Ossi Nurmi

Paper

Improving the accuracy of outbound tourism statistics with mobile positioning data

Survey response rates are declining at an alarming rate globally. Statisticians have traditionally used imputing and recalibration of weights to deal with non-response. In case survey response rates are well below 50 %, these methods may often result in little more than guesswork. Alternative data sources are needed to improve the accuracy of statistical estimates.

Tourism surveys are typically conducted by National Statistics Institutes or other research institutes using a combination of computer assisted telephone interview (CATI) or web interview (CAWI) methods. Respondents are asked to provide information on their tourism trips, including the destination, duration, purpose, mode of transport and accommodation used as well as expenditure associated with different products and services consumed.

Mobile positioning data can be considered as an alternative data source for measuring the number and duration of tourism trips. 'Traces' of the mobility of people, including tourism trips, are registered by the mobile network operators (MNOs) in the form of CDRs (call detail records) and DDRs (data detail records) and they are generated by the activities of mobile devices. The MNOs are obliged to store these events for a certain time period, typically one year.

Since 2016 Statistics Finland has worked with Finnish MNOs in order to translate these CDRs and DDRs into tourism specific monthly aggregates such as number and duration of trips by destination country.

This paper provides a summary of the methodological process that the Finnish MNOs have followed to compile statistics specifically for outbound tourism, i.e. the trips abroad by the subscribers of the Finnish MNOs. Two out of three MNOs were able to provide full 2017 data for outbound tourism.

A similar process may be adopted by National Statistics Institutes or other organizations who are approaching their national MNOs with the intent of obtaining data. The paper also provides guidelines on how the data can be used to better calibrate the existing tourism surveys in order to improve the estimation of tourism trips to different destination countries.

Processing raw data from MNO into tourism trips

During 2016 up to 2018, Statistics Finland carried out the work in two phases within the context of Eurostat's ESSNet Big Data –project. The focus in these projects was on three statistical domains: inbound tourism, outbound tourism and seasonal population.

The first phase focused on negotiations with national authorities and MNOs in order to set up such a process that is feasible from a legislative and technical point of view. The second phase was to carry out the process, collect data from each MNO and analyze the results.

The chosen approach relies on the operators to process the data and aggregate it for Statistics Finland. In the current Finnish legislation, only the operator is allowed to process the raw data using automatic means. The size of the raw data is also massive, with annual data consisting of several billions of events per operator.

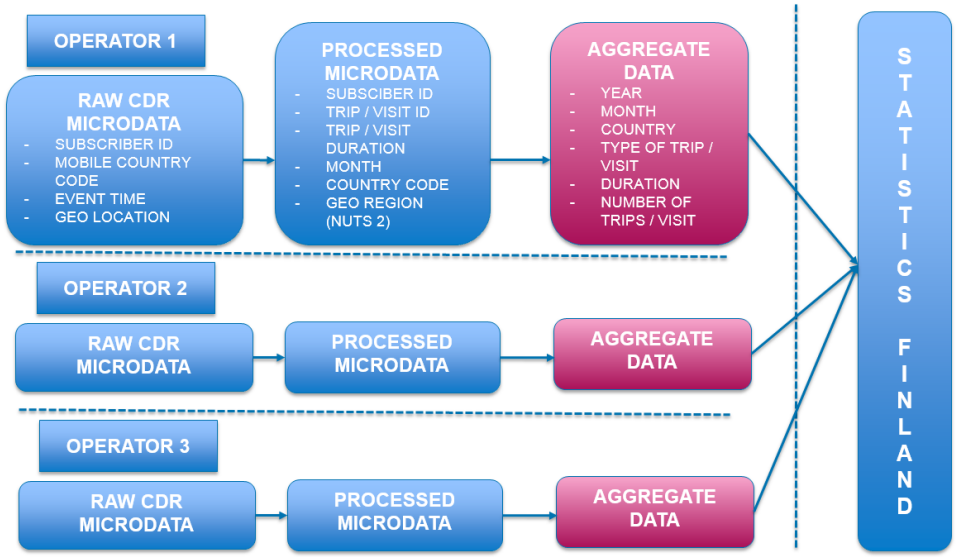

Figure 1 Process from raw microdata to aggregated trips

The starting point for each operator are the raw data of each of their subscribers. To be more precise, the subscriptions are associated with sim cards found on mobile devices. Machine-to-machine sim cards are excluded from the data as they do not represent the movement of people.

In case of outbound trips, the raw data consists of roaming events (calls, sms, mobile data) taking place outside of the subscriber's home network, in other words, a foreign country where the event took place. Based on the time gaps between these events, individual trips of each subscriber can be recognized.

