TOURISM STATİSTİCS AND UNOBSERVED TOURİSM: EMPIRİCAL EVIDENCES İN SİCİLY

JEL classification: L83, C83

Abstract
Despite its strategic importance, accurately measuring visitor attendance has been a challenging and problematic exercise for tourism managers for decades. Consequently, the primary aim of tourism statistics of accurately quantify tourism flows has been only partially achieved; indeed at a lower territorial scale tourism statistics appear less precise and accurate. The aim of this paper is: to introduce the concept of unobserved tourism, by highlighting the main limits of official statistical systems (with a special focus on the European statistical system on tourism statistics; to formalize a theoretical model in which tourism nights and trips in a given destination are decomposed into observed and unobserved components, according to the current systems of tourism statistics. We define unobserved tourism, in terms of overnight stays, the sum of two components: the set of all the nights spent by tourists in unofficial establishments (unmeasured tourism), and the set of nights spent by tourists in official establishments, but deliberately concealed from public authorities, mainly for fiscal reasons (underground tourism). Some empirical evidences derived from surveys aimed at quantifying the magnitude of unobserved tourism in Sicily are described in order to illustrate the different approaches which can be adopted to explore the issue of unobserved tourism.

Keywords: accommodation statistics, sampling tourists, tourism indicators
1. **INTRODUCTION**

Having more and reliable statistics is essential for policy-makers to make effective decision, for designing marketing strategies, evaluating the efficiency and effectiveness of management decisions, and measuring tourism throughout the regional/local economy. In the last decades there was a growing awareness that the weakness of the statistical data in tourism needed some major initiatives. However, despite the efforts demonstrated by national and international institution (WTO, 1994; European Communities, 1994; UNWTO, 2008; European Parliament, 2011), for improving the reliability and the comparability of statistical information on tourism, current statistics produced by national institutes seem to be still inadequate for destination management purposes, mainly at a local (sub-regional) level. Moreover, the increasing importance of tourism in many urban and rural areas has called into question the adequacy of official statistical sources for specific local planning needs. To date, to answer satisfactorily to an apparent simple question such as “how many tourists visited in a given year a certain destination?” is still an open issue, both under the theoretical and the applied perspective, since simply counting the number of tourists in a destination is not as simple as one might initially think (Smith, 1995:16). At European level, the partial inadequacy of tourism statistics is demonstrated by the recent new Regulation (EU) No 692/2011 of the European Parliament and of the Council, concerning European statistics on tourism, which tries to establish a common framework for the systematic development, production and dissemination of European statistics on tourism (European Parliament, 2011: art.1). The new Regulation repeals Council Directive 95/57/EC and, on the one hand, highlights that the Union’s tourism industry occupies an important place in the economy of the Member States, with tourist activities representing a large potential source of employment. On the other hand, the Regulation affirms that any appraisal of its competitiveness requires a good knowledge of the volume of tourism, its characteristics, the profile of the tourist and tourism expenditure and the benefits for the economies of the Member States. It appears that due to: a) the growing importance of short trips and same-day visits contributing substantially in many regions or countries to the income from tourism, b) the increasing importance of non-rented accommodation or accommodation in smaller establishments, and c) the growing impact of the Internet on the booking behaviour of tourists and on the tourism industry, the production of tourism statistics should be adapted and the recommendation 95/57 CE overcome. However, the weakness of tourism statistics highlighted by the European Parliament Regulation, and by several other authors (Lickorish, 1997; Vaccina, et al., 2011), are not only due to the partial inadequacy of methodologies for the collection of information of the different Member States, since they are related also to the complex nature of tourism phenomenon itself, and many of these issues still need to be overcome, under the logical and the methodological perspective. The aim of this paper is to introduce the concept of un-observed tourism and the problems related to its estimation. Next section discusses the main limits of official statistical sources in Europe, and it defines the concept of
un-observed tourism, by formalizing a conceptual framework for un-observed tourism, both in terms of overnights, and in terms of tourists. The approaches undertaken in several empirical researches aimed at estimating the magnitude of unobserved tourism in Sicily and the main results are presented in the third section. Final comments conclude this work.

