Border surveys and Time Location Sampling (TLS): an application on incoming tourism in Sicily

Stefano De Cantis, Mauro Ferrante

Abstract   The increasing need of reliable and detailed information at a regional and subregional level poses some challenges related with the measurement of tourists flows, for tourism destination management, planning and promotion purposes,. The aim of this work is twofold; first, it intends to analyse and discuss the main features and limits of the official border surveys conducted in the majority of world's countries, within the framework of Time Location Sampling (TLS) design; second, it suggests the implementation of TLS design at a subnational level, by presenting the sampling design of the survey on incoming tourists in Sicily, carried out between 2009-2010, aimed at estimating the unobserved tourism and at analysing tourist mobility in the island

Keywords: Frontier survey, mobile population, time-space sampling, tourism statistics

1   Border surveys and Time Location Sampling

The importance of gathering data on international visitors at frontier is confirmed by the number of countries using border surveys, that has virtually tripled over the past twenty years. Border surveys are particularly important, since the information on foreign tourists deriving from supply-side statistics only give a partial coverage of the phenomenon, due to several problems which affect accommodation statistics (e.g. the use of relative or friends houses, the use of rented and not registered accommodations, the double counting effect, the lack of information about the purpose of stay and on expenditure, etc., see UNWTO, 2005:56-57; Vaccina, Parroco, 2005).

The limitations of the actual tourism statistics, in general, and of border surveys, more particularly, affect reliability and comparability of official statistical information in tourism,

---

[1]   Stefano De Cantis, Mauro Ferrante Dipartimento di Scienze Economiche e Finanziarie, Università degli studi di Palermo; email: stefano.decantis@unipa.it;mauro.ferrante@unipa.it; *The paper is a common work of both the authors; however, sections 1 is due to Mauro Ferrante, and section 2 is due to Stefano De Cantis.*

as also highlighted by the "Comparative study of international experiences in the measurement of travel flows at national border" made in 2003-2004 on forty-eight world's countries by the World Tourism Organization (UNWTO, 2005). The study analyses the main features and technical details of the methodologies used in the estimation of the international tourism flows. Based on this comparative study, on an analysis of border surveys methodologies, and on our practical experience, the most critical aspects to face with when designing and conducting a border survey can be summarized as follows: a) the selection of entry/departure points; b) the choice of the time spread and the periodicity of information collection; c) the moment in which the interview has to be made (when tourists enter or leave the country); d) the definition of the main goals of the survey and of the main variables of interest; e) the choice of the sampling design strategy, e.g. number of stages, stratifications, stratification variables, etc.; f) the choice of the survey instrument (face-to-face interview, CAPI, etc.); g) other only apparently less important aspects which characterize the research design, such as the techniques adopted to approach tourists, the choice and the availability of prior information to define the sampling design; h) finally, the statistical procedures of estimation (the estimator used, the definition of selection probabilities, the determination of sampling weights, the estimation of standard errors, the calibration techniques, etc.). All these technical and practical difficulties are mainly due to the special nature of tourists population, which is by definition a mobile population. As for many other mobile or hard-to-reach populations, one of the main problem in sampling is represented by the absence of a complete list of the units. These issues, which are discussed mainly in epidemiological studies (Muhib et al., 2001; Kalsbeek, 2003; Kakinami, Conner, 2010), have been approached through Time Location Sampling (TLS) design. TLS offers an alternative solution to the more classical schemes, since it tries to recruit respondents in places and times where they would reasonably be expected to gather and ask them about their experiences within the place or space (Kalton, 2009). TLS is used to sample a population for which a sampling frame cannot be constructed but locations are known in which the population of interest can be found, or for which it is more efficient to sample at these locations (Karon, 2005). TLS is a probabilistic method used to recruit members of a target population at specific times in set venues. The sampling framework consists of venue-day-time units (VDT) – also known as time-location units – which represent the potential universe of venues, days and times. Surprisingly, even if many of the border surveys conducted by most of the national institutions could be set within a TLS framework, in most of the border surveys reviewed, there seems to be a lack in the formalized specification of the sampling design adopted. Consequently, the estimation procedures used, do not take adequately into account, at least explicitly, for the specific sampling design. The following section shows the TLS design implementation for the survey on incoming tourists in Sicily, by paying a particular attention to the main problems faced, and the solutions found in practice.

