Same-day visitors crossing borders: a big data approach using traffic control cameras

M. Izquierdo Valverde[1], J. Prado Mascuñano[2], M. Velasco Gimeno[3]

[1] *National Statistical Institute (INE), Madrid, Spain; maria.izquierdo.valverde@ine.es*
[2] *National Statistical Institute (INE), Madrid, Spain; jesus.prado.mascunano@ine.es*
[3] *National Statistical Institute (INE), Madrid, Spain; maria.velasco.gimeno@ine.es*

Abstract

One pilot project recently implemented in INE (Spain) using big data is related to the task of building the frame of visitors crossing the borders by road. Due to Schengen Area, there is no direct control of people crossing the borders with France and Portugal.

The register of traffic loops provides the number of vehicles for each crossing-road, in both ways (going in and out of Spain), per quarter of an hour and according to vehicles length. This information is completed by the traffic control cameras installed in the border lines. The last is a huge database of number plates of vehicles crossing the borders. Combining both sets of big registers (responsibility of the Traffic Control Directorate), INE builds the population frame to estimate the number of foreign vehicles that enter in Spain monthly, broken down by nationality.

An important methodological challenge currently under research is the assessment of the potential of the cameras database to estimate international same-day visits.

These estimations are difficult to obtain from traditional outbound tourism sample surveys, as international same-day visitors concentrate along the borders, while sample design of these surveys (with great potential for mirror-statistics) has more general objectives, and there is no subsample big enough to characterize these trips. Besides, due to the difficulties and high costs of conducting surveys at road borders the quality of the estimations of inbound same-day trips could be enhanced with this new information.

Tracking an anonymized number plate through the camera registers database is the first step of the analysis. Those databases contain for each vehicle the number plate, the frontier point, the nationality of the plate and the moment when it is crossing the border.

Defining a same-day trip in terms of number of hours (less than 24 and more than 3) between the moment of going out (coming in) Spain and the moment of coming back (going back), and extending the tracking to detect vehicles that go in and out through different points, we can produce estimates of the ratios of same-day trips in respect of the total frame numbers. Thus we can derive subtotals to analyze with more precision this specific subset of tourism. The first pilot experiences presented in the paper produce the inputs for analysis and decision making in order to adjust parameters as border points,

segments and time intervals choices, to approximate an optimum for implementation.

Keywords: big data, traffic loops, traffic control cameras, same-day visitors

1. Introduction

Tourism is one of the most dynamic industries in many economies. According World Tourism Organization (WTO)[1], it represents 10% of the world GDP, it generates one in eleven jobs and international tourist arrivals have increased 4% to 1.2 billion in 2015. Estimation of tourism flows come mainly from border crossing and accommodation statistics, and from household surveys to resident population. The interregional component of this phenomenon makes comparability an essential feature of the reliability of data, which has been developed within the frame of WTO and Eurostat manuals, guides and recommendations. Thus, these sources of primary information, which have been providing data for many years, have a strong methodological base and fulfil high quality standards.

However, in a world where mobility has increased to its highest levels in few years and border controls have disappeared in neighbouring areas, such as the Schengen Space in Europe, border crossing surveys are becoming more costly and difficult to conduct, and many countries are looking for alternatives and complementary information.

In this context, data generated not from purely statistical sources but from events intimately linked to the tourism phenomenon appear as a source of information that can improve the relevance, opportunity and punctuality of the products offered under the quality standards of official statistics. Examples of these new data sources are registers from traffic loops and traffic control cameras capturing flows of vehicles, records of mobile phones travelling from one place to another, activity of credit cards during a trip, among others.

This paper focuses the attention on registers from traffic loops and traffic control cameras and the opportunity of getting information on same day visits of non-residents in Spain visiting our country.

Although the analysis is carried out on inbound tourism, similar studies are being developed to take advantage of traffic control cameras to improve the estimation of same day visits from the outbound tourism point of view.

[1] Source: International Tourism Arrivals infographics, WTO (http://media.unwto.org/content/infographics)

2. Statistic of Tourist Movements on Borders (FRONTUR)

The statistic of Tourist Movements on Borders (FRONTUR) is carried out by INE to reach the following objectives:

- To measure the number of non-resident visitors coming into Spain each month, distinguishing the different access routes (road, airport, port and railway), further distinguishing between tourists (those who make at least on overnight stay in Spain) and excursionists (one-day visitors, who do not stay overnight).
- To describe the main characteristics of the trips made by those visitors: main destination in Spain, type of accommodation, country of residence, purpose of travel, the organization of the trip (with or without a package tour).

