# Mobile Phone and Credit Card Data: Experience from 10 years of Public Private Partnership

Jaanus Kroon*
Eesti Pank / Bank of Estonia, Tallinn, Estonia – jaanus.kroon@eestipank.ee

## Abstract

Since 2008, Eesti Pank, the central bank of Estonia, has been cooperating with private companies to source big data for official statistics. The central bank uses mobile phone data from Estonian mobile network operators to quantify inbound and outbound travel, and credit card payment data to calibrate expenditure figures. The combination of these two datasets and several other sets of reference data allows timely and efficient production of trade in services statistics, used in the compilation of the balance of payments and the national accounts. The cooperation has stood the test of time as part of the statistical business process at the central bank.

After a full decade of cooperation between official statistics and big data providers in Estonia, we look back at the experience gained and the lessons learned:
- the evolution of mobile phone data and credit card data over the years
- the ways both types of data can be used for official statistics
- lessons from the partnership

This best practice can be used by statisticians in accessing and adopting data sources to produce official statistics in their own countries.

**Keywords:**
big data; mobile positioning; central bank statistics; balance of payments; cooperation for official statistics.

## 1. Introduction

Globalisation, the blurring of borders and the complexity of measuring cross-border transactions have posed a considerable challenge for external statisticians for some time now. The rapid growth in worldwide travel, membership of the Schengen Area, where there are no border controls and so no data collection at borders, and the discontinuation of regular border surveys by Statistics Estonia because of budget cuts forced Eesti Pank as the institution responsible for external sector statistics to find an alternative way to continue the border-crossing time series. Border-crossing data are an important input in the compilation of the country's monthly and quarterly balance of payments, where exports and imports of travel services play a major role.

Many alternative options were explored to meet the demand for a high quality and efficient data source at a reasonable cost and with low labour intensity. The table 1 briefly describes these alternatives and lists the advantages and disadvantages of each, given the situation at that time.

Eesti Pank opted for mobile positioning as it was the simplest statistics instrument and relatively low-cost. The choice was largely determined by the availability of the potential partner, the Department of Geography of the University of Tartu in Estonia, whose spin-off company Positium OÜ had had regular experience in using mobile positioning data in urban and regional geography and planning since 2001. Building on the regular data exchange that was already established with the biggest mobile operator in Estonia and on the availability of related calibration surveys, remarkable scale effects were expected. In 2008, the central bank started working with researchers at the University of

Table 1. List of possible alternatives

| Alternatives | Pros | Cons |
|---|---|---|
| Taking over the border-survey from Statistics Estonia or financing it | - Two years of experience, routine<br>- Can be partly integrated with visitor motivation interview surveys | - Time and labour intensive<br>- Disproportionately expensive<br>- Insufficient reliability: coverage, sampling, grossing-up, etc. |
| Setting up an accommodation statistics-based assessment model | - Monthly frequency<br>- Easy to run<br>- Reliable geographical allocation<br>- Low costs | - Does not cover outbound tourists<br>- Additional costs for a regular calibration survey (private vs. hotel stays; visitors vs. tourists, etc.)<br>- Estimation errors |
| Using a Road Office sensor data-based assessment model (car-counters on the road at border-crossings) | - High periodicity<br>- Supplements existing harbour and airport business data on border-crossings | - Very limited coverage<br>- Additional costs for a calibration survey (geo allocation, number of passengers, length of stay)<br>- Estimation errors and quality issues |
| Using credit and debit card data from Northern European Transaction Services (NETS Estonia) | - High periodicity<br>- Gives estimation on total expenditures and indirect geo allocations<br>- Coverage (only one service provider in Estonia)<br>- No administrative burden<br>- Low costs for compiler<br>- Long experience with payments statistics | - Additional costs for a calibration survey (card vs. cash, expenses and visitors on "behalf of family", etc.)<br>- Negative example from neighbouring countries<br>- "Noise" related to e-commerce |
| Introducing the methodology based on mobile network roaming information | - High periodicity<br>- Representativeness (almost everyone has a mobile phone)<br>- Operational information in time and space (including geographical allocation)<br>- No major administrative burden | - Lack of experience and practice<br>- Undefined cooperation model with data providers<br>- Additional costs for the calibration survey of mobile usage patterns<br>- Substantial IT resources needed for data processing |

Tartu and Positium OÜ to develop the new data collection methodology and models. Methods for inbound travel were fixed in 2008-2009, and those for outbound travel in 2009-2010. The complete methodology was revised and updated in 2015.