2. THE TRIPS-ARRIVALS (T-A) MODEL FOR THE ESTIMATION OF TOURISM TRIPS AT LOCAL LEVEL

The collection of information on tourism, at European level, is related both to the demand side, and to the supply-side of tourism market. However, statistical sources on the demand side do not give any information at sub-regional level, since they are based on sampling surveys that are not designed to give local information. This imply that the only available local information are provided from the supply-side statistics on guests in collective establishments. However, they are affected by several problems: first, no information on the motivation of the stay is collected from the supply-side, making it impossible to distinguish tourists from other guests (e.g. seasonal workers, students, etc.). Second, not all tourists stay at collective accommodations, and those who do not might have very different patterns of behaviour than those who do. Some kind of accommodations (e.g. non-collective, and private accommodations), in fact, are not included in the survey from the supply-side at all, such as second houses, vacation houses, boats, relatives and friends houses, and so on (Hall, Müller, 2004; Gallent, Tewdwr-Jones, 2000). We will call this component of tourism demand “unmeasured tourism” (Parroco, Vaccina 2004), according to the terminology used in the field of un-observed economy (OECD, 2002). Third, as for many other economic activities, accommodation manager may choose to declare only part of their guests in order to avoid direct or indirect taxation. We will call this component “underground tourism” (Parroco, Vaccina 2004). Fourth, visitor while on a trip might stay in more than one collective accommodation, resulting in an overestimation (i.e. the “double counting effect”) of the number of visitors and an underestimation of the total duration of the visit within the destination considered (Pearce, 1995; Lickorish, 1997; Parroco, Vaccina 2004). Given these considerations, we formalize a conceptual model of actual tourism in a destination/region, by expressing the above problems in terms of parameters and/or quantities to be estimated.
2.1. **The Nights-Presences (N-P) equation**

The unavailability of direct information (derived from demand-side surveys) on tourism flows at sub-regional level, determines the habits of using supply-side statistics in order to quantify the magnitude of tourism at a destination/local level. Indeed, the above cited problems which affects statistics on guests in collective establishments led us to define the concept of unobserved tourism, as that part of tourism which cannot be measured through supply-side statistics. In order to formalize the differences between the number of guests in collective establishments and the number of tourists in a given destination, the following framework aims at highlighting the main differences between guests and tourists in terms of quantities and parameters that need to be estimated.

Let \( \text{obs} P_{i,t} \) be the total number of nights spent by guests in official collective establishments in the \( i \)-destination/region, in the time interval \( t \) (e.g. one year), some of these nights can be made by tourists (in a proportion equal to \( \alpha_0 \)), and some other by other kind of guests (e.g. seasonal workers, crews on public modes of transport, students, etc.) according to UNWTO definition of classification of inbound travellers (UNWTO, 2008:18). So the nights spent by tourists in official establishments (\( \text{obs} N_{i,t} \)) will be equal to \( \alpha_0 \text{obs} P_{i,t} \); where \( 0 \leq \alpha_0 \leq 1 \).

As above told, the un-observed tourism can derive both from the nights spent in un-official establishments (e.g. private houses, boats, second houses, etc.), and from the nights concealed from public authorities mainly for fiscal reason (OECD, 2002, 13). We will call the former component “unmeasured tourism” (\( \text{unm} N_{i,t} \)), and the latter “underground tourism” (\( \text{und} N_{i,t} \)) (Parroco, Vaccina, 2004). These two components constitute what we call un-observed tourism (\( \text{unobs} N_{i,t} \)). So the actual number of nights spent by tourists in a given destination/region \( i \), in the time interval \( t \) considered, would be equal to:

\[
\text{tot} N_{i,t} = \text{obs} N_{i,t} + \text{unobs} N_{i,t} = \alpha_0 \text{obs} P_{i,t} + \text{unm} N_{i,t} + \text{und} N_{i,t} \tag{1}
\]

where \( \text{unobs} N_{i,t} \) represents the un-observed nights spent by tourists in the destination/region \( i \), during the time interval \( t \). For simplicity we will call this expression as the Nights-Presences (N-P) equation. However, if we consider the available official information provided, only the first aggregate is known \( \text{obs} P_{i,t} \). The motivation coefficient (\( \alpha_0 \)), the unmeasured, and the underground components are unknown, and need to be estimated (for a brief review of the methods proposed for the estimation of these components, see Vaccina et al., 2011).
2.2. The Trips-Arrivals (T-A) equation