The following example presents an imaginary case of the roaming events and outbound trips deduced for a single subscriber during a 30 days period. This subscriber is a particularly active traveler and five outbound trips are registered during the 30 day period.

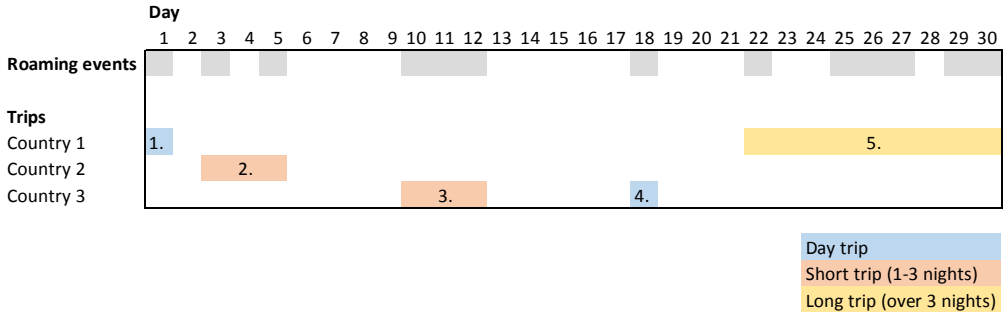

Figure 2 – Example of roaming events and identified outbound trips during a 30 day period

Roaming events are registered whenever the subscriber is making calls, sending sms or using data in a mobile network abroad. The first event in a country indicates the beginning of the trip to that particular country.

The end of the trip is determined based on time gaps and the events following those gaps. In case there is a long enough gap between these roaming events, such as in days 13 to 17, the assumption is that the subscriber has left the country after day 12 when the last event of trip 3. was registered. The first event of day 18 is thus considered as the beginning of a new trip to Country 3.

Based on the events of this subscriber, five outbound trips to three different countries are identified:

- Trips 1. and 4. are day trips, where all events occur during the same day
- Trips 2. and 3. are short trips
- Trip 5. is a long trip

This logic is described in greater detail instructions in the data request sent by Statistics Finland to the MNOs. Using this logic, the Finnish operators have processed the roaming events of all their subscribers in 2017 into outbound trip data. The final data is aggregated by month, country of destination, and trip duration.

Methodology for estimating outbound tourism trips

In the context of outbound tourism, the target population consists of all outbound tourism trips made by Finnish residents. According to the EU Regulation on Tourism Statistics[1], each member state of the EU shall transmit data on the national tourism trips to Eurostat annually. The annual sample size of the Finnish Travel –survey is 28,200 persons and roughly 14,700 of them were interviewed by phone concerning trips that ended during 2017. In total, the respondents reported roughly 2,700 outbound trips for 2017.

The data provided by the Finnish MNOs should be treated as samples of all outbound trips made by Finnish residents in 2017. Two out of three operators have provided the data and the market share of each operator is roughly one third of all subscribers. The source data thus contains the outbound trips of two thirds of the Finnish residents and one third are missing.

It should be pointed out that using the current approach, it's not possible to distinguish tourism trips from other trips in the MNO data. In tourism statistics, tourism is defined as visitors taking trips outside of their usual environment, which is defined as 'the geographical area, within which an individual conducts his regular life routines and shall be determined on the basis of the following criteria: the crossing of administrative borders or the distance from the place of usual residence, the duration of the visit, the frequency of the visit, the purpose of the visit'. Certain trips included in mobile positioning data may not qualify as tourism in case they are too short, too frequent or part of the routines of regular life. This potential source of overestimation of trips in the MNO data is often related to the neighboring countries of Finland with an open land border, namely Sweden and Norway. There are also other potential sources of over- and underestimation mainly related to border noise, zero or multiple devices, transit corridors such as airports and so forth.

As with every sample, it is of limited use by itself and coefficient weights are needed to estimate the target population: all outbound tourism trips. According to the Finnish Travel –survey, Finnish residents made 10.5 million outbound trips in 2017. This is a relatively stable figure based on 2,700 trips reported by the respondents.

In order to avoid many of the pitfalls related to over- or underestimation in mobile positioning data, this annual figure of 10.5 million outbound trips, is used as a frame and mobile operator data is weighted using the following formula:

$$Weight\ coefficient\ = \frac{Annual\ outbound\ trips\ in\ tourism\ survey}{\sum(Mobile\ network\ operator\ annual\ trips)_n}$$

In other words, the weight coefficient is obtained by dividing the annual outbound tourism trips by sum of trips made by the subscribers of each MNO. In the case of Finnish data from two (out of three) operators, the coefficient for year 2017 is less than 1.3.

