

**UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE**

**CONFERENCE OF EUROPEAN STATISTICIANS**

**Seminar on New Frontiers for Statistical Data Collection**  
(Geneva, Switzerland, 31 October-2 November 2012)

Topic (i): New data sources

**MOBILE POSITIONING AS A POSSIBLE DATA SOURCE FOR  
INTERNATIONAL TRAVEL SERVICE STATISTICS**

**Contributed Paper**

Prepared by Jaanus Kroon, Eesti Pank (Bank of Estonia)

**I. Introduction**

1. Globalisation, the blurring of borders and the complexity of measuring cross-border transactions have been posing a considerable challenge for external statisticians for some time now. Border crossing statistics and travelling geography as the basis for measuring international travel services industry in the balance of payments current account have not been left untouched either. Rapidly growing worldwide travels, membership in the Schengen Area, where no regular border controls are used between the member countries, and the termination of regular border surveys by Statistics Estonia due to budget cuts have forced Eesti Pank, who is responsible for external sector statistics, to adopt an innovative and cost-effective data source to replace and continue the border crossing time series.
2. Although the area is not a core domain of the central bank statistics itself, the time series of inbound/outbound travellers supplemented by their expenses *per capita* is an important input for statistical models measuring international travel consumption. The paper provides an overview of the key aspects and features related to the experience of Eesti Pank in adopting mobile positioning as a possible official statistics data source, paying attention to the historical measuring of visitor flows in Estonia and to related issues, the nature of mobile positioning and its applicability in border-crossing statistics based on the Estonian experience.

**II. Background and challenges**

**A. Measuring visitor flows in Estonia and the way forward**

3. In 2004, when Estonia joined the EU, the geo-political environment changed. The existing administrative border-control system as an accurate detailed data provider of border crossing statistics for Statistics Estonia, responsible for the domain, was simplified, and in 2007, abolished for intra-Schengen visits. Due to the incomplete coverage of the Police and Border Guard Board records, the data source was replaced by a) new quarterly frontier surveys for estimating international arrivals, and b) the existing quarterly household survey, which was adopted to estimate outbound visits. These data were supplemented by official accommodation statistics and figures provided by Tallinn Airport and harbours. Although the mentioned surveys raised the

question of a possible imbalance in between costs and quality, they still were a sufficient input for the Estonian balance of payments statistics, where travel related items have always played a key role: international tourism makes up about 25% of Estonia's services exports.

4. The economic crisis and remarkable government budget cuts in 2009 also affected the financing of official statistics. Revisions of the official statistical programme by setting negative priorities broke the border crossing time series. It was agreed that Eesti Pank would seek possibilities to resume the latter, at least within its competence, to compile external sector statistics.
5. Since there was demand for high quality and an efficient data source at a reasonable cost (incl. low labour intensity), many alternative options were explored. The following table briefly describes the latter and lists the advantages and disadvantages of each, taking into account the current situation at that period of time:

<i>Alternatives</i>	<i>Pros</i>	<i>Cons</i>
Take over the border-survey from Statistics Estonia (or to finance it)	<ul style="list-style-type: none"> <li>- 2 years of experience, routine</li> <li>- Possibility to partly integrate with visitor motivation interview surveys</li> </ul>	<ul style="list-style-type: none"> <li>- Time and labour intensive</li> <li>- Unreasonably expensive</li> <li>- Insufficient reliability: coverage, sampling, grossing-up, etc.</li> </ul>
Implement an accommodation statistics-based assessment model	<ul style="list-style-type: none"> <li>- Monthly frequency</li> <li>- Easy to implement</li> <li>- Reliable geographical allocation</li> <li>- Low costs</li> </ul>	<ul style="list-style-type: none"> <li>- Additional costs regarding regular calibration survey (private vs. hotel stays; visitors vs. tourists, etc.)</li> <li>- Does not cover outbound tourists</li> <li>- Estimation errors</li> </ul>
Participate in the project initiated by the Road Office to install car-counters on the road at border-crossings	<ul style="list-style-type: none"> <li>- High periodicity</li> <li>- Supplement to harbour and airport statistics</li> </ul>	<ul style="list-style-type: none"> <li>- Investment costs</li> <li>- Additional costs regarding calibration survey (geo allocation, number of passengers, length of stay)</li> <li>- Estimation errors and quality issues</li> </ul>
Credit-card information, based on Northern European Transaction Services (NETS)	<ul style="list-style-type: none"> <li>- High periodicity</li> <li>- Gives estimation on total expenditures and indirect geo allocations</li> <li>- Coverage (only one service provider in Estonia)</li> <li>- No administrative burden</li> <li>- Low costs for compiler</li> </ul>	<ul style="list-style-type: none"> <li>- Additional costs for calibration survey (card vs. cash, expenses and visitors on "behalf of family", etc.)</li> <li>- Negative example from neighbouring countries</li> <li>- "Noise" related to e-commerce</li> </ul>
Derive travelling statistics from mobile-phone roaming information	<ul style="list-style-type: none"> <li>- High periodicity</li> <li>- Representativeness (almost everyone has a cell phone)</li> <li>- Operational information in time and space (incl. geographical allocation)</li> <li>- No administrative burden</li> </ul>	<ul style="list-style-type: none"> <li>- Lack of experience and practices</li> <li>- Undefined co-operation model with data providers (3 mobile operators)</li> <li>- Additional costs regarding the calibration survey of mobile usage pattern</li> <li>- Remarkable IT resources for data processing</li> </ul>

