cost of the survey is accepted, counting operations must be limited to a sample of crossings, those that are more relevant in terms of traffic flows. Moreover, in terms of cost sustainability, counting operations are conducted only for a certain number of days

[5] "Mirror data" refer to the estimation of the same item provided by a different BoP compiler. Indeed, travel expenditures from country A to country B are recorded by both economies: as debits for A and as credits for B. Mirror data play a crucial role as they allow verifying, validating, and improving the reliability, accuracy and consistency of international data.

each month and for a daily time span usually ranging from 6 to 8 hours. Finally, it should be noted that counting operations on roads, in the absence of checkpoints, involve counting vehicles and the number of occupants, using the license plate as a proxy for residency status. In some situations, counting operations may be challenging as, due to safety concerns, they cannot take place near the road where vehicles are passing. Additionally, information about road checkpoints is largely missing from administrative data sources.

- On the contrary, MPD can potentially cover all road checkpoints and monitor them continuously. Therefore, an advantage is expected in terms of a estimation of both the universe size and the traveller's residency distribution.
- Railroad crossing points: counting operations are conducted directly on sampled trains, for which administrative data are available, and, as a rule, they manage to cover all passengers on the train. Therefore, in this case the expected advantage of MPD lies mainly in the fact that telephone data enable to monitor all trains and not just a sample of them.
- Port crossing points: counting operations are conducted in boarding and disembarkation areas, which are subject to access restrictions for security reasons. Additionally, it is often prohibited to approach docking areas during disembarkation phases, as they must occur as quickly as possible.
- MPD can monitor ports without the need to access restricted areas and with equal effectiveness during both the boarding and disembarkation phases. Considering that only a minority of ports provide administrative data with timing useful for BoP use, MPD should lead to an improvement in both the estimation and qualification of the universe.

5. The Use of Mobile Phone Data by the Bank of Italy for Compiling the "Travel" Item of the Balance of Payments

5.1. The MPD used and the main characteristics of the MNO

MPD provided by Vodafone, the same MNO that supplied data during the experimental phase and the pandemic period, have been integrated into the estimation process since the end of 2020.

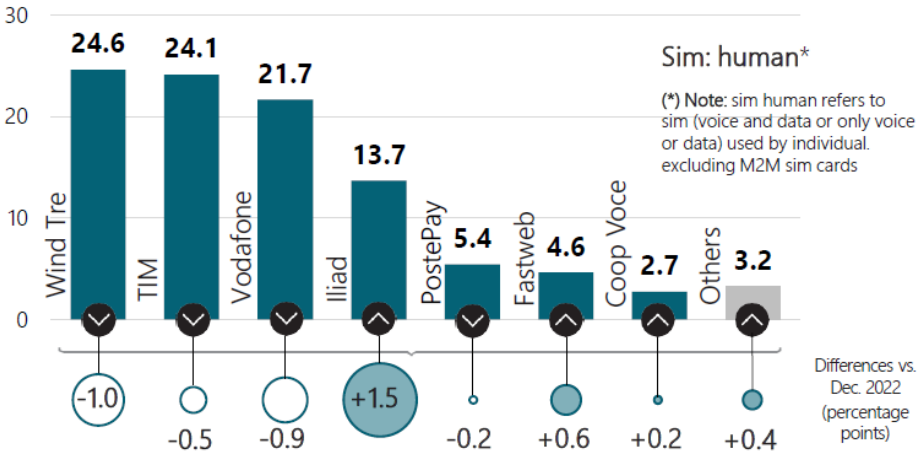
In September 2023, the Italian market share of total SIM cards issued by this MNO was 27.5% (Figure 1). At that time, the company had more than 200,000 cell towers and covered more than 99% of the Italian territory. Since the operator extrapolates the data about SIMs detected on its network to the target population (total SIMs detected on Italian mobile networks) based on its market share, this factor is crucial to ensure sufficient representativeness and adequate expansion of the findings to the entire universe. In the case of non-residents, the proportion of users connecting to the specific MNO is determined by considering the likelihood of their devices connecting to the network of the mobile carrier in the specific cell near the border. This probability also depends on the presence of competing operators' transmitters in the area, as well as on existing commercial agreements between mobile carriers across different countries.

The information provided by the MNO leverages on the data traffic that each SIM, limited to the human-type[6] ones, continuously maintains with the operator's network. Billions of positions are collected daily, with a frequency of one minute and a spatial resolution ranging from a few hundred meters in urban areas to several kilometres in rural areas.

The data are compliant with the Italian privacy law. With specific reference to the data transmitted to the Bank of Italy, they are irreversibly anonymized and aggregated into subsets that cannot be formed by less than 15 SIM cards. If this is not the case, the information is not provided.

[6] Human-type SIM cards are associated with physical persons, while machine-to-machine (M2M) SIM cards are used for communication between devices.

Figure 1 - Total SIMs: market shares of the main Italian MNO.



Source: the Italian Regulatory Authority for Communications (AGCOM).

5.2. The structure of MPD: a comparison with traditional counting operations

The mobile phone data were requested from the MNO in such a way as to approximate the structure of the information obtained from counting operations, which MPD are meant to complement.

Data derived from counting operations differ for each type of border crossing (road, rail, etc.); this is because different information is collected depending on the mode of transport used, such as the type of vehicle boarded on a ship, the class of a train carriage, the type of flight (scheduled or charter). All this detailed information cannot be captured through mobile telephony. The use of MPD thus results in an objective loss of secondary information, which is counterbalanced, as will be discussed later, by a greater reliability of the most relevant information.

Other advantages of MPD relate to the much higher frequency at which data are collected, as well as to the wider spatial coverage. Indeed, counting operations are exclusively conducted at the border points sampled in the survey, based on a monthly calendar which includes a certain number of counting shifts at each crossing point for a certain number of hours. In this respect, MPD are much more efficient, as they can continuously cover not only each sampled crossing point, but potentially all of them (although smaller crossings are typically grouped in clusters based on type and geographic location). As a second example, regarding the determination of the traveller's residency, counting operations only sample these data on a subset of counted individuals.[7] In contrast, MPD considers the nationality of the SIM card issuer for every SIMs detected on the network.