Mobile owners are a representative large sample whose spatial behaviour and characteristics in time can be extended to the entire population. The data are readily available, which makes data collection faster and more cost-effective. As mobile phones are widely used, the resulting data set is comprehensive; it minimises the human factor affecting interviewer interpretation in surveys, and ensures homogeneity. So the data are more accurate and of better quality than those collected by traditional data collection methods.

Although each transaction with a payment card would give a location-based fact similar to the roaming transaction of a mobile owner, ten years ago it was obvious that the facts from card data were not as numerous as those detected by mobile positioning. That is why it was decided to continue payment card data collection on an aggregated basis for estimating travel expenditures and to use it as one of the main alternative data sources that can help to validate the border-crossing aggregates derived from the mobile positioning data (MPD).

## 2. Mobile positioning based statistics of border crossing: methodological aspects

The jointly developed methodology is based on the use of readily available log files from Mobile Network Operators (MNOs), registering the information needed for billing incoming and outgoing roaming activities like voice calls, SMSs and MMSs, mobile-data usage, and mobile supported GPS

usage. These activity events are called Call Detail Records (CDR). The parameters that the methodology needs for each call activity are:

- SIM card ID, replaced by a randomly generated pseudonymous ID for statistical use;
- date and time;
- antenna ID with location data;
- country ID.

In line with the Balance of Payments (BoP) methodology, mobile positioning determines the residence of a traveller using the permanent residence criterion, regardless of the resident's citizenship or nationality. As a rule people sign a contract with a mobile company in the country where the phone will be used most frequently, and so the phone owners are presumed to reside in the country where their SIM card is registered. This approach is supposed to give even more precise results under the residency concept of BoP statistics.

The amount, length and nature of the visits of Estonian residents and non-residents are determined by the location-based anonymised use patterns of mobile phones derived from the roaming activities in the reporting resident operator network, and operator clients' roaming activities in networks abroad. The statistics on inbound and outbound travel reflect both same-day and overnight visits:

- The number of visitors is determined from the encoded number IDs.

- The duration of a visit is determined from the temporal distribution of the call operations of an individual mobile phone. If all the call activities are within a single day, the recorded duration of the visit is one day. If there are call activities on several days, the number of the days with call activities and any 'empty' days in between is assumed to be the number of days of the visit. If there are no call activities for seven consecutive days, the person is assumed to have left the country. For outbound travel, all the visits are compared against the activities in the same period in Estonia, and if necessary the initial visit is split into several visits.

Data processing broadly consists of the following steps, some of which are country-specific:

- Quality control of the data collected from the operator's system. Since there is a huge amount of data, filters have been developed to find and correct errors based on data characteristics.

- Filtering and evaluation of the roaming data in order to ensure representativeness and quality of data.

- Geographical and temporal interpolation, i.e. linking additional parameters to ensure administrative and chronological comparability.

- Elimination of border bias:

  o the registration of the roaming activities by Estonian residents who are incidentally in the coverage area of foreign mobile operators or

  o the registration of the roaming activities of the customers of foreign mobile operators who are incidentally in the coverage area of mobile phone masts located near Estonian borders, such as ships' crews and passengers on passing ships, or residents of the neighbouring countries.

- Elimination of transit travel:

  o For inbound travel, travellers detected in Estonia's main transit corridors if the stay in Estonia is shorter than a certain limit; there are 10 such transit areas, including the Tallinn-Ikla and Tallinn-Narva roads and the Estonian section of the Riga-Pskov road, Tallinn Airport, and the major ports.

  o For outbound travel, trips through foreign countries. Countries visited without an overnight stay, which do not comply with the criteria of a destination country, are classified as transit countries. One of the criteria for determining visiting and transit countries is the distance from Estonia and the length of stay in hours.

- Profiling and segmenting of individual trips for single-day and multiday visitors; the number of visits to and from the country, the country of origin (inbound) and destination (outbound) and the number of nights and days spent are calculated. Certain exceptions apply to long-term students and border workers:

  o Long-term stay of non-residents. Students are expected to have stayed in Estonia for over 183 days during the preceding 12 months. Two mutually independent criteria have been chosen for distinguishing non-resident border workers: either based on the number of visits to Estonia in a year or on the duration of the stay. A visitor is considered to be a non-resident worker when they have stayed in Estonia on seven or more occasions in a year or for more than 91 days during the past 12 month (derived from experience).

  o Long-term visits by Estonian residents who work or study abroad, if they stayed in the destination country for over 91 days during the past 12 months.

- Grossing-up to the total population with a special penetration model which takes into account the following:
  o Market shares of the mobile network operators (MNOs) covered
  o Penetration rate of SIM cards in the country with roaming services
  o Calibration survey results for differences in phone usage between residents of different countries, seasonality, and over or under coverage of SIM cards to cover double or travel SIMs
  o Available reference data
  o Estimate using the calibration coefficients matrix

Theoretically, it would be possible to use random sampling of SIM cards instead of the whole raw data-file, which could be done with weights in grossing-up, but given the small size of Estonia and the small population, it is considered impractical.