If for some purposes it could be more important to know the number of nights spent by tourists in a destination/region, for many planning and management issues, it is essential to know the number of tourism trips made, in a given time interval, in the destination/region considered. The problem of converting available information on guests arrivals in trips, is not only related with the lack of the information on guests’ motivations, but also with the implications of tourists mobility. To convert guests into trips, it is necessary to introduce also a coefficient which take into account for the average number of establishments (official and un-official) used by tourists during their visit within the destination/region considered.

Let be \( G_{i,t}^{obs} = \alpha (A_{i,t}^{obs}) \) the number of tourists arrivals registered in official accommodation establishments (where \( 0 \leq \alpha \leq 1 \) represents the proportion of guests arrivals \( A_{i,t}^{obs} \) with touristic motivations), in the \( i \)-destination/region, during the time interval \( t \); and let be \( \beta \) the average number of establishments used by tourists during their visit within the destination/region considered \( (\beta \geq 1) \). The number of tourism trips \( (TRIPS_{i,t}) \) in the destination/region \( i \), during the time interval \( t \), would be equal to:

\[
TRIPS_{i,t} = \left( G_{i,t}^{obs} + G_{i,t}^{unm} \right) / \beta = \left[ \alpha (A_{i,t}^{obs}) + G_{i,t}^{unm} \right] / \beta
\]

where: \( G_{i,t}^{unm} \) is the number of tourists which used establishments for which information on arrivals and nights spent are not collected ("unmeasured tourism"); and \( G_{i,t}^{unm} \) is the number of tourists which used official accommodation establishments, but were not declared to public authorities, mainly for fiscal reasons ("underground tourism"). For simplicity we call this expression as the Trips-Arrivals (T-A) equation. Supply-side surveys usually provides information only on the number of guests arrivals in official establishments (i.e. \( A_{i,t}^{obs} \)); on the contrary, the remaining aggregates needs to be estimated. Moreover, information about the coefficient \( \beta \) (i.e. the average number of accommodation establishments used by tourists during their stay) become really relevant. This issue falls into the broader phenomenon of tourists mobility (Lue et al., 1993; McKercher, Lew 2004); a topic which is almost ignored by actual official statistics, but which have important implication not only for the estimation of tourism trips at sub-regional level, but also for logistic and tourism services provisioning and management.
2.3. The Average Duration of Visit – Average Length of Stay (ADOV-ALOS) equation

Finally, another important aggregate usually considered as an indicator of tourism behavior, is given by the so-called “average length of stay” (ALOS), defined as the ratio between the overnight stays and arrivals: $^{obs}ALOS_{i,t} = \frac{^{obs}P_{i,t}}{^{obs}A_{i,t}}$. However, this index, usually read as an indicator of the length of the trips in the destination/region considered, for the problems above highlighted (i.e. unobserved tourism, and tourists mobility), is only a measure of the “average length of stay” in official accommodation establishments. On the contrary, the “average duration of visit” (ADOV), in the destination/region $i$, during the time interval $t$ considered, according to our framework would be given by the ratio between equation (1) and (2). Subsequently, we have:

$$ADOV_{i,t} = \frac{^{off}N_{i,t}}{TRIPS_{i,t}} = \frac{\beta \alpha_i \left(^{obs}P_{i,t}\right)}{\alpha_i \left(^{obs}A_{i,t}\right) + ^{unobs}G_{i,t} + ^{underground}G_{i,t}}$$

$$= \frac{\beta}{\alpha_i} \left[ \frac{\alpha_i ^{obs}ALOS_{i,t}}{^{obs}A_{i,t}} \times \alpha_i \left(^{obs}A_{i,t}\right) \right] + \left( ^{unobs}ALOS_{i,t} \times ^{unobs}G_{i,t} \right) + \left( ^{underground}ALOS_{i,t} \times ^{underground}G_{i,t} \right)$$

We call this expression as the Average Duration of Visit-Average Length of Stay (ADOV-ALOS) equation, which expresses the average duration of visit, in the destination/region $i$, during the time interval $t$ considered ($ADOV_{i,t}$), as a weighted mean of the average length of stays (of tourism trips) in the different types of establishments/situations considered (official, unmeasured, underground), multiplied by the $\beta$ coefficient. The weights are given by the number of guests arrivals (official, unmeasured, underground). It should be noted that if the $\beta$ coefficient would not be taken into account, the weighted mean of the different $ALOS$, represents the average length of stay in the different establishments/situations considered (official, unmeasured, underground). Also in this case, the only known component is $^{obs}ALOS_{i,t}$.