5.3. The impact of MPD on counting operations

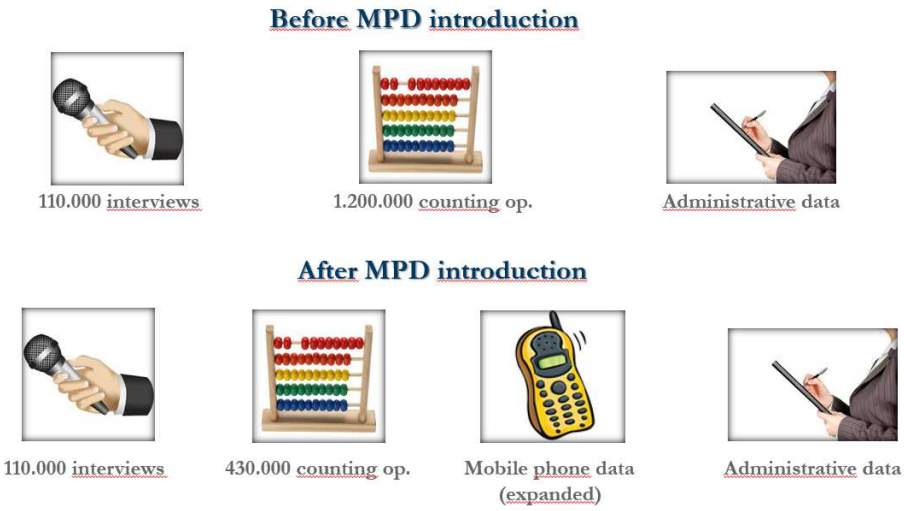
Mobile phone data have been integrated into the estimation process, replacing part of the counting operations conducted at border crossings. Consequently, the counting operations have decreased from 1,200,000 to about 430,000 per year (Figure 2), thus allowing for a significant reduction in the costs associated with the survey.

It is expected that, with the continuous improvement of methodologies for the exploitation of MPD, the need for counting operations will further decrease over time. However, a certain number of counts will always be necessary, mainly as a control tool of mobile data, especially for frontier points without administrative data sources.

[7] For example, at a road crossing with a count step of 5, the nationality of the vehicle's license plate will be recorded once for every five vehicles that pass through. The counting step varies, based on traveller flow intensity and the specific crossing point characteristics.

Counting operations are also important for gathering evidence on specific secondary phenomena, which are difficult or impossible to capture with MPD, such as travellers from smaller countries without a resident MNO (San Marino, Monaco, Andorra, etc.); or to gather additional information not detectable through MPD, such as the type of vehicle used at road border crossings.

Figure 2 - Sources for compiling the "Travel" item of the BoP, before and after the introduction of MPD in the estimation process.



5.4. How to make the most of mobile phone data

In order to use telephony data for tourism statistics, it is necessary for the MNO to isolate the SIMs that contribute to aggregates that correspond as closely as possible to the definitions of official statistics.

Listed below are the main points that were worked on, in collaboration with the telephone company, in order to limit the possibility of incorrect selection of relevant SIMs.

5.5. Identification of the destination country for Italians and of the origin country for foreigners

In BoP statistics identifying the counterpart country is fundamental. In tourism statistics, this means determining the foreign country (or countries) which Italian residents visited, and the country of residence of non-resident travellers who visited Italy.

To determine the destination country of an Italian resident, the "Welcome message" received by each SIM when connecting to a foreign operator's network is used. If multiple countries are visited, the number of welcome messages corresponds to the number of countries visited. The date of receipt of these messages could also be used, in principle, to identify the period of stay in each of the visited countries. The quality of this information is still being explored.

For the country of origin of non-resident inbound tourists, the residence of the SIM card issuer is used as a proxy. This approximation is of good quality in most cases. However, certain specific phenomena can be systematically distorting. A typical case is that of individuals originally from other

countries who are resident[8] in Italy and use a SIM card issued in their home country. In Italy, there are numerous and large communities of foreign residents, so this phenomenon cannot be overlooked. To correct this type of false positive, an analysis of the SIM card's behaviour over several months is used: if a SIM card issued abroad is detected on the Italian network for more than six out of the twelve months preceding the one under examination, then the SIM card is deemed to be associated with a resident in Italy.

Regarding the use by foreign residents of SIMs issued in Italy, no specific filters have been introduced at this stage.

5.6. Estimating Overnight and Same-Day Trips

The BoP records the monetary flows between two economies, therefore in travel estimation it is crucial to distinguish between overnight and same-day travellers, as their spending profiles are dramatically different. For instance, in 2023, same-day travellers represented around one third of non-resident travellers visiting Italy but only 4% of the total expenditure. With appropriate measures, MPDs can identify the two categories of travellers.

The general criterion adopted to associate a SIM issued by a non-resident MNO with an overnight trip to Italy is that at least 24 hours must elapse between the time of entry into the Italian territory and the time of exit. However, there are cases where a traveller can stay in Italy for less than 24 hours but still spend a night in the country. We detect these situations by requiring that the traveller spend at least 6 hours during the night in a single location in Italy ("6-hour-detection criterion").

For SIMs issued by resident MNOs, a loss of signal for at least 24 hours is used to identify overnight travellers abroad, while the 6-hour-detection criterion is not applied. This is because, unlike non-resident SIMs, for which a journey is identified based on connection to the network, for resident SIMs, the journey is defined by disconnection from the network. The methodology is continuously monitored, and further adjustments may be considered in the future.

5.7. Determining the Province Visited or of Residence

Determining the province visited (for inflows) or of residence (for outflows) is not relevant for the compilation of BoP statistics, but it is of considerable interest for analysing tourism-related phenomena.

To determine the province of residence associated with a (resident) SIM user, the following procedure is followed: for each day, the municipality where the user spent the most time between 8 p.m. of the previous day and 8 a.m. is defined as the "prevalent night-time location". The "residence" is then defined as the municipality that most frequently in the previous twelve months (or in the number of months available in the case of newly activated SIMs) has been the "prevalent night-time location". Although, at the level of MPD, information is collected at a more detailed territorial level than the province, data are then aggregated to this last dimension for two reasons: a) to contain the otherwise high number of masked records which the MNO cannot provide due to confidentiality reasons;[9] b) the impossibility, on the part of the MNO's network, to reconstruct the administrative subdivision into municipalities with precision, particularly for smaller municipalities.