While this coefficient could also be calculated separately for main destination countries or regions, it is still beneficial to start with a top-down analysis, using a single coefficient for all outbound tourism destinations. This kind of analysis can reveal countries that are over- or underestimated in the MNO data.

Top-down analysis of outbound tourism

The main strengths of mobile positioning are related to data granularity and timeliness. The MNOs possess a data source that is updated close to real time and it covers millions of trips to all countries in the world. In the current Finnish travel -survey Statistics Finland does not publish data of less than 100,000 trips due to small sample sizes. On an annual level, this means that only roughly 20 countries can be published.

This chapter presents the results of the top-down analysis where Finnish Travel -survey data is compared to MNO data.

Outbound tourism trips by month

The total number of outbound trips in 2017 was 10.5 million according to the Finnish-travel survey. The monthly distribution of these trips is shown below separately for the survey as well as the data provided by both MNOs.

[1] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2011.192.01.0017.01.ENG&toc=OJ:L:2011:192:FULL

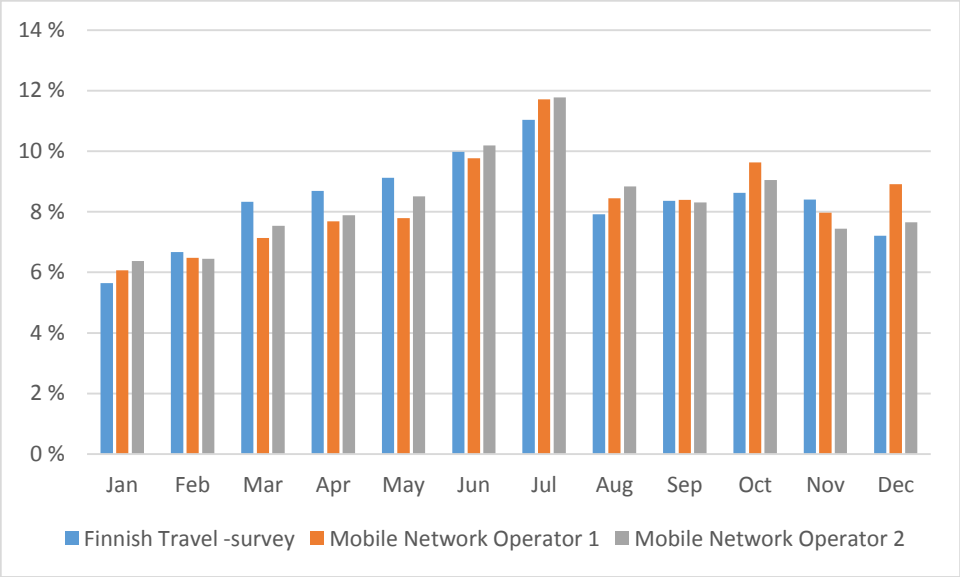

Figure 3 – Monthly seasonality of outbound trips using different sources in 2017

The monthly seasonality of outbound trips is strikingly similar for both MNOs. Both register the summer holiday months of July and June as the peak months of outbound tourism. The third biggest month for both operators is October, which is the month of autumn holidays in Finnish schools. The Finnish Travel –survey registers more outbound trips than the MNOs during the months of March through May.

Outbound trips by country of destination - Estonia and Sweden

Neighboring Estonia and Sweden are by far the most important outbound tourism destinations. 45 per cent of outbound trips are made to these two countries.

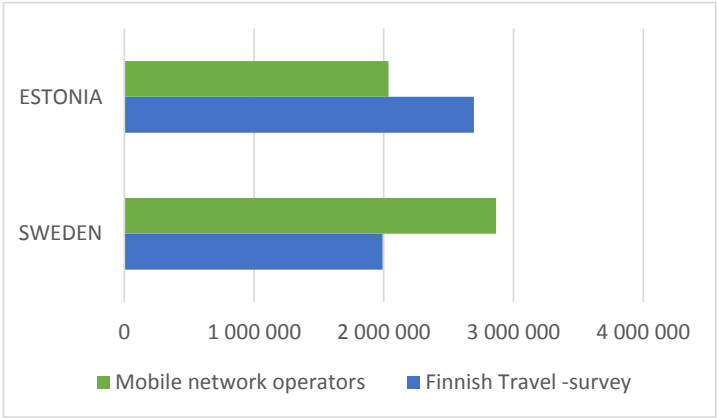

Figure 4 – Trips to Estonia and Sweden in 2017

There were 2.6 million trips to Estonia in 2017 according to Finnish travel –survey. Using the top-down approach for mobile positioning data, there were 24% less trips. This indicates that the top-down approach underestimates tourism to Estonia, as the 2.6 million trips from the survey is a very stable figure with only small changes annually.

