

Calibration as a tool to enhance coherence at regional level in tourism statistics

C. Fullea Carrera¹, M. Izquierdo Valverde², M. Velasco Gimeno³

¹ National Statistical Institute (INE), Madrid, Spain; carlos.fullea.carrera@ine.es

² National Statistical Institute (INE), Madrid, Spain; maria.izquierdo.valverde@ine.es

³ National Statistical Institute (INE), Madrid, Spain; maria.velasco.gimeno@ine.es

Abstract

Tourism occurs at local level, with different incidence and characteristics in each destination. In many countries, estimation of tourism flows come from household surveys (outbound and domestic), border crossing (inbound) and accommodation statistics. These sources of primary information have a strong methodological base and fulfil high quality standards. But they usually present limitations in the analysis at sub-national level.

In Spain, surveys on accommodation establishments conducted by INE provide monthly information of the number of tourists and overnight stays with much geographical detail, reaching even the municipality level. However, to complete the picture of the tourism flows in a determined area, information on the tourists staying in non-market accommodations is also needed.

Demand-side surveys on resident and international tourists cover the total tourism flows, and also allow to explore the characteristics of the visitors and their trips in terms of duration, expenditure, transport, accommodation, activities, etc. Nevertheless, the geographical detail of this kind of surveys is quite limited, covering only the regional level. Besides, figures for market accommodation in a region obtained from these surveys do not always match those provided by accommodation statistics, more robust at that level as they are specifically designed to fulfil that objective. The analysis of the regional figures becomes then difficult for interested users.

Calibration is a reweighting method widely used to improve the accuracy and coherence of official statistics. Sample weights are adjusted to match population totals from external sources. When the variables considered are highly correlated with the objective of the survey, bias and variance of the final estimates are reduced. In the case of the demand surveys, if figures obtained from the accommodation sector at regional level are introduced in the process, more accurate estimates are obtained and the coherence of the different sources at regional level is enhanced, making them more meaningful.

This paper aims to describe i) the conceptualization and operational exercise needed to make variables from the demand and supply side comparable, ii) the process developed (for the main destinations) and the results obtained and iii) its extension to less popular destinations combining different reference periods. This exercise can also be

analysed in terms of the potential use of estimates derived from big data sources as an input for calibration.

Keywords: calibration, regional, coherence, disaggregation, tourism demand, accommodation

1. Introduction

Tourism occurs at local level, with different incidence and characteristics in each destination. In many countries, estimation of tourism flows come from household surveys (outbound and/or domestic), border crossing (inbound and/or outbound) and accommodation statistics (inbound and/or domestic). These sources of primary information have a strong methodological base and fulfil high quality standards, developed under internationally agreed guidelines and recommendations, such as the International Recommendations for Tourism Statistics (UNWTO, 2008) or the Methodological manual for tourism statistics (Eurostat, 2014). However, these surveys usually present limitations in the analysis at sub-national level.

Nights spent at tourism accommodation establishments is one of the more commonly used tourism indicators worldwide. Data on tourism accommodation establishments are often obtained from samples that combine a census for establishments that rank higher in the classification systems of tourism accommodation with a random sample of other establishments. Depending on the country, this means that the biggest hotels are self-represented in the sample, allowing to obtain meaningful results of the performance of this industry in the territories where tourism activity is more intense.

In Spain, surveys on accommodation establishments conducted by INE provide monthly information of the number of tourists and overnight stays with much geographical detail, reaching even the municipality level on a monthly basis. However, to complete the picture of the tourism flows in a determined area, information on the tourists staying in non-market accommodations is also needed. Figures 1 and 2 show the share of nights spent at market accommodation establishments for domestic and inbound tourism, respectively.

Demand-side surveys on resident and international tourists cover the total tourism flows, and also allow to explore the characteristics of the visitors and their trips in terms of duration, expenditure, transport, accommodation, activities, etc. Border surveys are very costly from an operational perspective and household surveys are affected by the intensity of the phenomenon in the country. The less the residents travel, the bigger samples are needed, therefore the higher the cost. For this reason, the geographical detail of this kind of surveys is quite limited. In Spain they cover only the regional level: monthly estimates can be obtained for the main regions of destination whilst quarterly or annual estimates are obtained for the less popular ones.