As for determining the province visited by non-resident travellers, the "prevalent province" is defined as the one where the travellers spent the most time during their trip to Italy. Due to confidentiality reasons, we do not pursue the full list of visited provinces, since this would considerably increase the number of "masked" data that do not meet the requirements of the Italian privacy regulation.

[8] It is important to remember that the concept of residence used here is economic (the centre of predominant economic interest) rather than administrative, as outlined in the Balance of Payments and International Investment Position Manual Sixth Edition (BPM6) of the IMF.

[9] It is recalled that, in compliance with privacy regulations, it is not possible to provide information on phenomena for which, due to the attributes considered (e.g., date, municipality visited, nationality associated with the SIM, etc.) the number of SIMs detected is smaller than 15.

5.8. The roaming effect near road borders

A relevant problem concerns data roaming near land borders, where an Italian SIM could connect to a foreign network, or vice versa, without an actual border crossing, thereby generating a false positive. This phenomenon can occur multiple times within a single day, with frequent switching between resident and non-resident operators. This type of error must be minimized as much as possible because it can lead to significant distortions in estimates, particularly for same-day travellers.

In order to avoid this type of error, specific filters were incorporated into the MNO's algorithms for monitoring land borders. These filters were developed through a trial and error process involving close collaboration with the MNO. Given their importance, these filters continue to be the subject of ongoing study and fine-tuning interventions.

A first filter concerns the minimum duration for which a signal must be acquired (foreigners in Italy) or lost (Italians abroad) on the MNO's network. This minimum duration has been set to 30 minutes. Only if the signal exceeds this threshold, the SIM movement is considered indicative of an actual border crossing.

The filter shows good effectiveness in the vast majority of cases, but further analysis is required for borders located close to urban centres. In such cases, the threshold could exclude international travellers that spend less than 30 minutes in Italy (or abroad); on the other hand, going below the threshold of 30 minutes leads to a very large increase in the number of SIMs under consideration, with the risk of including a good number of false positives. To address this issue, the possibility of introducing time filters differentiated according to the distance of the borders from residential areas is being considered, as well as using external source benchmarks, such as mirror data. It should be noted that while this phenomenon may impact statistics related to the number of same-day trips, it has minimal effect on BoP statistics, as the spending profile associated with travellers on short or very short stays is very small.

A second filter concerns the daily frequency with which a SIM connects to or disconnects from the MNO's network. If the frequency is high, it is assumed that the SIM is providing wrong signals due to prolonged presence in areas near the border. The frequency filter prevails over the temporal one. Therefore, if a foreign SIM connects to the Italian MNO's network for over 30 consecutive minutes but for a large number of times on the same day, it is discarded as a false positive. Similarly, this applies to Italian SIMs whose signals are lost.

Finally, a third filter is defined by tracking the movements of SIMs through virtual "corridors" of adjacent telephone cells, defined by the operator for each border crossing (for instance, along the route of a motorway leading to a border crossing). This filter can concern both a foreign SIM connecting to the Italian network and an Italian SIM disconnecting from it. If the SIM follows these corridors, connecting progressively to cells further from the borders for foreigners or closer for Italians, it is considered a true positive, with a prevalence criterion over the two previous filters. While this filter is particularly effective, it is not reliable when a foreign SIM crosses the border and remains nearby, or when an Italian SIM starts its movement close to the border.

5.9. Identifying arrivals and departures at airports

Airports can represent very complex realities, covering vast areas where thousands of people move, only a portion of whom are travellers, and among them, only a fraction are international travellers. It was therefore necessary to define, in close collaboration with the MNO, a number of criteria in order to identify arrivals and departures related to international flights and to exclude all other numerous scenarios (domestic travellers, airport personnel, transportation personnel to and from the airport, people passing near the airport without accessing it, etc.).

The first step was to define two main detection areas: the airport itself and the surrounding area. This allows the recording of whether a SIM has been detected inside the airport (thus excluding nearby transits) and whether this event was preceded or followed by movement in the areas adjacent to the border crossing. This reasoning applies to both resident and non-resident international

travellers, whether they are arriving or leaving Italy. A set of consecutive filters allows each SIM to be assigned a specific status, as shown in the scheme below.

- A SIM was detected in the area adjacent to the airport, then inside the airport, and finally disappeared from the network:
    - If the SIM is not detected again in the following hours, it is associated with an international departing traveller.
    - If the SIM is detected again in the following hours (at the same or another national airport), the data is discarded.
- A SIM not present on the network "appeared" inside the airport and then was detected in the adjacent area: the SIM card is associated with an incoming international traveller.
- A SIM not present on the network "appeared" inside the airport and then disappeared again from the network: the SIM card is associated with an international traveller in transit.

The described methodology, progressively refined over time, is able to provide estimates on the number of travellers for each airport that deviate only slightly from the official administrative data. The error decreases as the passenger flow increases. More detailed data on this aspect will be provided below.

The possibility of associating the described SIM behaviours with specific international flights, based on the flight schedules of each airport, has also being studied. The evidence collected so far indicates that this methodology appears effective for smaller airports, where flight arrivals and departures are sufficiently spaced out. In larger airports, where arrivals and departures often occur almost simultaneously, the methodology proves to be significantly less effective, making it practically unusable.

5.10. Purchase of SIM cards issued in the visited country

It can happen that an international traveller purchases a SIM card issued in the country he/she is visiting upon arrival. In the past, this behaviour was widespread, as the additional costs of international roaming services were high. Over time, international agreements have been developed among MNOs to reduce these additional costs and, starting from June 2017, the citizens of the European Economic Area[10] (EEA) can utilize roaming services within this area without incurring additional costs compared to their operator's tariff plans[11]. This has effectively mitigated the phenomenon of purchasing SIM cards issued in the visited territory for movements within the EEA. For travels between the EEA and other countries, the use of a SIM card issued by an operator in the traveller's home country may still incur high additional costs. To address this issue, one can use SIM cards issued in the visited country, typically in the form of prepaid SIM cards or eSIMs, which are virtual SIMs activated on one's mobile device.