In contrast, the outbound tourism to Sweden comprises of only 2.0 million trips in the survey, while the top-down approach estimates 44% more trips to Sweden. As mentioned, the neighboring countries with open land borders have many possible sources of overestimation such as frequent non-tourism trips, border noise etc.

Seasonality of trips to Estonia

There is yet another reliable data source for providing the monthly seasonality of outbound trips to Estonia, the main outbound tourism destination. Nearly all passengers to Estonia use one of the ferries that operate between the capital cities of Helsinki and Tallinn. The Finnish Transport Agency compiles statistics on the total monthly passengers departing to Estonia, including passengers of all nationalities.

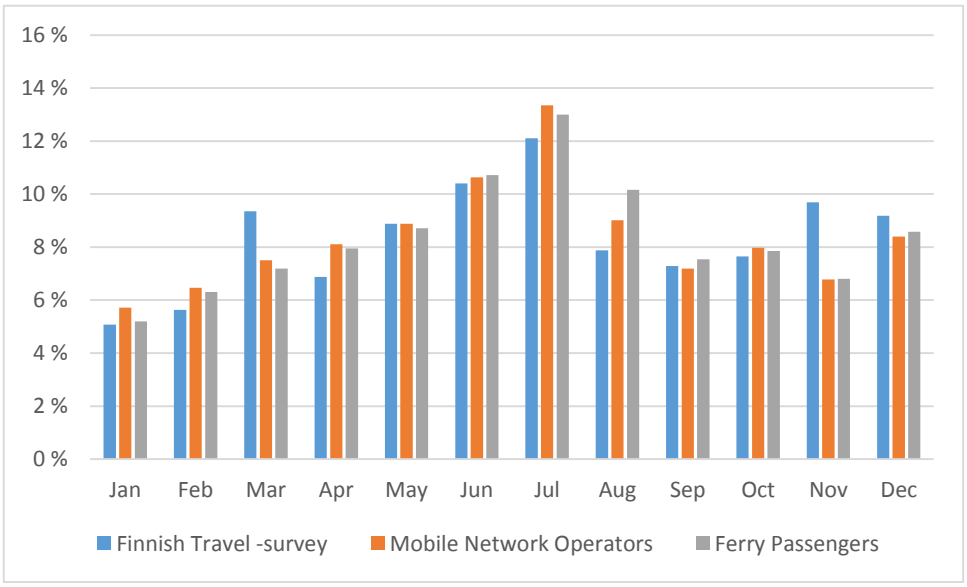

Figure 5 – Monthly seasonality of trips to Estonia in 2017

The seasonality of ferry passengers is in line with the outbound tourism data provided by the MNOs. In contrast, the seasonality of the Finnish Travel –survey seems to be affected by randomness. Some months in the survey are exceptionally high (March, November) and some too low (April, August) when compared to ferry passengers and MNO data.

Although the MNO data underestimates the total number of trips to Estonia, it seems to be a better data source for estimating the monthly seasonality. Unlike the survey, it's not affected by the randomness caused by a small sample size. This suggests, that MNO data should be used to adjust the seasonality of the survey.

Other top 20 destination countries

The top 20 destination countries make up 90 per cent of all outbound tourism trips. In other words, only 10 per cent of trips are made outside of these top 20 countries.