3. EMPIRICAL EVIDENCES FROM THE SURVEY ON INCOMING TOURISM IN SICILY

In order to determine the actual magnitude of tourism in Sicily, and to quantify the relevance of unobserved tourism in the island, in the period between 2009 and 2010, thanks to a research project co-founded by the Italian Ministry of University and Research, the research group of the University of Palermo and Catania, composed mainly by social statisticians, planned a survey covering the whole Sicily. The survey aimed to estimate the actual magnitude of tourism in the Island, trying to quantify two of the main biases related to statistics on guests
arrivals: the double counting effect (i.e. the $\beta$ parameter), and the “un-observed tourism” (particularly, the un-measured component, i.e. $uG_{i,t}$, and $uN_{i,t}$). For the survey on incoming tourism in Sicily, a complex Time Location Sampling (TLS) design was adopted, given the mobile and particular nature of tourists population (see De Cantis, et al. 2010; Kalsbeek, 2003). The units of interest were represented by Italian (not resident in the Island) and foreign tourists leaving the Island 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 Sicily, through a direct interview, allowing to reduce the recall bias, which usually affect many demand-side surveys (Rylander et al., 1995). A detailed description of the sampling design is contained in De Cantis, et al. (2010). The insularity of Sicily allowed us to select almost all the places from which it is possible to leave the Island, namely: the airports of Palermo, Catania, and Trapani, the ports of Palermo and Catania, and the Strait of Messina (only the two airports of the two small islands Pantelleria and Lampedusa were not included in the survey). 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 official tourists flows are concentrated. The research instrument was represented by a questionnaire of 29 items. The questionnaire was divided into different sections: filter questions and organization of the trip; motivations and expectations; type of holiday (sea and sand; cultural, etc.); intra-regional mobility and type of establishments used; expenses; satisfaction. The specific section of the questionnaire related to the collection of information on tourism mobility and on un-observed tourism is presented in figure 1.

**Fig. 1. Questionnaire section on tourism mobility**

In this section, the tourist was asked to specify all the places (municipalities) which he/she visited during his/her trip, with at least one overnight stay. For each places visited he/she was asked to specify the number of nights spent, and the type of accommodation establishment used, to be able to distinguish between official and un-official establishments. Through this section it was possible to relate the information collected to the two topics of interest: tourism mobility (i.e. the $\beta$ parameter) and un-measured tourism (i.e. $uG_{i,t}$, and $uN_{i,t}$).
Between Summer 2009 and Spring 2010 a total of 3,935 valid interviews were collected (i.e. incoming tourists in Sicily). Although the survey had several research aims, which are presented in another work (Oliveri, De Cantis, 2013) the in this paper we present some first results in order to implement the T-A model, by highlighting both the un-measured component of tourism demand in Sicily, and the effects produced by tourists mobility within the Island. In table 1 data related with the number of visits, with the nights spent, and with the average length of stay of tourists interviewed in Sicily, by accommodation establishment category are presented.

Table 1. Results in terms of stays and overnight stays by accommodation establishment category, from 3,935 interviews to incoming tourists in Sicily, Summer – Autumn 2009, Spring 2010.