From the perspective of tourism statistics based on MPD, this primarily leads to a false negative, as a traveller from another country is recognized as a resident. Issues can also arise regarding the identification of the country visited by resident travellers, in cases where the SIM from the resident MNO is never activated in the foreign country, thereby making it impossible to receive the "Welcome message".

[10] European Union countries, Great Britain, Norway, Iceland, Lichtenstein.

[11] It is possible for operators to offer their customers alternative roaming rates, even differentiated by country. However, their application requires explicit authorization from the customer.

In Italy, over 80 percent of travels, both with and without overnight stays, originate from European countries; similarly, over 80 percent of travels abroad by residents in Italy are directed towards this area. Therefore, the concerns described above are relatively minor. Efforts to address them are under way, but results are still too preliminary to be disclosed.

6. Quality Assessment of MPD

The quality control of data provided by the MNO can only be achieved through comparison with other data sources. Among these, administrative data play a central role since, as previously discussed, they are considered as accurate information. In the absence of administrative data, those derived from the survey are used.

In cases where MPD for a specific border crossing are not consistent with administrative data, they are considered incorrect, and a revision process of the algorithms, agreed between the MNO and the Bank of Italy, is initiated. The type of the detected differences guides the search for corrective actions to be implemented. This process is repeated until the estimates derived from the MPD show a satisfactory convergence with administrative data. For this purpose, a deviation of the estimate of less than 10 per cent is considered acceptable. This constraint is relaxed for crossings with very low traveller flows, where the percentage deviation has little informational value, and the comparability of levels is more useful. Once accurate estimates are obtained for the revised crossing, they are closely monitored for several months to ensure that the convergence of the estimates towards the administrative values is genuine and not due to random effects.

In the absence of administrative data, the comparison is made with the number of travellers estimated using counting operations from the survey. For each border crossing, MPD are considered of good quality if they approximately replicate the volumes and residence distributions of the survey data, without introducing predefined quantitative thresholds. If sufficient approximation is not achieved, in the absence of additional informative sources to support the analysis (such as mirror data), specific experiments can be conducted at the concerned border crossings. These experiments involve a counting session at the crossing that covers most of a single day (approximately 12-18 hours). The results of such experiments are considered accurate and are then compared with the MPD collected at the same crossing during the same time frame. This approach helps to determine whether a revision of the MPD production algorithms is necessary or if the MPD have indeed improved the previous estimates based on counting. In the few cases where such in-depth studies have been required, the MPD have demonstrated superior performance.

An additional quality check for MPD is conducted by examining data over a longer time frame than a single month. Over such a period, it is reasonable to expect that the number of incoming foreign SIMs will roughly equal the number of foreign SIMs departing from Italy (a similar and reciprocal expectation applies to SIMs issued in Italy). This is based on the assumption that a non-resident visiting Italy will leave within a few days, weeks, or, in rarer cases, months. The comparison should be made across the entirety of the flows, as the entry border crossing may differ from the exit one. Nevertheless, even control at the level of individual border crossings has its validity, as significant imbalances between incoming and outgoing flows could indicate potential issues. This examination must be conducted with care, as such imbalances could also be an inherent characteristic of the specific border crossing[12].

6.1. MPD Quality in 2023

The previous section outlined the main techniques employed in order to mitigate errors and assess the quality of estimates derived from the use of MPD. This section shows the results of the comparison between Vodafone data and administrative data on the sampled airports for the year 2023 (as provided by Assaeroporti, the Italian association of airport operators; see Annex I).

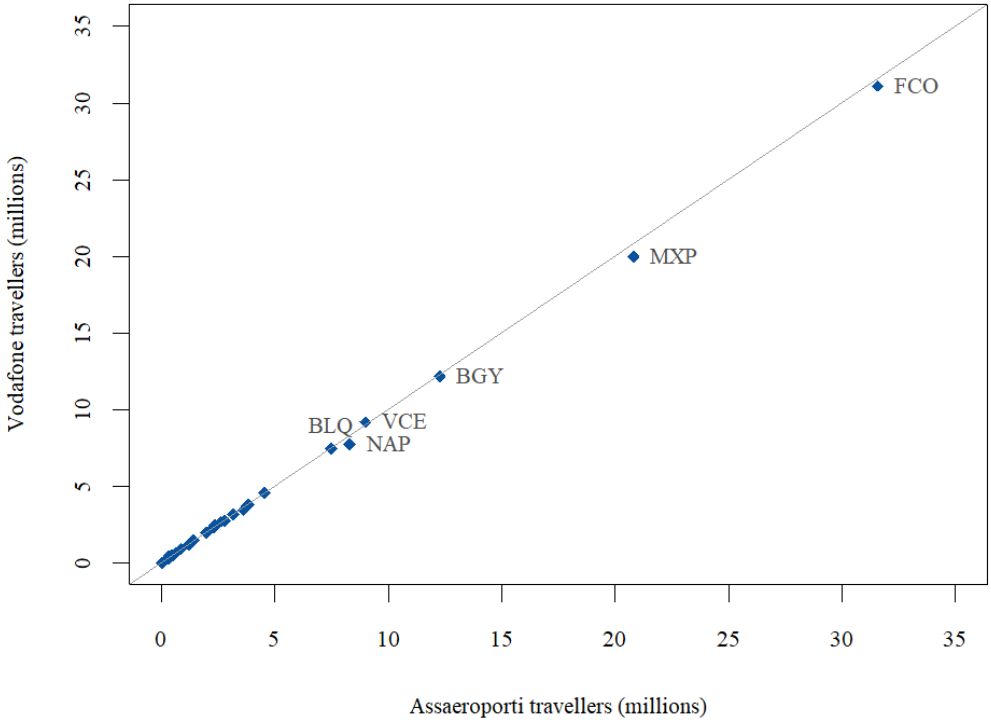
[12] For example, let's consider a port border crossing that serves as the final destination for cruises; travellers might then use an airport border crossing to return to their country of residence.

Administrative data provide information on the number of travellers on international flights, both inbound and outbound, without the country of residence qualification. The comparable telephony aggregate includes the set of SIM cards detected entering and exiting the airports, obtained by adding both those issued in Italy and those issued abroad. These aggregates differ from that used in the BoP statistics, where only inbound SIMs of resident MNOs and outbound SIMs of non-resident MNOs are recorded.