6. To this end, Eesti Pank opted for mobile positioning as the simplest and a relatively low-cost statistics instrument. The choice was largely determined by the availability of potential partner, the Department of Geography, University of Tartu (Estonia) whose spin-off company, Positium LBS, has regular experience in using mobile positioning data in urban and regional geography/planning since 2001. Relying on the already established regular data exchange with the biggest mobile operator of Estonia and on the availability of related calibration surveys, the scale-effect was expected to be remarkable. In 2008 the central bank set up cooperation with scientists at the University of Tartu to develop the new data collection methodology and models. Methods for inbound travel were fixed in 2008-2009, and for outbound travel in 2009-2010.

## **B. The phenomena of mobile positioning**

7. Mobile positioning is a synthesis of two traditional data collecting techniques – census and global positioning systems (GPS). Through tracking the location of mobile phones and diffusions in space and time, it is possible to study the underlying geographical phenomenon and human mobility patterns with three remarkable merits: large sample size, high degree of accuracy and high spatio-temporal dynamics.
8. Mobile positioning is technically possible from almost any mobile network and it can be performed in two ways:
  - (a) *Active mobile positioning* is used for tracking the location of mobile phones in real time through a network of antennas. More precise methods use trilateration by obtaining the distance from many antennae using direction and time lag of signals. An alternative option is Network Assisted GPS positioning and the navigation of smart-phones if the service is activated by the user.
  - (b) *Passive mobile positioning* uses location information from historical log files stored by mobile service providers. Operators' systems generate a large amount of phone usage data as a by-product with location information attached to it. Mostly, this data is used internally for business purposes: charging clients for services, providing usage statistics, marketing, etc., but it also holds valuable information for anonymous statistics about space-time movement of phones without significant marginal costs.
9. From the statistical point of view, mobile owners are a representative large sample within a statistical population whose spatial behavior and characteristics in time can be extended to the entire population. Theoretically, data are readily available, and this makes data collection faster and more cost effective. Because cell phones are widespread, the resulting data set is comprehensive, minimizing the human factor (interviewer interpretation in surveys), and ensuring homogeneity, which improves the accuracy and quality of the data compared to traditional data collection methods.
10. Though being a promising technique, mobile positioning also entails some limitations and weaknesses, such as issues regarding surveillance and privacy, the business secrets of mobile operators, and the peculiarities of data processing.
11. Nevertheless, the legal constraints for official statistics should be eliminated at the legislator level. According to the European Statistics Code of Practice, the statistics-related costs and response burden should be effective and optimized. The producer of official statistics should primarily use data collected from the administrative records and other databases of government agencies and other legal persons, if statistics complying with the quality criteria of official statistics can be produced on the basis of these data. Cell phone usage records as a data source should be in accordance with the mentioned criteria. The development of advanced data processing techniques and models should be regarded a challenge.