2 Sampling incoming tourists in Sicily

The partial nature of accommodation statistics (Istat, 2011a), and the rough estimates of the foreign tourists provided by the Bank of Italy (Bank of Italy, 2011) and of Italian tour-

ists published by the Italian National Statistical Institute (Istat, 2011b) do not provide an adequate support for tourism analysis at subregional level (Vaccina, Parroco, 2005). Given these premises, in 2009-2010 a survey on incoming tourists in Sicily was performed in order to gather information on tourist mobility and to quantify the magnitude of the unobserved tourism in the island. About 4.000 interviews were collected (for details on the entire research project, sampling design, and first results see Vaccina, Parroco, 2010), and a complex TLS design was adopted. The units of interest were represented by Italian (not resident in the island) and foreign tourists leaving Sicily at the end of their vacation. In this way it was possible to collect direct information (from the demand side) related to the whole period spent in the island, by minimizing the recall bias. Considered the insularity of Sicily, according to the TLS design, almost all the places from which it is possible to leave the Island were selected: the three main airports (Palermo, Catania and Trapani airports), the ports of Palermo and Catania, and the Strait of Messina. The periods covered by the survey were selected according to official data on tourists flows in the island: spring, summer, and autumn, during which more than the 80% of the official tourists flows are concentrated. The specific TLS implementation was treated as a two-stage stratified sampling design with unequal selection probabilities for the first-stage units, and with constant selection probabilities for the second-stage units. The first-stage units are constituted by the combination of places, days and hours, i.e. Venue-day-time units (VDT, i.e. primary sampling units: PSU). Let be M the total number of VDT. The second-stage units are constituted by tourists, which were selected within the first stage units through a systematic selection procedure. An integration of the official data coming from different statistical sources was used: tourism surveys from the demand side (Istat, 2011b), from the supply side (Istat, 2011a), data collected by the Bank of Italy on international tourism (Bank of Italy, 2011), data on daily air passengers and on daily ferries leaving the Strait. First-stage units selection probabilities of the VDT units (for the i-th place, and the j-th time) were consequently determined (i.e. $P_{i,j}$). For the second stage units, a pseudo-probabilistic approach was adopted, through a pseudo-systematic selection of the units in the days and places selected, according to strict rules given to the interviewers (De Cantis, Ferrante, 2011). After the survey, ex-post information on number of passengers within each VDT unit selected were asked to official authorities and managers of the different places selected (e.g. GESAP s.p.a. for Palermo Airport, SAC s.p.a. for Catania Airport, AirGEST Trapani Airport, and the Port Authorities for the Port of Palermo, Catania, and Messina). Considered the information on the number of people travelling in each VDT selected, a monthly estimate of the proportion of incoming tourists and of other travellers categories was used to determine the number of tourists within each VDT selected (i.e. second-stage units population), and, subsequently, the second-stage units selection probabilities ($\pi_{h|i,j}$, i.e. the probability of selecting the h-th tourist within the selected $VDT_{i,j}$) (for a more detailed description on the auxiliary information, see De Cantis, Ferrante, 2011). Once obtained the set of $y_{h,i,j}$ values (where Y is the variable of interest), a direct estimator of Hansen-Hurwitz class is given by the following (for simplicity, stratification is ignored):

$$\hat{Y} = \frac{1}{m} \sum_{i=1}^{s} \sum_{j=1}^{t_i} \sum_{h=1}^{n_{i,j}} \frac{y_{h;i,j}}{\pi_{i,j} \pi_{h|i,j}} = \frac{1}{m} \sum_{i=1}^{s} \sum_{j=1}^{t_i} \frac{1}{P_{i,j}} \frac{N_{i,j}}{n_{i,j}} \sum_{h=1}^{n_{i,j}} y_{h;i,j} = \sum_{i=1}^{s} \sum_{j=1}^{t_i} \sum_{h=1}^{n_{i,j}} w_{i,j} y_{h;i,j}$$

where m represents the number of first stage units included in the sample; for each $VDT_{i,j}$ selected, $n_{i,j}$ represents the number of second stage units included in the sample, among the total $N_{i,j}$ units; finally, $w_{i,j} = N_{i,j} / (m\ P_{i,j}\ n_{i,j})$ represent the final sampling weights.