Airport data are generally available 20-25 days after the reference month, while the Vodafone estimate is delivered to the Bank two weeks after the reference month, when administrative data are not yet available. This makes the ex-post verification of the MNO estimates significantly probative.

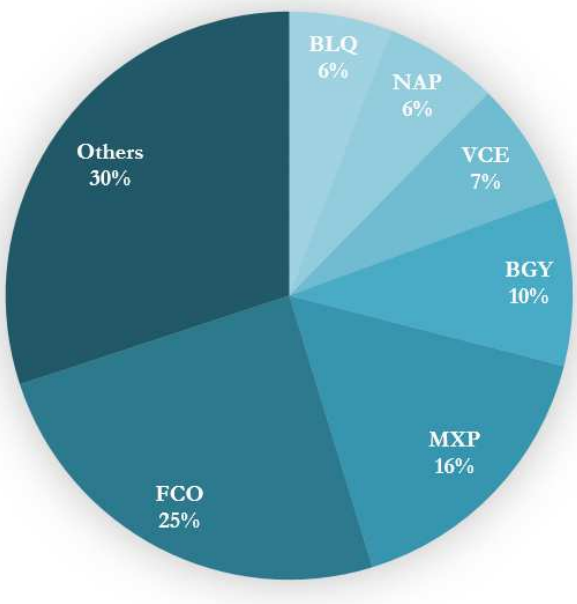
The number of international travellers recorded for single airports by each source appears to be very similar (Figure 3**Errore. L'autoriferimento non è valido per un segnalibro.**). The two main airports, Rome Fiumicino (FCO) and Milan Malpensa (MXP), stand out from the rest of the distribution, together serving over 40 percent of international travellers (Figure 4). In addition, an intermediate cluster of airports (Bologna, Naples, Bergamo, Venice) with a significant volume of international travellers (between 7,4 and 12,4 million) can be identified, while the remaining 23 airports examined here record a flow of international travellers of less than 5 million.

Figure 3 - Comparison of Vodafone and Assaeroporti data for all sampled airports (2023).



Source: MPD and Assaeroporti; refer to Appendix for airport codes.

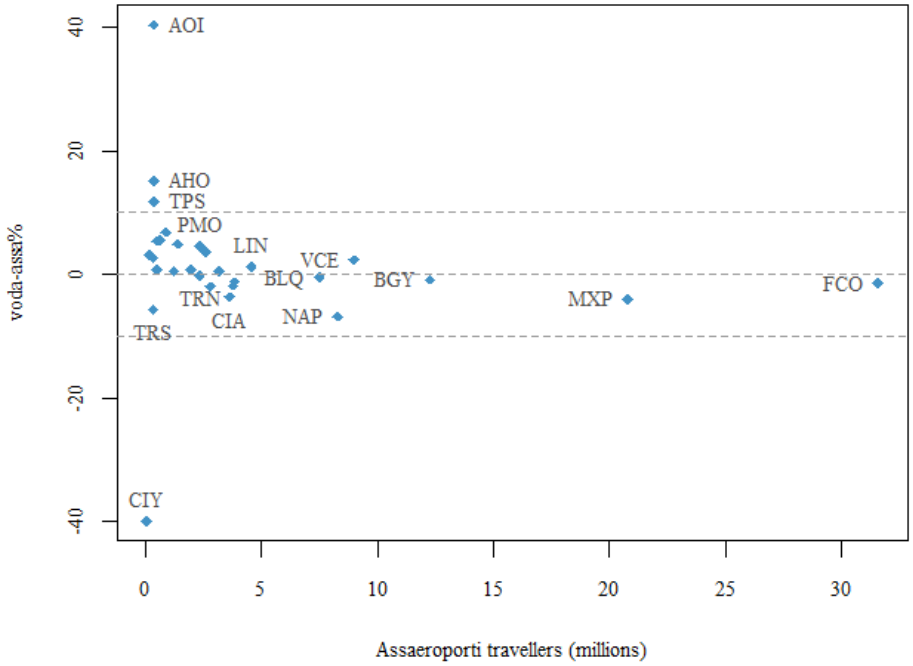
Figure 4 - Composition of international travellers visiting Italy across the sampled airports 2023.



Source: Assaeroporti; refer to Appendix for airport codes.

Deviations between the two sources in percentage terms (Figure 5) are almost always below 10 percent. The few cases where the deviation exceeds this threshold pertain to smaller airports with a very limited flow of international travellers, for which, as already stated, percentage-based analysis loses its informative capacity.

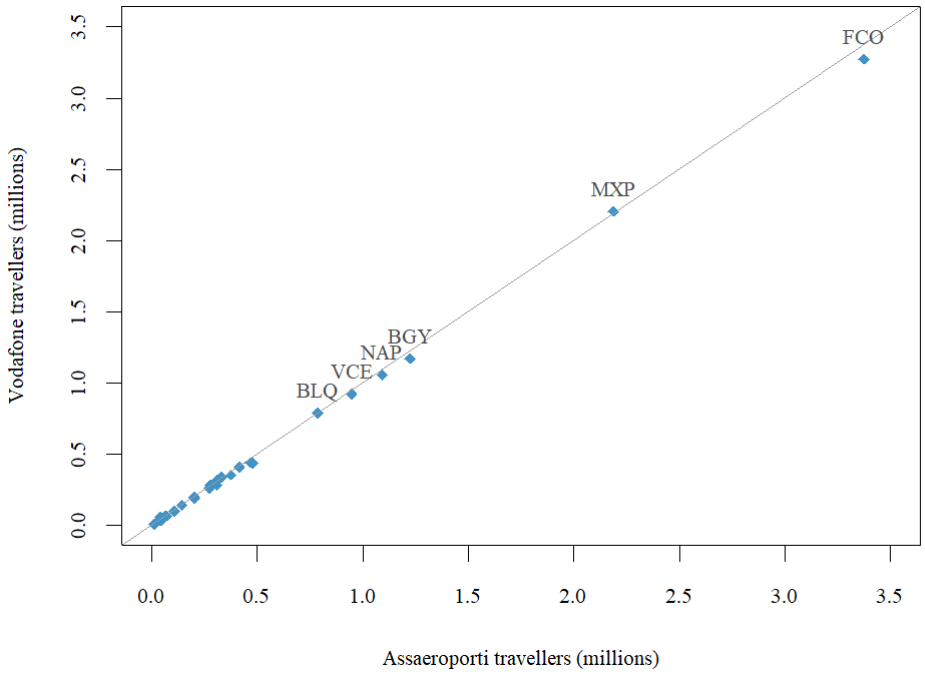
Figure 5 - Percentage differences between Assaeroporti and Vodafone data (2023).