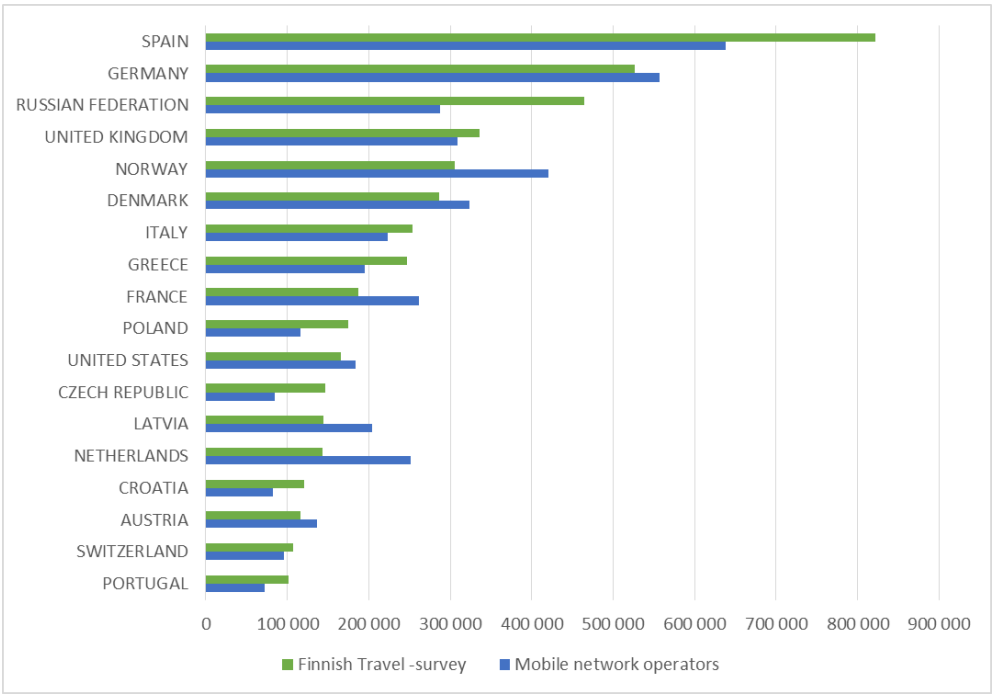

Figure 6 –Outbound trips to top 20 countries in 2017

Assuming that the estimate provided by the survey is accurate, the top-down approach estimate for mobile positioning is within the range of 42 per cent less trips, in case of Czech Republic, up to 75 per cent more trips in case of Netherlands.

Norway, another neighboring country of Finland with an open land border, has 38 per cent more trips in mobile positioning data.

The last of the top 20 countries, Portugal, has just over 100 000 trips according to the survey. This estimate is based on 27 interviews of people, who reported a trip to Portugal during 2017. The average weight of each trip was more than 3 700. An estimate based on 27 observations is not very accurate and subject to randomness. In other words, the Finnish Travel –survey does not succeed very well in estimating Finnish tourism to Portugal.

Summary of the top-down approach

The main findings from the top-down approach are the following:

1. The data from two different operators are very consistent between each other. The monthly seasonality is nearly identical and correlates well with outbound tourism based on Finnish Travel –survey.
2. Depending on country of destination, the top-down approach seriously over- or underestimates the total number of outbound tourism trips to that country. There are many known sources of bias in mobile data: non-tourism trips, border noise, devices switched off, multiple devices, transit corridors etc.
3. Mobile positioning data provides a better estimate on the monthly seasonality of outbound tourism. The monthly estimates of Finnish Travel -survey are affected by randomness due to small sample size.

The strengths and weaknesses of both data sources can thus be summarized as follows:

Table 1 – Strengths and weaknesses of survey and mobile positioning data

	Finnish Travel -survey	Mobile positioning data
Strengths	**Scope** is clean: only tourism trips are included **Provides supporting information** of the trip (ie. purpose of trip, expenditure, means of transport and accommodation)	**Granularity:** millions of observations covering nearly all destination countries **Monthly seasonality** of tourism is more accurate.
Weaknesses	**Granularity;** very few observations per year, covering only a few destination countries **Monthly seasonality** estimates are affected by randomness	**Scope** is not clean, there are many sources of over- or underestimation **No supporting information** of the trip

Combining the data sources for improved outbound tourism statistics

The aim of this chapter is to outline a method for providing more accurate outbound tourism estimates for each country. The method aims to combine the strengths of both the Finnish Travel -survey and MNO data.

Methodological requirements

The proposed method for recalibrating the outbound trips data has to provide at least the following estimations:

1. Annual number of all outbound trips
2. Monthly seasonality of outbound trips to each country
3. Annual number of outbound trips to each country
4. Year-on-year change in the number of outbound trips

For annual number of outbound tourism trips (1.) the Finnish Travel -survey provides a solid estimate as shown earlier. For monthly seasonality (2.) the MNO data is more robust as it is not affected by survey randomness. For trips to each country (3.), the best data source depends on the country of destination.

At present, it's not possible to evaluate the year-on-year change as MNO data is available only for 2017. The method can be further improved in this area as 2018 data is collected from the MNOs. It is expected, that due to the higher granularity of the MNO data, the absolute year-on-year change percentages will be smaller and closer to reality compared to the survey. In this case, the MNO data is most likely a better source for measuring the annual change in the number of trips to each country.

The selection criteria for choosing the best country specific trip estimate

The accuracy of statistical estimates is often expressed using the concept of confidence interval. The 95% confidence interval has a lower and upper limit meaning that there is only a 5% chance that the actual figure is outside of this range. For example, the Finnish Travel –survey estimates that there were 100 000 trips to Portugal. This estimate has a 95% confidence interval with a lower limit of 63 000 and an upper limit of 140 000. In the case of Portugal, the estimate from MNO data is 72 000 trips, placing it within the confidence interval. All outbound destination countries can be analysed in a similar fashion.