According to this estimator, the formula for standard error is well known (it is not reported here for simplicity); moreover it is particularly interesting that, for this specific sampling design (two stage sampling with unequal first stage units probabilities with replacement) the standard error does not depend on the second stage sampling design. Finally, the estimation procedure is completed by the use of calibration techniques. Particularly, the availability of auxiliary information on official establishments (i.e. arrivals and overnight stays in official establishments) allows the determination of post-stratification weights and the implementation of calibration procedures. This was possible thanks to a specific section of the questionnaire, in which arrivals and presences spent in official accommodation establishments (e.g. hotels and similar establishments), were asked to each tourist interviewed, separately to the nights spent in unofficial establishments (e.g. relative or friends houses, second houses, etc.).

In a nutshell, in the implementation of the survey on incoming tourists in Sicily, some main methodological contributions can be highlighted: the identification of the sampling design within the general framework of TLS design; the absence of qualified counting of "all the tourists" within the VDT selected (for a discussion about qualified counting see Parsons et al., 2008); the use of an estimator with well-known properties; the use of calibration techniques based on accurate ex post information on guests arrivals and overnight stays in collective establishments. The TLS implementation tried to overcome practical and technical difficulties, with relatively low expenses, and it could be repeated in other temporal and spatial contexts in order to improve the knowledge of tourists behaviours at subregional level.

References

1. Bank of Italy: Survey on Italian international tourism, (2011)
2. De Cantis S., Ferrante M.: The probabilities of selection of first and second stage units in the Time Location Sampling (TLS) design in the survey on incoming tourism in Sicily and Sardinia. In Asero V., D'Agata R., Tomaselli V. (eds), Turisti per caso?...Il turismo sul territorio: motivazioni e comportamenti di spesa, Bonanno editore, Roma, pp.27-33, (2011)
3. Istat: Capacity and occupancy in collective accommodation establishments 2010, Istat, Rome, (2011a)
4. Istat: Trips and holidays in Italy and abroad 2010, Istat, Rome, (2011b)
5. Kakinami L., Conner K.: Sampling strategies for addiction research. in Miller P.G, Strang J., Miller P.M. (eds) Addiction Research Methods, Blackwell (2010)
6. Kalsbeek W.D.: Sampling minority groups in health surveys, Statistics in Medicine, 22, 1527-49, (2003)
7. Kalton G.: Methods for oversampling rare subpopulations in social surveys, Survey Methodology, Vol. 35, n.2, pp.125-41 (2009)
8. Karon J.M.: The analysis of time-location sampling study data, Proceedings of the Survey Research Methods Section, ASA (2005)
9. Muhib F.B., Lin L.S., Stueve A., Miller R.L., Ford W.L., Johnson W.D., et al.: A venue-based method for sampling hard-to-reach populations. Public Health Report, 116 (Suppl 1), pp. 216–222 (2001).
10. Parroco A.M., Vaccina F.: Mobilità ed altri comportamenti dei turisti: studi e ricerche a confronto. Collana di Scienze del Turismo, McGraw-Hill (2010)
11. Parsons J.T., Grov C., Kelly B.C.: Comparing the effectiveness of two forms of time-space sampling to identify club drug-using young adults. J Drug Issues, 38 (4), pp. 1061–1081 (2008)
12. UNWTO: Tourism as an International Traded Service – A Guide for Measuring Arrivals and Associated Expenditures of Non-Residents, United Nations World Tourism Organization, Madrid (2005)
13. Vaccina F., Parroco A.M.: Referring To Space And Time While Using Territorial Data: The Case Of Touristic Arrivals. ISI2005, 55th session of ISI. 5-12 April 2005, Sydney, (2005)