Source: MPD and Assaeroporti; refer to Appendix for airport codes.

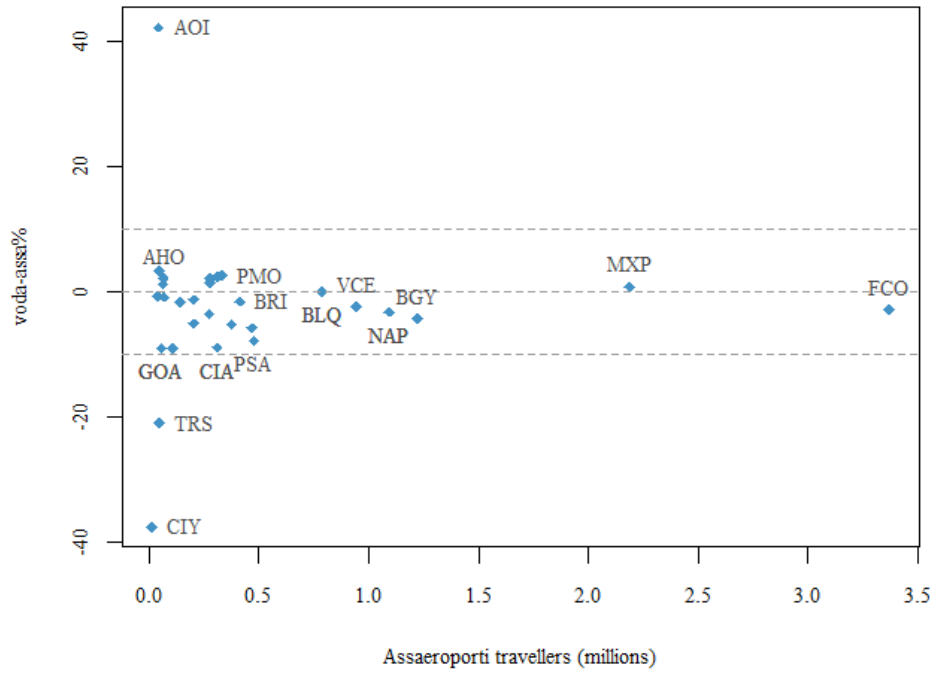
The quality of the estimates is confirmed even at a lower level of data aggregation. Indeed, if we examine the data for August 2023, the month with the largest flows in that year, we obtain results (Figure 6) not dissimilar to those at the annual level shown in Figure 3: the discrepancy between the MPD and the administrative source is almost always less than 10 percent (Figure 7).

Figure 6 - Comparison between Vodafone and Assaeroporti data for all sampled airports (August 2023).



Source: MPD and Assaeroporti.

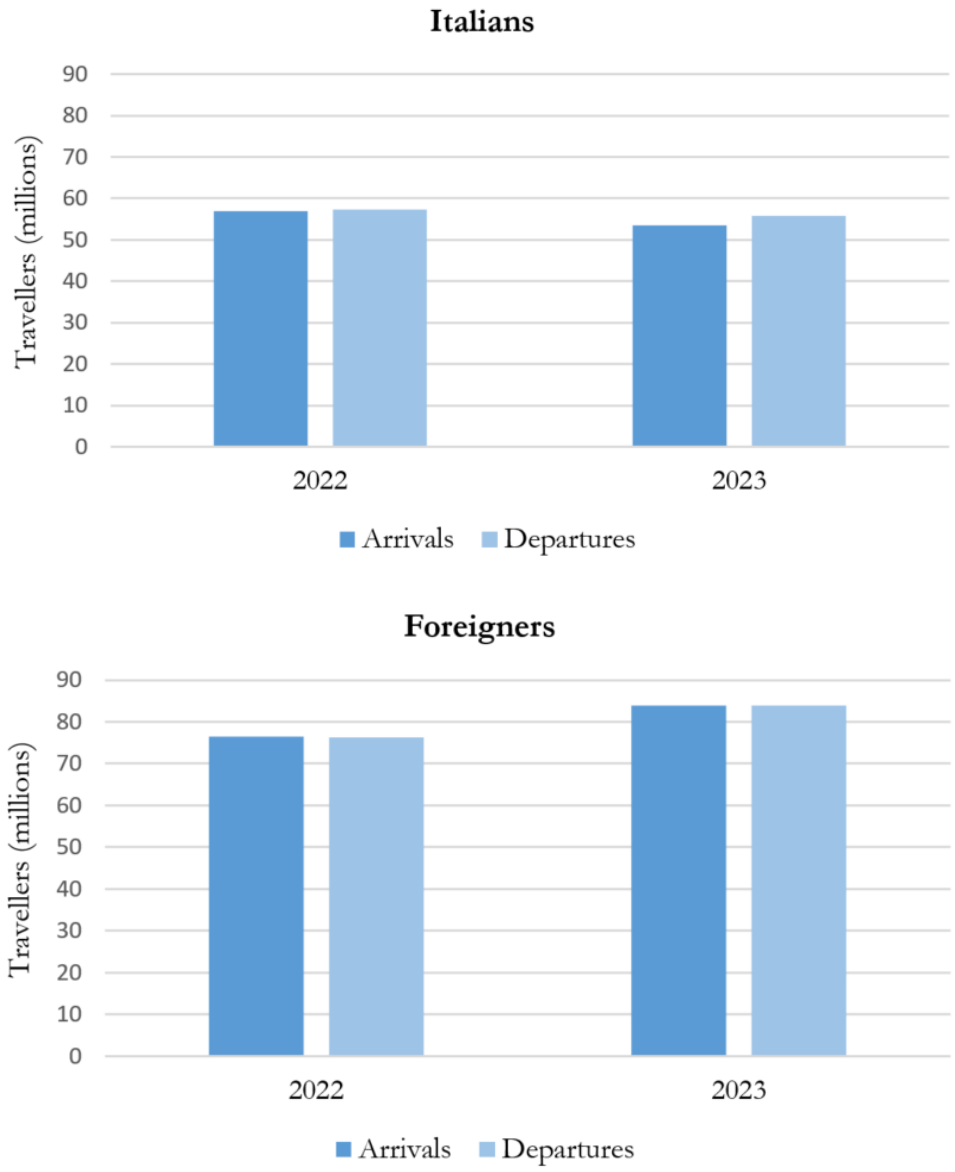
Figure 7 - Percentage differences between Assaeroporti and Vodafone data (August 2023).