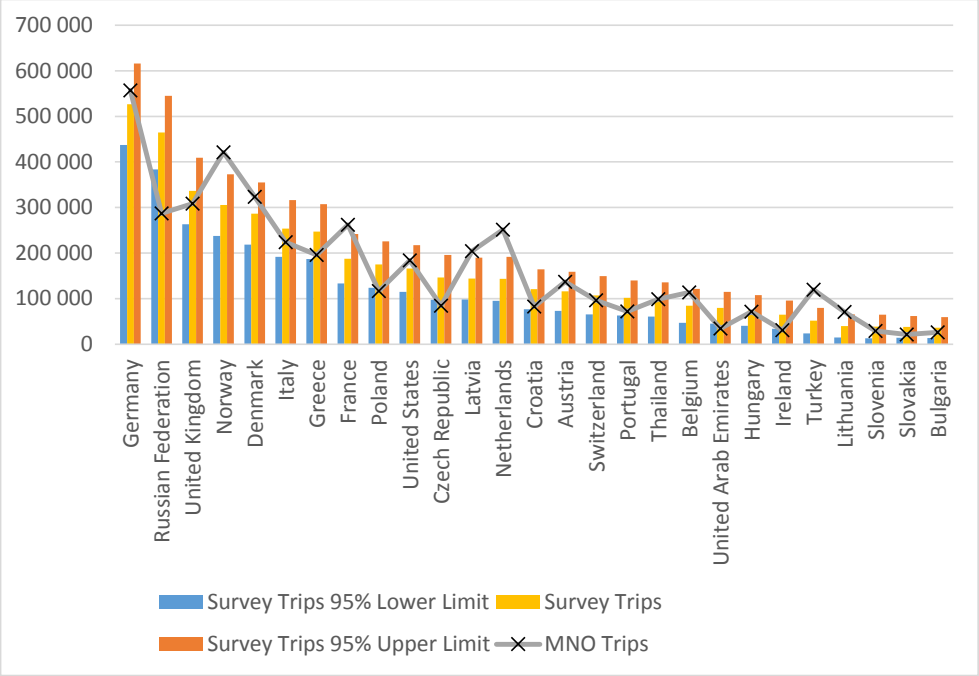

Figure 7 – 95 % Confidence intervals for top 30 destination countries (excluding top 3)

The figure presents the number of outbound trips to each country based on the Finnish Travel –survey as well as the upper and lower bounds of the 95% confidence interval. The number of trips based on MNO data is plotted against them as a line graph.

Using this kind of method, the MNO data provides a 'second opinion' to the survey confidence intervals for each country. In case the MNO trips are outside of the confidence interval for a certain country, the MNO data most likely includes serious sources of over- or underestimation for that country.

On the other hand, the survey estimates also become rather useless if the 95% confidence interval is too large. At present, countries with less than 170,000 trips have 95% confidence interval limits of plus or minus 30 per cent. For most of such small destination countries, the MNO trips can still provide a better estimate, as long as the MNO trips are within the confidence interval.

In light of this, the proposed criteria, for choosing the data source representing total annual trips to each country, could follow such guidelines:

1. **Use survey data** for all countries with more than 200,000 trips annually. This estimate is typically based on more than 50 survey interviews and has a 95% confidence typically in the range of 150,000 to 250,000 trips.
2. For countries with less than 200,000 trips (or 50 interviews) there are two options
    a. In case the number of trips in MNO data is outside of the 95% confidence interval of the survey trips, **use survey data** as there are probably major over- or underestimation issues in the MNO data for that country.

b. In case the MNO trips of trips is within the 95% confidence interval of survey trips, **use MNO data** as it provides a better estimate compared to the survey data with a large confidence interval for this country.

While this type of approach is by no means perfect, it brings certain benefits such as being able to provide outbound tourism estimates to most countries in the world instead of just the top 10-20 destination countries.

Recalibration of weights for annual trips per country

Currently the survey data is weighted based on sample size and response rates and stratified based on three different criteria: gender (male/female), age (15-84 in 10 year groups) and place of residence (5 different regions). These weights currently range from under 2,500 to over 6,000 depending on the respondent.

These weights should be recalibrated based on a coefficient factor calculated for those countries where MNO data is used. The annual coefficient factor is calculated as:

$$Coefficient = \frac{Annual\ outbound\ trips\ in\ MNO\ data}{Annual\ outbound\ trips\ in\ tourism\ survey}$$

The end result of this recalibration is summarized in the table below:

Table 2 – Summary of recalibration

	Estimated based on survey data	Estimated based on MNO data	Total
Number of trips by country	9,02 million	1,46 million	10,48 million
Number of destination countries	24	129	153

In other words, outbound trips to 24 countries would continue to be estimated based on the Finnish Travel -survey and they represent 86% of all outbound trips. Meanwhile, 129 countries will be entirely based on MNO data but they represent only 14% of all outbound trips. The total number of all outbound trips will remain at 10.5 million in 2017.