### **III. Official mobile positioning based travel statistics of Estonia: methodology and practice**

#### **A. General overview**

12. One of option of the passive mobile positioning helps to determine a person's outbound travel during the person's stay outside the country of residence in the coverage area of foreign mobile operators on the basis of phone activities: calls, SMSes/MMSEs, mobile-net usage, data transmission operations, etc., when roaming.
13. Traditional GSM Roaming is defined as the ability for a mobile-phone customer to automatically make and receive voice calls, send and receive data, SMSes/MMSEs, or access other services, when travelling outside the geographical coverage area of the home network, by means of using a visited network. Roaming activities and related information are available in the 'billing log' for customer what makes the approach easy. The registered parameters for each call activity are the following:
  - (a) SIM-card ID (replaced by randomly generated pseudonymous ID of the phone for statistical use);
  - (b) date and time;
  - (c) antenna ID with location data;
  - (d) country ID.
14. The methodology is based on the anonymised moving patterns according to the roaming activities of mobile phones in reporting resident operator network and operator clients' roaming activities in networks abroad with the patterns being aggregated and extended to the total population by using simple statistical models.
15. In compliance with the balance of payments methodology, mobile positioning determines the residence of a traveller by the permanent residence criterion, regardless of the resident's citizenship or nationality. As a rule, entering into a contract with a mobile company is most favourable in the country where the phone is used most frequently, therefore the presumable residence of phone owners is determined by the registration country of the SIM-card. The approach is supposed to give even more precise results according to the residency concept.
16. Data processing broadly consists of the following steps:
  - (a) Automatic quality control of the data collected from the operator's system. Since the data amount is huge, filters to find and correct errors were developed, based on the characteristics of the data.
  - (b) Filtering and evaluation of the roaming data in order to ensure representative data and scientific quality.
  - (c) Geographical and temporal interpolation i.e., linking additional parameters to ensure administrative and chronological comparability.
  - (d) Estimating the amount, length and nature of the visits of Estonian residents abroad and of non-residents to Estonia by algorithms determined by the location-based usage patterns of mobile phones, depending on the country and region of use.
  - (e) Extension to the entire population by a special penetration model which takes into account the penetration rate and market share of mobile operators within roaming service, the differences of phone usage between residents from different countries, seasonality, etc.

17. The data model is a part of special data mediation software called Positium Data Mediator, created in collaboration between Positium LBS and the Department of Geography of University of Tartu, applied by Positium LBS, permanent outsourcing contractor of Eesti Pank, in providing location based services to the public.
18. Monthly results are verified and validated by Eesti Pank while also taking into account other official data sources (number of passengers in the Port of Tallinn and Tallinn Airport, crossings of frontiers on the Estonian/EU administrative border, official accommodation statistics, the press, etc.). Data are used as an input for monthly and quarterly external sector statistics and published as official statistics on Eesti Pank's website on a quarterly basis.
19. According to the definition, international travel is a visit of a resident of one country to another country, while the length of the visit is shorter than one year and the main purpose of the visit is not gainful activities or long-term studies in the country of destination. When determining calculation algorithms by the location-based use patterns of mobile phones, the country specific aspects should be taken into account for both inbound and outbound visits. The most important of these are described below.

## **B. Inbound travel**

20. Statistics on inbound travel reflect only the *same-day* and *overnight* visits of non-residents to Estonia.
  - (a) The *number of visitors* is determined on the basis of the encoded number IDs of roaming mobile phones in Estonian operators' networks.
  - (b) The *duration of a visit* is determined on the basis of the temporal distribution of call operations performed by an individual mobile phone. If the call activity is performed in a single day, the recorded duration of the visit is one day. If call activities are performed on several days, the number of consecutive days and "empty days" between is assumed as the number of days of the visit. If call activities are not performed on seven consecutive days, it is assumed that person has left the country and returned. This choice is based on the statistics of average duration of visits and repeated visits. Intervals exceeding 7 days imply a higher probability of repeated visits and such cases are not automatically regarded as a single long visit.
21. The following data are excluded from the source data:
  - (a) "*cross-border noise*", i.e., the registration of phone roaming by the customers of foreign mobile operators that are incidentally in the coverage area of mobile phone masts located near borders (e.g., crew members and passengers of passing ships, residents of neighbouring countries);
  - (b) *transit travel* registered in Estonia's main transit corridors (10 transit areas have been distinguished: Tallinn-Ikla and Tallinn-Narva roads and the Estonian section of Riga-Pskov road, Tallinn Airport and major ports) in case the length of stay in Estonia is shorter than a certain limit;
  - (c) *long-term stay* of non-residents. Students are expected to stay in Estonia for over 183 days during the past 12 months. Two mutually independent criteria have been chosen for distinguishing non-resident workers. Workers are distinguished either on the basis of the number of visits to Estonia in a year or on the basis of the duration of the stay. A visitor is considered to be a non-resident worker when he/she has stayed in Estonia on more than 7 occasions in a year (inclusively) or for more than 183 days during the past 12 month.