Figure 1. Share of nights spent at market accommodation establishments by region. Domestic tourism. Residents Travel Survey. 2017

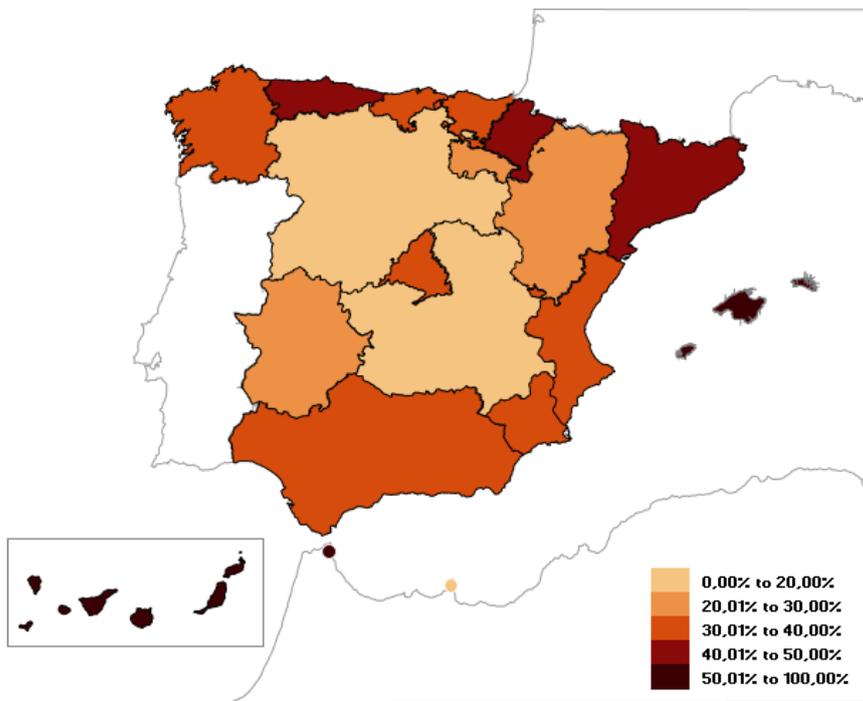
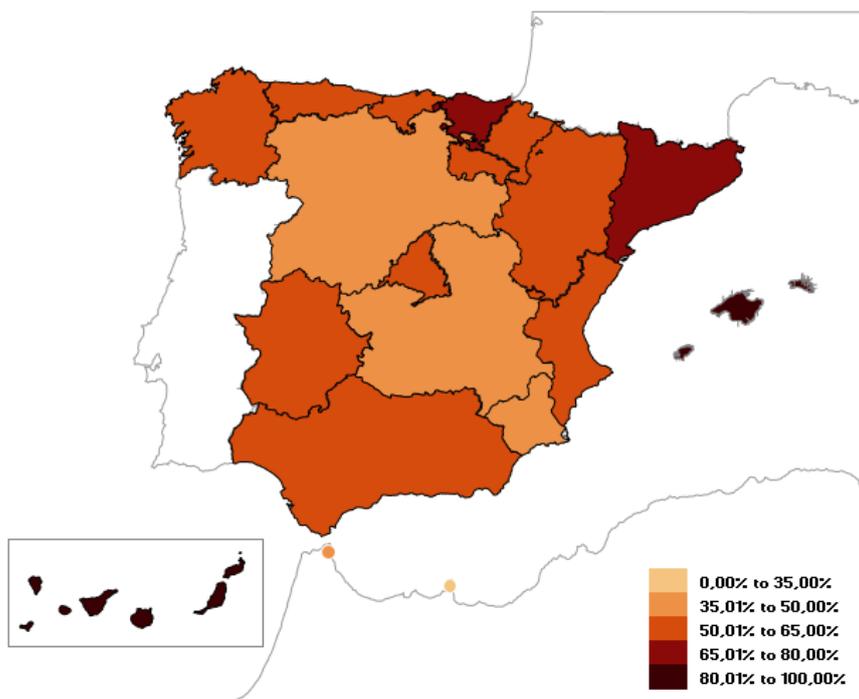


Figure 2. Share of nights spent at market accommodation establishments by region. Inbound tourism. EGATUR. 2017



When figures are available from both demand and supply sources, some differences may arise when comparing them. Figures for market accommodation in a region obtained from demand surveys do not always match those provided by accommodation statistics, the latter being more robust at that level. The analysis of the regional figures becomes then difficult for interested users.