Source: MPD and Assaeroporti

We now go back to the consideration that, in a sufficiently long run, the number of incoming foreign SIMs will roughly equal the number of foreign SIMs departing from Italy (a similar and reciprocal expectation applies to SIMs issued in Italy). From the comparison of international arrivals and departures detected by MPD, separately for Italians and foreigners, a good similarity between the total annual values is observed for both 2022 and 2023 (Figure 8), in particular, for foreigners.

Figure 8. Comparison between the number of arrivals and departures (millions) for Italians and foreigners in 2022 and 2023.



Source: MPD.

To summarize, our analysis shows that the estimates derived from MPD closely approximate administrative data, even at the level of individual month and individual airport, with lower performance only for airports with very limited flows of international travellers. Furthermore, MPD exhibit good internal coherence, as shown by the full comparability of arrivals and subsequent departures.

7. Conclusions

Tourism statistics are notoriously complex, and their measurement and definition pose a constant challenge. Traditionally, tourism studies have relied on specifically designed and implemented sample surveys to collect pertinent data. Since 1996, the Bank of Italy has been conducting the Survey on International Tourism, a valuable tool for gathering essential information for the compilation of the “Travel” item of the balance of payments. The Survey is based on two main pillars: interviews to gather information about expenses and counting operations, which is integrated by administrative data, to estimate the size of the reference universe.

Despite the invaluable nature of this survey, it presents some significant drawbacks, including high costs, the objective impossibility of sampling all borders at a high frequency, and the difficulties related to border crossings that are hard to reach (mainly, border points located in mountain areas far from populated centres) or lack administrative data sources (essentially all road borders and some seaports).

The removal of many checkpoints, following the introduction of the Schengen agreements, has made the data collection process through interviews and counts even more complex. The global health emergency caused by the COVID-19 pandemic further highlighted the limitations of this approach, exemplifying the impossibility of conducting the survey during national emergencies.

In this context, the Bank of Italy started using new, less expensive, timelier and resilient information sources compared to traditional ones. After a period of experimentation, MPD have been integrated into the BoP estimation process since 2020, replacing part of the counting operations. This has led to significant cost savings. We have shown that MPD align well with available administrative sources and show internal consistency, thus being of good quality. Additionally, one of the main features of MPD is their ability to continuously monitor the entire territory, providing more accurate data for complex border crossings or those lacking administrative sources.

However, the use of MPD carries significant challenges, especially in translating data originally designed for telephony into tourism statistics aligned with official definitions. This process requires constant coordination with the MNO to refine estimation algorithms and filtering criteria for raw data. It is essential to precisely define all metrics, filters, adjustments, and methodologies for the correct extraction of information. Without this level of collaboration and coordination, the accuracy and reliability of the information collected through MPD could be compromised. It is probably for this reason that only a few countries have yet integrated such data into the official statistics production process.

In conclusion, the integration of MPD into the production of official statistics is a relatively new activity that is already producing satisfactory results but still has margin for improvement. The effective use of MPD requires significant commitment in terms of collaboration with MNOs and methodological development. However, the benefits achieved justify this effort. Tourism statistics, thanks to the integration of MPD, can become much more precise. Furthermore, technological and methodological progress will likely overcome many of the current limitations of MPD, making them an increasingly indispensable tool for the production of official statistics.