### 3. Card payment statistics: methodological aspects

Eesti Pank has compiled and disseminated detailed monthly payment statistics since 1998. The statistics are based on aggregation of the detailed input data reported directly by credit and other payment institutions and as a part of the reporting obligation of monetary and financial institutions, following the census principle.

The following data are reported for card payments:
  o Date [Month ]
  o Type of payment [Cash deposit/withdrawal | Point of sales (POS) payments | E-commerce payments, which are card payments in internet-based stores or elsewhere in the online environment]
  o Residency of card issuer [ISO country code]
  o Country of payment [ISO country code]
  o Number of transactions
  o Total turnover

All payments with similar identifiers are aggregated in the data model. No granular payment data are currently reported to keep the reporting volumes low. The biggest limitation of using the card data for travel statistics is that the residency of the card issuer is not always a good proxy for the residency of the card holder, i.e. the traveller.

Despite this, card payment statistics are a good data source for quantifying inbound and outbound travel and credit card payment data to calibrate expenditure figures. The dynamics of card transaction volumes and turnover correlate strongly with the dynamics of visits. Card expenditures at home and abroad correlate strongly with BoP travel exports and imports.

Card payment statistics could be developed further by exploiting other information stored by the card service provider for each card payment, such as Merchant Category Code (MCC), assigned by the

acquiring bank when the business applies for a merchant account, and Transaction Category Codes (TCC) groups according to ISO 18245. Such data are readily available and could give additional information needed for the estimation of BoP sub-categories, and provide important detail for economic flash forecasts and other users of statistics.

## 4. Data validation and cooperation model

The daily cooperation takes the form of a Public Private Partnership (PPP), outsourcing data processing contracts from Positium OÜ. Three-year contracts have been announced in 2009, 2012, 2015 and 2018.

The work has been arranged according to the Generic Statistical Business Process Model (GSBPM) as described in table 2.

Table 2. Work arrangements

| Positium OÜ | Eesti Pank |
|---|---|
| | - Specifying needs and defining business case |
| - Development of methodology | |
| - Design | |
| - Build | |
| - Data collection and processing | |
| - Calibration surveys | - Data analysing and validation |
| - Revising Design and Build | - Data dissemination |
| | - Feedback for fine-tuning Design and Build |

Monthly results provided by OÜ Position are verified and validated by Eesti Pank. Along with mobile data, other official data sources are used (e.g. the number of passengers in the Port of Tallinn and Tallinn Airport, crossings of the Estonian/EU administrative border, official accommodation statistics, the press, etc.). The most relevant time series for data validation is provided by card payment statistics (figures 1 and 2).



**Figure 1: Comparision of different data sources (inbound volumes)**

**Figure 2: Comparision of different data sources (outbound dynamics)**
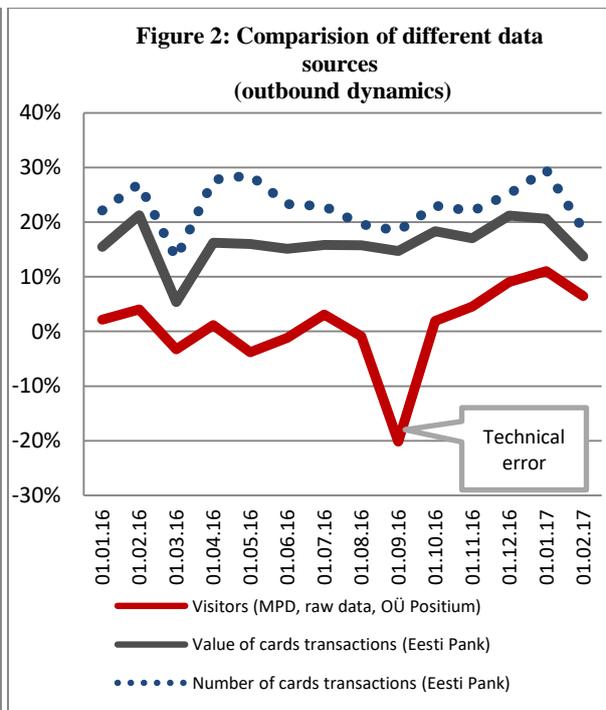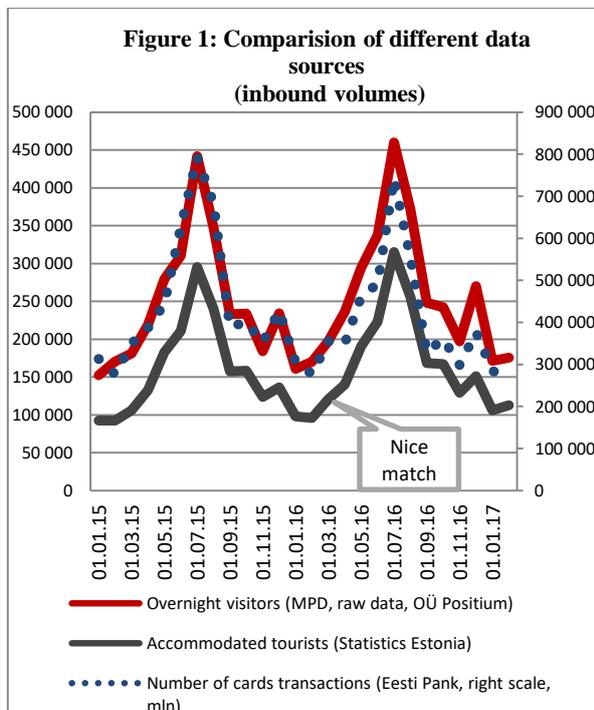
In case of doubts about the quality of the original data, an enquiry is addressed to our partner, who carries out his own analysis and, if necessary, forwards the enquiry to the mobile operators. If the data