This statistical operation follows the definitions, classifications and best practices recommended by UNWTO, which are described in the documents *International Recommendations for Tourism Statistics 2008 (IRTS-2008)* and *Tourism Satellite Account. Recommendations on the Methodological Framework 2008.* Similarly, *Eurostat Methodological Manual for Tourism Statistics* is also used as reference.

The methodology of this survey can be consulted in the following link:

Methodology FRONTUR

The topic of this paper is focused in the sub-operation carried out when the borders are crossed by road. Specific characteristics of the sub-operation are described now:

2.1. Sample framework: the theoretical population framework is people crossing borders (by road) into Spain. There is no framework to select the sample in the traditional sense used in sample surveys. INE have access to information of administrative records and databases managed by different agencies or originations, which are the basis for estimating flows of entry into our country. In the specific case of roads, the General Directorate of Traffic (DGT) provides information on the number of vehicles entering and leaving all border points. These data come mainly from traffic loops installed at these points; the loops count vehicles and classify them according their length as small, medium and large. Furthermore, this information is contrasted with data from video cameras that are installed at major points of entry into our country.
In order to estimate the number of persons crossing the border, a sample operation called *Aforo*, is carried out to estimate the number of travellers entering Spain by type of vehicle and nationality (according number plates).
More details of this process are explained in chapter 3.

2.2. Sample size: the theoretical sample size is more than 85.000 surveys per year, distributed by stratum and month, taking into account also the type of visitor (tourist and day tripper) and nationality (11 nationalities are considered).

3. Traffic loops and traffic control cameras

The first experience of INE using big data in tourism statistic is related to the task of building the frame of people crossing the borders by road. Due to Schengen Treat, there is no control over people that cross the border from France or Portugal to Spain (and vice versa).

The register of traffic loops provides the total number of vehicles that cross the border for each crossing-road, in both ways (going in and out of Spain) by hour and classifying the vehicles according their length (short, medium and large). This information is completed by the traffic control cameras that are installed in the border lines (both registers are managed by Traffic General Direction). It is a complete database of number plates of vehicles that come into our country. Combining both sets of big data we can estimate the number of foreign vehicles that enter in Spain monthly broken down by vehicle nationality.

The next step to know the number of persons that come into our country is transforming the *Vehicles Frame* in a *Travelers Frame*. To get this aim, sample data of vehicles by type of vehicle, nationality (of number plate) and number of occupants per vehicle are collected (this sample operation has been mentioned in 2.1, *Aforo)*. Using the collected information, an occupancy rate of vehicles, by type of vehicle and nationality is calculated. Mixing both data the *Travelers Frame* mentioned before is calculated.

This is the general schema that is carried out to get this basic information to estimate the number of foreign visitors (tourist and same-day visitors) that come to Spain every month by road. In this case we don't have to face problems related to different definitions used in the register of traffic loops and traffic control cameras, vehicles crossing border is the counted unit in both registers. But the coverage of these sets of data sometimes is not exactly the same, due to technical problems that are being solved.

Tracking an anonymized number plate through the camera registers database will allow new studies about same-day visitors.

4. Same-day visits using camera registers

The camera registers include the following information for each vehicle:

- Border crossing
- Year
- Month
- Day
- Hour
- Number plate
- Nationality of the number plate
- Direction of the road (in or out)
- Validity

For this analysis the vehicles coming into Spain by one border crossing are selected, and tracking their number plates in the registers of vehicles going out the country (by any border crossing) during the same day.

For each border crossing, *i*, the number of vehicles coming into Spain counted by cameras (VC_i) is:

$$VC_i = ECNR_i + TCNR_i + ECR_i + TCR_i + NOIDEF_i$$

Where:

$ECNR_i$ = number of vehicles classified as *same-day visits* with foreign-number plate coming into Spain by border crossing *i*.

$TCNR_i$ = number of vehicles classified as *non-same-day visits* with foreign-number plate coming into Spain by border crossing *i*.

ECR_i = number of vehicles classified as *same-day visits* with Spanish-number plate coming into Spain by border crossing *i*.

TCR_i = number of vehicles classified as *non-same day visits* with Spanish-number plate coming into Spain by border crossing *i*.