From 2015 onwards, the National Statistics Institute of Spain (INE) has tried to integrate the different sources to give a comprehensive picture of tourism at regional level. This paper aims to describe i) the conceptualization and operational exercise needed to make variables from the demand and supply side comparable, ii) the process developed to achieve consistency of the figures for the main destinations and the results obtained and iii) its extension to less popular destinations combining different reference periods. This exercise can also be analysed in terms of the potential use of external estimates derived from big data sources as an input for calibration.

Section 2 focuses on domestic tourism and section 3 is dedicated to inbound tourism.

2. Domestic tourism: calibration of the demand survey to accommodation figures

2.1. Residents Travel Survey

The Residents Travel Survey (RTS) is a household survey whose main objective is to estimate the total number of tourism trips made by the population resident in Spain. Both trips within the national territory and abroad are studied, and also their main characteristics: nights spent and associated expenditure. Information on the main elements of the trip (e.g. origin, destination, length, main purpose, accommodation, means of transport, etc.) and the visitor is also available to analyse the travelling behaviour of the population.

The methodology of this survey follows the guidelines determined in the Methodological Manual for Tourism Statistics (Eurostat), developed under the Regulation 692/2011 concerning European statistics on tourism.

Monthly and quarterly estimates are published on a quarterly basis. Although the destination of the trip is collected with municipality detail, the geographical dimension is only explored at regional level with quarterly reference.

Around 8,200 interviews are conducted every month. Respondents give information on the trips they have finalized in the two previous months. Thus, estimates of the number of trips made in one month are based on a sample of around 16,400 questionnaires. After a process of editing and imputation, data are weighted to obtain the final figures. A methodological reference of the survey is available [online](#)¹.

Although the number of interviews is considerably big, the sample of trips obtained is significantly smaller. As an example, in 2017 in the Spanish survey, on average, just the 24% of the sample had made at least one trip during the period of reference of the

¹ http://www.ine.es/en/daco/daco42/etr/etr_metodologia_en.pdf

questionnaire (the two previous months). Obviously, that percentage decreases when focusing on specific kind of trips such as trips abroad, trips using hotels as the main accommodation, etc. This leads to high sample errors, as shown in Table 2.

Table 1. Share of trips by month and main accommodation.

	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12
Total trips	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Hotels and similar establishments	19,3	23,9	24,0	23,1	24,3	25,2	19,6	17,8	21,5	24,2	25,4	21,2
Other market accommodation	10,1	10,6	12,0	12,8	12,7	14,7	18,6	21,1	16,2	13,7	13,0	13,2
Non-market accommodation	70,6	65,6	64,0	64,1	63,0	60,1	61,9	61,1	62,3	62,1	61,6	65,5

Source: Residents Travel Survey, 2017.

Table 2. Trips by month and main accommodation. Coefficient of variation (%)

	M01	M02	M03	M04	M05	M06	M07	M08	M09	M10	M11	M12
Total trips	2,72	2,73	2,78	2,12	2,55	2,45	2,19	1,72	2,38	2,50	2,86	2,47
Hotels and similar establishments	5,15	5,01	5,51	3,59	4,28	4,23	3,58	3,26	3,78	4,19	4,80	4,37
Other market accommodation	9,41	8,69	8,04	5,95	8,86	6,88	4,47	3,67	5,88	6,35	8,47	6,95
Non-market accommodation	3,31	3,50	3,55	2,85	3,20	3,24	3,08	2,37	3,26	3,41	3,77	3,22