need to be corrected, each specific case will be considered separately and adjusted as appropriate. If it is not possible to identify the specific reasons for the discrepancy, the data will be corrected based on the data from the same period of the previous year, multiplied by a coefficient equal to:

- o the average increase/decrease of available payment statistics and other reference data or
- o the average change in the last eleven months.

The main inconsistencies in the data relate to a) errors in the source data files provided by MNOs (missing or invalid data); b) changes in MNOs market shares or roaming rates; c) under coverage caused by improved availability of alternative networks (wifi) or d) new sources of noise, incl. overflights.

The data are used as an input for monthly and quarterly external sector statistics and have been disseminated as official statistics on Eesti Pank's website on a quarterly basis since 2012. A press-release on this is attached when disseminated.

Border-crossing statistics is included in the list of Eesti Pank's statistical activities, which forms a part of the official statistical programme of Estonia. Compared with earlier expenditures on similar statistics, involving regular border-crossing surveys and the related interviews, the current approach is remarkably cost-effective.


## 5. Conclusions

A decade of experience of using mobile data for external sector statistics has shown that the methodology offers a reliable overview of travellers crossing the Estonian border. When Eesti Pank started publishing the time series of international travel in 2012, Estonia was one of the first few countries publishing official statistics based on big data. It should be pointed out, however, that the methodological concepts cannot be uniformly applied in every country. Local conditions and the particular statistical goals should be taken into account.

The biggest advantage of the method is its speed, as it uses existing information, stored by mobile operators as potential respondents for statistics. There are neither direct costs associated with a network of interviewers nor a burden for travellers as potential respondents. In comparison with earlier expenditures on regular border-crossing surveys with interviews, the current approach is remarkably cost-effective.

The current PPP-based collaboration model enables both parties to take advantage of their specialisation. While Eesti Pank has experience in producing national statistics in traditional fields, OÜ Positium as a spin-off company of the Department of Geography at the University of Tartu is better equipped to model human mobility from different facets. As border-crossing statistics is not a core activity for the central bank, this cooperation enables us to draw on OÜ Positium's long-term experience in the field of direct processing of big data (based on the Hadoop software framework).

**References:**
1. Using mobile positioning data for tourism statistics: methodology and data model. OÜ Positium LBS, Eesti Pank; Tartu 2009 (in Estonian).
2. Ahas, R. Aasa, A., Roose, A., Mark, Ü., Silm, S. 2008. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. Tourism Management 29(3): 469–486.
3. Using mobile positioning data for tourism statistics: updated methodology and data model. OÜ Positium LBS, Eesti Pank; Tartu 2015 (in Estonian).
4. Eurostat. (2014). Feasibility study on the use of mobile positioning data for tourism statistics.Reports on Eurostat Contract No 30501.2012.001-2012.452.
5. Establishment of reports on payment statistics of credit institutions. Decree of the Governor of Eesti Pank No 8 of 02/06/2014
6. Payment guides. Centus. [www] https://centus.com/uk/a-guide-to-mcc-codes-tcc-codes-and-mcg-groups (12.11.2017)
7. International Travel Statistics.(2019) Eesti Pank. [www] http://statistika.eestipank.ee/#/en/p/MAKSEBIL_JA_INVPOS/1410 (29.03.2019)