Seasonality adjustments

Finally, the monthly seasonality of outbound tourism is adjusted for each country based on mobile positioning data. This is done by calculating the share of each month and applying this monthly seasonality pattern over the seasonality obtained from the survey, which is affected by randomness. This provides a more reliable picture of tourism seasonality and takes advantage of the granularity available in the MNO data. The over- or underestimation inherent in the MNO data does not play that big a role here, since the number of trips is based on the survey for those countries that.

The following graphs present the monthly seasonality for selected destination countries before (Survey) and after (MNO) the seasonality adjustments. The name of the country is followed by (n), which is the number of interviews the survey data is based on.

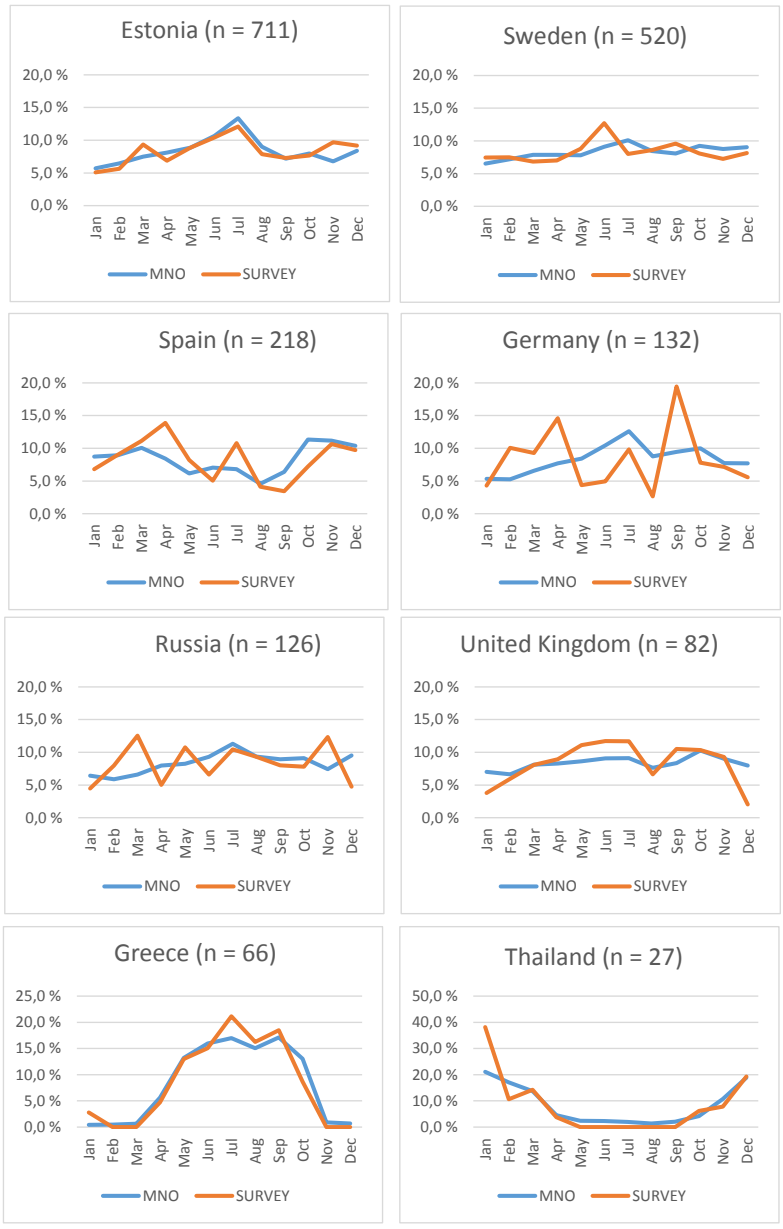

Figure 8 - 2017 monthly seasonality before (Survey) and after (MNO) adjustments

Outbound tourism seasonality is highly country specific and this is well reflected in the MNO data. Greece is a summer destination with most of trips taking place between May and October. Thailand is almost the opposite, with most trips in the winter season between November and April. Spain falls somewhere between these two as many trips are made to the Canary Islands during the winter months. Seasonality to some countries is relatively flat with trips taking place throughout the year. This is the case with Sweden and the UK.