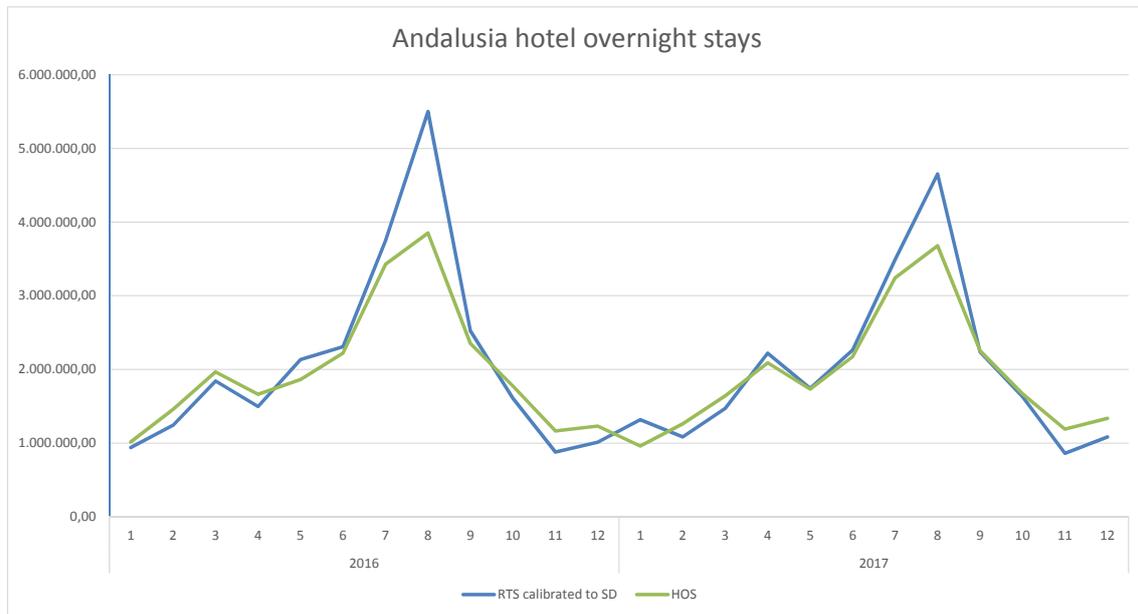
Source: Residents Travel Survey, 2017.

Measures are taken during the survey design to improve the efficiency of the sample: households with higher propensity to make trips are oversampled. At a later stage, weights are adjusted to take into account that oversampling. Calibration to sociodemographic variables is also applied to correct non-response biases.

At this first stage, RTS results can be compared with those obtained from the Hotel Occupancy Survey (HOS), the survey measuring overnight stays at hotels and similar establishments based on the information provided by hotels.

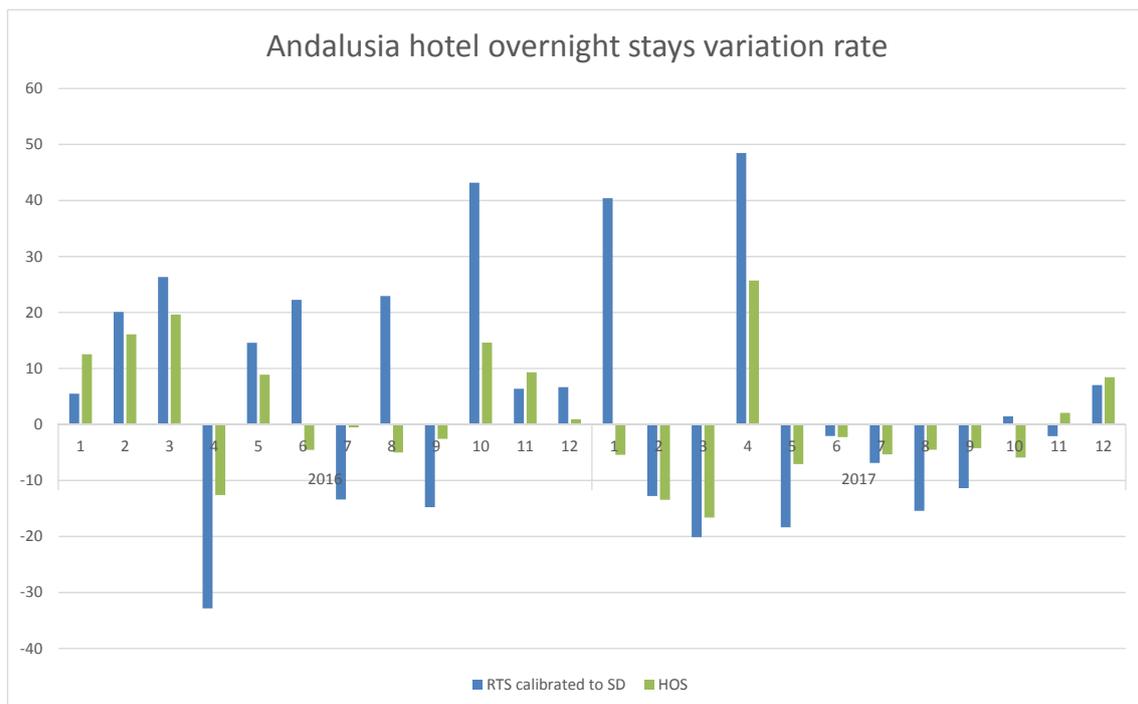
Figure 3 shows the series of overnight stays at hotels and similar establishments obtained from both sources for the most popular region of destination. From a statistical perspective, coherence of the two series can be considered very good taking into account the methodological and conceptual differences behind and also the sampling problems described above. High values observed during the summer months in the RTS survey have a conceptual explanation: RTS measures overnight stays of trips finishing within the month of reference whilst HOS measures overnight stays really made during that month.