<table>
<thead>
<tr>
<th>Accommodation establishment category</th>
<th>Stays</th>
<th>Overnight stays</th>
<th>Average length of stay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Official establishments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rural establishments</td>
<td>152</td>
<td>589</td>
<td>3.88</td>
</tr>
<tr>
<td>Holyday camps</td>
<td>24</td>
<td>200</td>
<td>8.33</td>
</tr>
<tr>
<td>Hotels</td>
<td>2,615</td>
<td>11,071</td>
<td>4.23</td>
</tr>
<tr>
<td>Camping</td>
<td>377</td>
<td>1,183</td>
<td>3.14</td>
</tr>
<tr>
<td>Bed and Breakfast</td>
<td>1,023</td>
<td>3,359</td>
<td>3.28</td>
</tr>
<tr>
<td>Youth hostels</td>
<td>46</td>
<td>129</td>
<td>2.80</td>
</tr>
<tr>
<td>Un-official establishments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>House or room rented</td>
<td>461</td>
<td>4,607</td>
<td>9.99</td>
</tr>
<tr>
<td>Relative and friends houses</td>
<td>1,354</td>
<td>12,587</td>
<td>9.30</td>
</tr>
<tr>
<td>Owned houses</td>
<td>307</td>
<td>4,502</td>
<td>14.66</td>
</tr>
<tr>
<td>Other un-official establishments</td>
<td>126</td>
<td>417</td>
<td>3.31</td>
</tr>
<tr>
<td>Total</td>
<td>6,485</td>
<td>38,644</td>
<td>5.96</td>
</tr>
</tbody>
</table>

Average duration of visit in Sicily (ADOV\textsubscript{i,t}) = 38,644/3,935 = 9.82

According to the T-A model, it is possible to quantify some of the aggregates contained in the three equations of the model. Regarding the N-P equation, a total of 38,644 nights (\textsuperscript{\text{tot}}N\textsubscript{i,t}) were spent by incoming tourists sampled in Sicily, during the time-interval considered. These are only partially measured by official statistics on guests arrivals, since un-official establishments are not covered by supply-side statistics (\textsuperscript{unm}N\textsubscript{i,t} = 22,113), resulting in a sampling share of unmeasured nights equal to the 57% of total nights spent in Sicily by tourists interviewed. Moreover, nights spent in official establishments can be partially concealed to public authorities for fiscal reasons, but our method and results do not allow to separate the underground and the official components, resulting in total of 16,531 nights which could be in some measure declared or underground (\textsuperscript{\alpha}_{0}\textsuperscript{obs}P\textsubscript{i,t} + \textsuperscript{unm}N\textsubscript{i,t} = 16,531). With reference to the second equation (i.e. the T-A equation) of the model, a total of 3,935 tourism trips (TRIPS\textsubscript{i,t}) sampled made 6,485 stays, distributed among several establishments categories, resulting in a value of the average number of stays (\beta) almost equal to 1.65. The stays in official establishments (i.e. \textsuperscript{\alpha}\textsubscript{i}\textsuperscript{\alpha}A\textsubscript{i,t} + \textsuperscript{unm}G\textsubscript{i,t}) resulted equal to 4,237, whereas 2,248 are the un-measured stays (\textsuperscript{unm}G\textsubscript{i,t}). Finally, for the ADOV-ALOS equation,
with a total of about 38 thousands nights spent in Sicily by tourists interviewed, it resulted an average duration of visit \((\text{ADOV}_{it})\) in Sicily equal to 9.82 nights. The average length of stay \((\text{ALOS}_{it})\), instead, varies among the different establishment categories, resulting in a value of 3.90, in official establishments, and in a value of 11.65, in un-official establishments.

As above told, official data on guests in accommodation establishments give information only on arrivals, presences, and average length of stay in official establishments), as shown in table 2. The average length of stay resulting from official data (table 2) is quite similar to the average length of stay in official establishments, derived from the survey on incoming tourists (table 1). However, the average length of stay in un-official establishments resulting from the survey is about three times the official ones. This example empirically shows the biases which occurs when data on guests arrivals, overnight stays and average length of stay, are used to quantify and characterize tourism flows in a given destination.