## C. Outbound travel

22. Statistics on outbound travel reflect the same-day and overnight visits of residents to foreign countries.
- (a) The *number of visits* is determined on the basis of the resident encoded number IDs of roaming mobile phones in foreign operators' networks abroad, charged by the resident mobile operator. Outbound travel by Estonian residents may include visits to a number of countries. The country of destination is the farthest country from Estonia or the country (countries) where at least two consecutive days have been spent.
  - (b) The *duration of a visit* is determined on the basis of the temporal distribution of call operations performed by an individual mobile phone abroad, without any concurrent activity in Estonia. If the call activity is performed in a single day, the recorded duration of the visit is one day. If call activities are performed on several days, the number of consecutive days and "empty days" between is assumed as the number of days of the visit. If call activities are not performed during six consecutive days, it is assumed that the person has returned. All calculated visits are compared to the activities carried out at the same period in Estonia, if necessary, the initial visit is split. The number of visits and the duration are usually separated, but in some cases the first and the last day of single visits may be overlapped.
23. The following data are excluded from the source data:
- (a) "*cross-border noise*", i.e. the registration of phone roaming by Estonian residents that are incidentally in the coverage area of foreign mobile operators;
  - (b) *transit travel* through foreign countries. Other countries which are visited without overnight stay and which do not comply with the criteria of the destination country, are classified as transit countries. One of the criteria for determining destination and transit countries is the distance from Estonia;
  - (c) *long-term visits* by Estonian residents who work or study abroad, in case they stayed there for over 183 days during the past 12 months.

## D. Legal aspects

24. Anonymous mobile positioning data are used in compliance with the Estonian and EU legislation on data protection: Estonian Personal Data Protection Act; EU directive 95/46/EC on privacy protection of personal data processing and movement of personal data; EU directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector. The mobile positioning methodology has been developed and is used in various fields in cooperation with the Data Protection Inspectorate regarding personal data protection aspects.

## IV. Conclusions

25. Based on two years' experience of using data for external sector statistics, the methodology is considered to offer a reliable overview of persons crossing the Estonian border to travel abroad or to Estonia. Eesti Pank started publishing the international travel statistics time series in 2012.
26. The characteristics and definitions of the methodology cannot be general and equally applied in every country. Local conditions and set goals to be explored should be taken into account.
27. The biggest advantage of the method is its speed, laying already existing information stored by mobile operators as potential respondent for statistics. There are neither direct costs associated with the network of interviewers nor a burden for travellers as potential respondents.

28. Eesti Pank chose the partner by a public procurement process and concluded with them a permanent outsourcing contract. Comparing current costs with earlier expenditures on regular border-crossing surveys with necessary interviews for producing similar statistics, the current approach is remarkably more cost-effective.
29. The greatest concerns regarding mobile positioning studies expressed by experienced scientists and experts involve sampling. As with several statistical surveys, there is no overview available of the total population. In addition, it is essential to know the usage pattern of mobile phones in systems which receive telephone usage data as input. It is essential to understand the logic of phone usage by different user groups as it is apparent that different age and social groups, various income groups, nationalities and tourist types use mobile phones differently. No accurate related scientific studies are currently available.
30. Is the “Big Brother” really watching? The most topical questions of mobile positioning are related to privacy. The fear regarding a possible abuse of government power, particularly in respect to civil liberties and mass surveillance. Fortunately, the most recent detailed personal studies and the Census show that public confidence in statistics has been improved. Fears should be refuted by the fact that data are anonymised and used only to investigate mass phenomena.
31. Though the collection of statistics on tourism is not a traditional task of Eesti Pank, and the latter has no ambitions to leave the boundaries of its competence, we may predict, on the basis of current experience, the relevance of mobile positioning as an important data source in official statistics, especially in international and domestic surveys on tourism and the movement of people. In Estonia it would be legally possible to request data on mobile positioning directly from the source for the sake of official statistics.
32. The world keeps changing and so does the usage of mobile telephony. The Internet, Skype, and so on, are more and more widely used. As a result, the models need to be revised to keep up with the times. Nevertheless, mobile positioning as an up-to-date data collection method has justified itself in every way and it has great prospects as a tourism statistics data source for official statistics.

## V. References

- International Recommendations for Tourism Statistics 2008. United Nations 2010.
- Mobiilpositsioneerimisel põhinev välisküllastajate turismistatistikute uuring. OÜ Positium LBS, Eesti Pank; Tartu 2009.
- Mobiilpositsioneerimisel põhinev välisküllastajate turismistatistikute uuring: penetratsioonimudeli kirjeldus. OÜ Positium LBS, Eesti Pank; Tartu 2009.
- Mobiilpositsioneerimisel põhinev Eesti residentide väliturismistatistikute uuring. OÜ Positium LBS, Eesti Pank; Tartu 2010.
- Tiru, M. & Ahas, R., “Using mobile positioning data for tourism statistics: methodological and legal issues”. 10th International Forum on Tourism Statistics, Portugal 2010.
- Ahas, R., Tiru, M., Saluveer, E. & Demunter, C., “Mobile telephones and mobile positioning data as source for statistics: Estonian experiences” Conference on New Techniques and Technologies for Statistics, 2011.
- 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.