References

1. R. Ahas, A. Aasa, A. Roose, Ü. Mark and S. Silm, 'Evaluating passive mobile positioning data for tourism surveys: An Estonian case study', *Tourism Management*, 29(3), 2008, pp. 469-486.
2. R. Ahas, A. Aasa, Ü. Mark, T. Pae, A. Kull, 'Seasonal tourism spaces in Estonia: Case study with mobile positioning data', *Tourism Management*, 28(3), 2007, pp. 898-910, ISSN 0261-5177.
3. R. Ahas, S. Silm, O. Järv, E. Saluveer, and M. Tiru, 'Using Mobile Positioning Data to Model Locations Meaningful to Users of Mobile Phones', *Journal of Urban Technology*, 17, 2010, pp. 3-27.
4. P. Bonnel, E. Hombourger, A. Olteanu-Raimond, Z. Smoreda, 'Passive Mobile Phone Dataset to Construct Origin-destination Matrix: Potentials and Limitations', Transportation *Research Procedia*, 11, 2015, pp. 381-398, ISSN 2352-1465.
5. N. Caceres, L. M. Romero, F. G. Benitez, 'Exploring strengths and weaknesses of mobility inference from mobile phone data vs. travel surveys', *Transportmetrica A: Transport Science*, 16:3, 2020, pp. 574-601-
6. A. Carboni, C. Doria, C. Catalano, 'How can Big Data improve the quality of tourism statistics? The Bank of Italy's experience in compiling the "travel" item in the Balance of Payment', *Questioni di Economia e Finanza (Occasional Papers)*, Bank of Italy, 761, 2023.
7. A. Carboni, C. Doria, S. Zappa, 'La produzione statistica nell'emergenza COVID19: la stima dei "viaggi" in bilancia dei pagamenti', *Note Covid-19*, Bank of Italy, 19th June 2020.
8. Q. Chen, Z. Hu, H. Su, X. Tang and K. Yu, 'Understanding Travel Patterns of Tourists from Mobile Phone Data: A Case Study in Hainan', *IEEE International Conference on Big Data and Smart Computing, 2018*
9. Q. Chen, L. Weifeng, D. Zhengyu, Y. Dongyuan, R. Bin, 'Using mobile phone data to determine spatial correlations between tourism facilities', *Journal of Transport Geography*, 92, 2021, 103018, ISSN 0966-6923, https://doi.org/10.1016/j.jtrangeo.2021.103018.
10. G. Cousin, F. Hillaireau, 'Can Mobile Phone Data Improve the Measurement of International Tourism in France?', *Economie et Statistique / Economics and Statistics, Institut National de la Statistique et des Etudes Economiques (INSEE)*, (505-506), 2018, pp. 89-107.
11. C. Demunter, 'Tourism Statistics: Early Adopters Of Big Data?', *Sixth UNWTO International Conference on Tourism Statistics*, 2017.
12. Eesti Pank (2019). International travel statistics. Retrieved 12 June 2019, from http://statistika.eestipank.ee/#/en/p/MAKSEBIL_JA_INVPOS/1410.
13. Eurostat. (2014) Feasibility Study on the Use of Mobile Positioning Data for Tourism Statistics: Consolidated Report. LU: Publications Office.
14. L. Grassini, G. Dugheri, 'Mobile phone data and tourism statistics: a broken promise?' *Annals of Mathematics and Artificial Intelligence*, 3(1), 2021, pp. 50–68.
15. T. Kubo, S. Uryu, H. Yamano, T. Tsuge, T. Yamakita, Y. Shirayama, 'Mobile phone network data reveal nationwide economic value of coastal tourism under climate change', *Tourism Management*, 77, 2020, 104010, ISSN 0261-5177, https://doi.org/10.1016/j.tourman.2019.104010.
16. Y. Leng, A. Noriega, A.S. Pentland, I. Winder, N. Lutz and L. Alonso, 'Analysis of Tourism Dynamics and Special Events through Mobile Phone Metadata', Proceedings of Data for Good Exchange (D4GX) 2016. New York, NY
17. T. Lestari, R. Rufiadi, E. Saluveer, S. Esko, S. Dimulyo, 'Indonesia's Experience of using Signaling Mobile Positioning Data for Official Tourism Statistics', 2018.
18. I. Noviyanti, et al. 'Towards Big Data as Official Statistics: Case Study of the Use of Mobile Positioning Data to Delineate Metropolitan Areas in Indonesia', 2020: 943 – 954.

19. S. Park, Y. Xu, L. Jiang, Z. Chen, S. Huang, 'Spatial structures of tourism destinations: A trajectory data mining approach leveraging mobile big data', *Annals of Tourism Research*, 84,2020,102973,ISSN 0160-7383 ,https://doi.org/10.1016/j.annals.2020.102973.
20. S. Phithakkitnukoon, T. Horanont, A. Witayangkurn, R. Siri, Y. Sekimoto, R. Shibasaki, 'Understanding tourist behavior using large-scale mobile sensing approach: A case study of mobile phone users in Japan', *Pervasive and Mobile Computing*, 18, 2015, pp. 18-39,ISSN 1574-1192, https://doi.org/10.1016/j.pmcj.2014.07.003.
21. E. Rahmadian, Y. Virantina, D. Feitosa, A. Zwitter, A., T. K. Lestari, 'Big Data Governance: The Case of Mobile Positioning Data for Official Tourism Statistics In Indonesia', In *The 64th World Statistics Congress,* 2023.
22. J. Raun, R. Ahas, M. Tiru, 'Measuring tourism destinations using mobile tracking data', *Tourism Management*, Volume 57, 2016, pp. 202-212, ISSN 0261-5177, https://doi.org/10.1016/j.tourman.2016.06.006 .
23. E. Saluveer, J. Raun, M. Tiru, L. Altin, J. Kroon, T. Snitsarenko, A. Aasa, S. Silm, 'Methodological framework for producing national tourism statistics from mobile positioning data', *Annals of Tourism Research*, 81, 2020,102895, ISSN 0160-7383, https://doi.org/10.1016/j.annals.2020.102895.
24. UN.ESCAP, 'Using big data for official statistics: key considerations when using mobile phone data', 2022. Retrieved from: https://hdl.handle.net/20.500.12870/5177.
25. UN, 'Handbook on the Use of Mobile Phone Data for Official Statistics. Global Working Group on Big Data for Official Statistic', 2019. Available from: https://unstats.un.org/bigdata/task-teams/mobile-phone/
26. Z. Wang, S. Y. He, Y. Leung, 'Applying mobile phone data to travel behaviour research: A literature review', *Travel Behaviour and Society*, 11, 2018, pp. 141-155, ISSN 2214-367X.
27. S. Walfish, 'A review of statistical outlier methods', *Pharmaceutical Technology*, 30, 2006, pp. 82-86.
28. Y. Xu, J. Xue, S. Park, Y. Yue, 'Towards a multidimensional view of tourist mobility patterns in cities: A mobile phone data perspective', *Computers, Environment and Urban Systems*, 86, 2021, 101593,ISSN 0198-9715,https://doi.org/10.1016/j.compenvurbsys.2020.101593.
29. Y. Xu, J. Li, J. Xue, S. Park, Q. Li, 'Tourism Geography through the Lens of Time Use: A Computational Framework Using Fine-Grained Mobile Phone Data', *Annals of the American Association of Geographers*, 111(5), 2021, 1420-1444.

Appendix

Correspondence table: airports' names – IATA codes.

Name	IATA code
Alghero	AHO
Ancona	AOI
Bari	BRI
Bergamo	BGY
Bologna	BLQ
Brindisi	BDS
Catania	CTA
Cagliari	CAG
Comiso	CIY
Firenze	FLR
Genova	GOA
Lamezia Terme	SUF
Milano Malpensa	MXP
Milano Linate	LIN
Napoli	NAP
Olbia	OLB
Palermo	PMO
Perugia	PEG
Pescara	PSR
Pisa	PSA
Rimini	RMI
Roma Ciampino	CIA
Roma Fiumicino	FCO
Torino	TRN
Trapani	TPS
Treviso	TSF
Trieste	TRS
Venezia	VCE
Verona	VRN