<table>
<thead>
<tr>
<th>Establishment category</th>
<th>Arrivals</th>
<th>Overnight Stays</th>
<th>Average Length of Stay</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hotels and similar establishments</td>
<td>2,491,373</td>
<td>8,325,020</td>
<td>3.34</td>
</tr>
<tr>
<td>Other collective establishments</td>
<td>319,508</td>
<td>1,320,144</td>
<td>4.13</td>
</tr>
<tr>
<td>Total</td>
<td>2,810,881</td>
<td>9,645,164</td>
<td>3.43</td>
</tr>
</tbody>
</table>


4. COMMENTS AND CONCLUSIONS

Increasingly regional tourism authorities are interested in regional statistics. However, as highlighted in this work, at sub-regional and local level demand-side statistics are not provided by the European system of tourism statistics. This determined the habits of destination managers, policy makers, and researchers, of using accommodation (supply-side) statistics to evaluate tourism demand. However, the use of supply-side information to evaluate demand-side features can determines conceptual and practical mistakes. Some of these issues are getting recognized by major institutions, such as the European Travel Commission (ETC) and Eurostat. For example, the recent quarterly report published by the ETC on European tourism – trend and prospects (ETC, 2010) in comparing the results of the US Department of Commerce, reporting a decline of US outbound travel to Europe, with the results of the TourMIS, which indicated an increase of US arrivals in Europe, and a reduction of the average length of stay, commented that: “one plausible way to read the data is that US travellers are participating in multi-leg European trips with shorter stay in each destination” (ETC, 2010:16). However, despite the phenomenon of multi-destination trip is
being recognized, there are still no official sources of information able to measure the magnitude and the features of this phenomenon, neither at a national (visits to several regions, municipalities, etc.), nor at an international (visits to several countries) level.

Regarding the un-observed component of tourism, related to the use of un-official establishments (namely, the un-measured tourism), the new Regulation (EU) No 692/2011 of the European Parliament, put the attention to the non-rented accommodations, meaning, inter alia, accommodation provided without charge by family or friends and accommodation in owner-occupied vacation homes, including time share properties (European Parliament, 2011:19). According to the new Regulation, the data to be transmitted by the Member States shall related not only to the capacity and occupancy of tourists accommodation establishments, but also to tourism nights spent in non-rented accommodation. To date, however, nor in Italy, nor in the other European countries, no information on un-measured tourism (nor on underground tourism) are available, and the way in which member countries will collect and provide these information is still an open issue.

The T-A model allows to face with the problem of quantifying the number of tourism trips made in a given destination by highlighting the lack of official supply-side tourism statistics. Next challenges are related to the estimation of coefficients and quantities presented in the equations, and to the linkages between supply-side and demand-side information. With reference to the motivation coefficients (i.e. $\alpha_0$ and $\alpha_1$) it should be kept in mind the characterization of the destination/region. In tourism resorts, it could be assumed that all guests are tourists (i.e. $\alpha_0 = \alpha_1 = 1$); however, this hypothesis would be unreliable in urban destinations where other guests (e.g. workers) are likely to visit the destination and stay in collective establishments. In these cases, an estimate of $\alpha_0$ and $\alpha_1$, obtained for example through a sample survey on official establishments would be required. The unmeasured component of tourism demand is closely related to the presence of the so-called “un-official establishments”, such as second houses, rooms or houses rented. The estimation of the number of second houses in a given destination/region, for example through information coming from the census on population and housing could help to understand the magnitude of the unmeasured tourism in the destination considered. The underground component is even harder to quantify, since deliberately concealed to public authorities by accommodation establishments manager. This issue falls into the broader issue of measurement of un-observed economy (OECD, 2002), and, to date, no direct solutions have been proposed. Finally, regarding the $\beta$ parameter, next to nothing is known about the number of destinations visited (nor of the number of establishments used) by tourists. However, for small areas, such as municipalities a value of $\beta = 1$, whereas for larger areas, such as tourism districts, or Provinces, an estimate of $\beta$ would be required. By concluding, a deeper knowledge of tourists behavior is required to determine the values of these parameters and the factors affecting their variability.
Furthermore, the changing nature of demand and the increasing segmentation of the holiday market are also raising the need for more accurate, destination-based, information which integrate quantitative information on the magnitude of tourism with other more qualitative aspects of tourism behaviors.

REFERENCES


European Communities (1994). Community methodology on tourism statistics, Office for Official Publications of the European Communities, Luxembourg


Gallent N., Tewdwr-Jones M. (2000). Rural second homes in Europe: examining housing supply and planning control, Ashgate.