$NOIDEF_i$ = number of vehicles with non-identified number-plate

$$VC_i = ECNR_i + TCNR_i + ECR_i + TCR_i + NOIDEF_i$$

5. Analysis

For each border crossing, *i*, the number of estimated vehicles by the traditional method in FRONTUR (VF_i) is:

$$VF_i = LFNR_i + PFNR_i + LFR_i + PFR_i$$

Where:

$LFNR_i$ = number of small vehicles with foreign-number plate coming into Spain by border crossing *i*.

$PFNR_i$ = number of big vehicles with foreign-number plate coming into Spain by border crossing *i*.

LFR_i = number of small vehicles with Spanish-number plate coming into Spain by border crossing *i*.

PFR_i = number of big vehicles with Spanish-number plate coming into Spain by border crossing *i*.

By the other side, for each border crossing, *i*, the number of vehicles counted by cameras (VC_i) is:

$$VC_i = ECNR_i + TCNR_i + ECR_i + TCR_i + NOIDEF_i$$

To integrate the data on vehicles according their size with the data from cameras (by type of visit and nationality of number plate), we assume, in this first phase of the study, that $NOIDEF_i$ are always vehicles with foreign-number plate.

Now the structure of *same-day visits* with foreign-number plate is going to be applied to $VF_i - RF_i$ (vehicles with foreign-number plate estimated by the traditional method):

$$\begin{aligned} VCAJ_i &= VF_i - RF_i \\ &= ECNR_i \times \frac{VF_i - RF_i}{ECNR_i + TCNR_i + NOIDEF_i} + TCNR_i \\ &\times \frac{VF_i - RF_i}{ECNR_i + TCNR_i + NOIDEF_i} + NOIDEF_i \\ &\times \frac{VF_i - RF_i}{ECNR_i + TCNR_i + NOIDEF_i} = ECNRAJ_i + TCNRAJ_i + NOIDEFAJ_i \end{aligned}$$

Then

$$VF_i = ECNRAJ_i + TCNRAJ_i + NOIDEFAJ_i + RF_i$$

The next step is to integrate this estimation with the breakdown of vehicles according the size:

$$VF_i = ECNRAJ_i + TCNRAJ_i + NOIDEFAJ_i + RF_i = LFNR_i + PFNR_i + RF_i$$

A ratio to calculate the tendency of same-day visits in big vehicles has to be defined. Implicitly, a ratio of the tendency of same-day visits in small vehicles is also calculated. For each border crossing, *i*, the ratio is:

$$RPEX_i = \frac{ECNRAJ_i}{ECNRAJ_i + TCNRAJ_i} \times \beta_i$$

where β is a correction factor that shows the greater tendency of same-day visits in big vehicles. Then, the estimation of big vehicles with same-day visitors in *i,* is:

$$EPNR_i = RPEX_i \times PFNR_i$$

If there would be a case where $EPNR_i > PFNR_i$, then it would assume that $RPEX_i = 1$

Now, the number of big vehicles classified as *non-same day visits* can be calculated by difference:

$$TPNR_i = PFNR_i - EPNR_i$$

Once calculated those estimations for big vehicles, the ones for small vehicles can be estimated. It is assumed that non identified number plates are foreign vehicles and that they are brokendown by same-day visits and non-same day visits with the same structure as the small vehicles:

$$ELNR_i = \left(\frac{ECNRAJ_i}{ECNRAJ_i + TCNRAJ_i} \times (ECNRAJ_i + TCNRAJ_i + NOIDEFAJ_i) \right) - EPNR_i$$

Then

$$TLNR_i = LFNR_i - ELNR_i$$

With the described process, for each border crossing it is available the total number of vehicles coming into Spain broken down by resident and non-resident, and same day visits and non-same day visits as follows:

$$VF_i = TPNR_i + EPNR_i + TLNR_i + ELNR_i + RF_i$$

Following a similar reasoning, the number of vehicle with Spanish number plate could be brokendown by the same classification.

6. Work in progress

Till this moment, a very theoretical explanation has been described based in hypotheses that imply that the registers (both, traffic loops and traffic cameras) are completed and coherent. But 'perfect world' doesn´t exist.

The flows of vehicles registered in both datasets are different. The devices (loops and cameras) have technical failures that produce lack of information.

Nowadays INE is working and carrying out different tests based in different hypotheses to solve these situations.

The results of these test will be shown in the presentation in the 14th Global Forum on Tourism Statistics.