The survey data is also able to present a rough view of this seasonality. In some cases, such as Greece, this picture is almost perfect. But in most cases, the survey is affected by randomness due to small sample sizes. The randomness causes big monthly changes with estimates going up and down from month to month. This is the case especially with Germany and Russia but also with other countries to a lesser extent. This type of random seasonality doesn't appear credible to the data user and doesn't fit well with what is generally known about tourism seasonality. In these cases, the MNO data provides a more accurate estimation of the tourism seasonality.

Summary

The process steps in the proposed method for improving outbound tourism estimates can be summarized as follows:

1. Use the **total number of annual outbound trips** from the (Finnish Travel) survey as a baseline
2. Determine the lower and upper limits of **the 95% confidence interval** for each destination country
3. **Select the most reliable source data** (survey or MNO data) for outbound trips separately for each destination country based on the confidence intervals
4. Calculate **a coefficient factor** for countries that will be based on MNO data
5. **Apply the monthly seasonality** trend based on MNO data separately for each country

The results from the method are summarized by country in the following table including top 50 destination countries that make up 98 per cent of outbound tourism. *Outbound trips* displays the number of trips based on the Finnish Travel –survey along with the lower and upper limits of the 95% confidence intervals. *MNO Top-down Outbound Trips* present the number of trips according to the top-down method. *Final Trips* is either based on the survey or MNO data, depending on which is the more reliable source. *Monthly seasonality* presents the adjusted seasonality trend for each country.

Using the outbound tourism to 24 countries (86 % of trips) is estimated based on survey and 129 countries (14 % of trips) are entirely based on MNO data. The seasonality patterns appear logical based on what is generally known about the tourism to each country. The winter destinations stand out from the summer destinations, while some countries have less seasonality.

Table 3 – Summary of results for 2017

Country	Finnish Travel -survey			MNO			Monthly Seasonality (from MNO)
	Outbound Trips (000)	95% CI Lower Limit	95% CI Upper Limit	Top-down Outbound Trips (000)	Selected Source	Final Trips	
TOTAL OUTBOUND TRIPS	**10 484**	**10 135**	**10 834**	**10 484**		**10 484**	
ESTONIA	2 695	2 519	2 871	2 036	STAT	2 695	
SWEDEN	1 991	1 833	2 150	2 865	STAT	1 991	
SPAIN	822	714	930	638	STAT	822	
GERMANY	526	437	616	557	STAT	526	
RUSSIAN FEDERATION	464	384	545	287	STAT	464	
UNITED KINGDOM	336	263	409	309	STAT	336	
NORWAY	305	238	373	421	STAT	305	
DENMARK	287	218	355	323	STAT	287	
ITALY	254	192	316	224	STAT	254	
NETHERLANDS	144	95	192	251	STAT	144	
GREECE	247	187	307	196	STAT	247	
LATVIA	144	98	190	204	STAT	144	
FRANCE	188	133	242	262	STAT	188	
UNITED STATES	166	115	217	184	MNO	195	
AUSTRIA	116	73	159	137	MNO	146	
TURKEY	52	24	79	120	STAT	52	
POLAND	175	124	226	116	STAT	175	
BELGIUM	84	47	122	114	MNO	121	
THAILAND	98	60	136	99	MNO	105	
SWITZERLAND	107	65	150	96	MNO	102	
CZECH REPUBLIC	146	97	196	85	STAT	146	
CROATIA	121	77	164	83	MNO	88	
PORTUGAL	101	63	140	72	MNO	76	
HUNGARY	74	40	108	71	MNO	76	
LITHUANIA	40	15	65	71	STAT	40	
CANADA	15	0	30	46	STAT	15	
UNITED ARAB EMIRATES	80	45	115	34	STAT	80	
IRELAND	65	34	96	31	STAT	65	
ICELAND	23	2	43	29	MNO	31	
SLOVENIA	39	13	65	29	MNO	31	
BULGARIA	36	13	59	26	MNO	28	
JAPAN	17	0	34	26	MNO	27	
CHINA	22	4	40	23	MNO	25	
SLOVAKIA	38	14	62	21	MNO	22	
LUXEMBOURG	8	0	18	20	STAT	8	
MALTA	32	9	54	20	MNO	22	
SINGAPORE	13	0	27	19	MNO	20	
HONG KONG	5	0	14	17	STAT	5	
CYPRUS	9	0	20	16	MNO	17	
ISRAEL	27	7	47	14	MNO	15	
INDONESIA	23	3	42	14	MNO	15	
AUSTRALIA	32	8	56	14	MNO	15	
GIBRALTAR	0	0	0	12	MNO	13	
ROMANIA	16	0	31	12	MNO	13	
MONTENEGRO	22	4	40	12	MNO	12	
UKRAINE	24	4	45	10	MNO	11	
INDIA	20	2	38	10	MNO	11	