Figure 3. Nights spent at hotels in Andalusia. HOS estimates and RTS estimates



Differences are more obvious when looking at evolutions (figure 4): although in most months variation rates obtained from both sources go in the same direction, RTS values are more extreme, due to the bigger sampling errors of its figures. Such behaviour is commonly questioned by users and somehow damages the reliability of the demand surveys. Using the statistical tools at hand, INE has tried to improve the consistency of these figures as it is explained in the next section.

Figure 4. Nights spent at hotels in Andalusia. Variation rate (%). HOS estimates and RTS estimates



2.2. Calibration. Description

Calibration is a reweighting method widely used to improve the accuracy and coherence of official statistics. Sample weights are adjusted to match population totals from external sources. When the variables considered are highly correlated with the objective of the survey, bias and variance of the final estimates are reduced. In the case of the demand surveys, if figures obtained from the accommodation sector at regional level are introduced in the process, more accurate estimates are obtained and the coherence of the different sources at regional level is enhanced, making them more meaningful.

In the previous section it was mentioned that original weights are calibrated to population totals to correct nonresponse biases by adjusting the sample structure to the population. Population totals are obtained from demographic statistics and the Labour Force Survey at regional level (NUTS 2):

- Population by nationality (Spanish or foreign)
- Population by gender and age (five-year groups)
- Population by educational level (primary, secondary and higher education)
- Population by provinces (NUTS 3 level)

This is a standard procedure in most household surveys at the Spanish National Statistics Institute. But, in the RTS, accuracy is also improved by enhancing the coherence with the Hotel Occupancy Survey (HOS). To make a correct use of this technique, variables considered in the survey should be conceptually equal to those used as auxiliary information. In this case, we cannot use direct estimates due to methodologic differences between both surveys, which must be solved before calibration. These differences are mainly two:

- The HOS considers the overnight stays spent in hotels on the reference month whereas the RTS assigns all the overnight stays linked to one trip to the month in which the trips ends. That means that, for a given month:
 1. RTS considers some overnight stays of previous months and HOS doesn't.
 2. RTS doesn't consider reference month overnight stays of trips ended in following months and HOS does.
- Systematic differences in the coverage as for example permanent residents in hotels, considered by the HOS but not by the RTS (because no trip was made).

The previous means that differences should remain in the total aggregates. Therefore, the solution adopted was to derive 'base estimates' for the year before the beginning of the series. Those estimates were achieved carrying out a transformation of the overnight stays provided by the HOS into RTS overnight stays, trying to remove the differences mentioned above. For the following years, RTS figures are adjusted to the estimates obtained applying the HOS variation rate to these comparable overnights estimated for the first year. In addition, to estimate the monthly auxiliary variables, the RTS overnight stays are truncated to the maximum number of days of the reference month, to be coherent with the marginals used.

When RTS questionnaire was designed, it was prepared to collect not only the main destination of the trip but also the number of nights spent at every municipality visited during the trip. Therefore RTS collects information of all the legs of the trip (destination and accommodation), allowing to estimate the hotel overnight stays spent in each region.

Six monthly variables are used to calibrate: hotel overnight stays in Madrid, Canary Islands, Andalusia, Catalonia, Valencian community and one which groups the others together. The reason to make this last group is that these communities don't have enough sample size to be calibrated by its own, causing problems in the convergence of the CALMAR algorithm.

2.3. Calibration. Results

For the sake of simplicity, the analysis of the effect of introducing population totals to calibrate the weights will be omitted, focusing only on the effect of the introduction of HOS figures in the process.

Figure 5 shows the series presented in figure 3 and adds a third one: the results of the calibration process. The red line represents the final figures for hotel overnight stays in Andalusia obtained from RTS. Although differences remain, a significant improvement of the coherence of the two sources is observed. Also for variation rates are now smoother (figure 6).

Figure 5. Nights spent at hotels in Andalusia. HOS estimates and RTS estimates (before and after calibration to HOS derived figures)

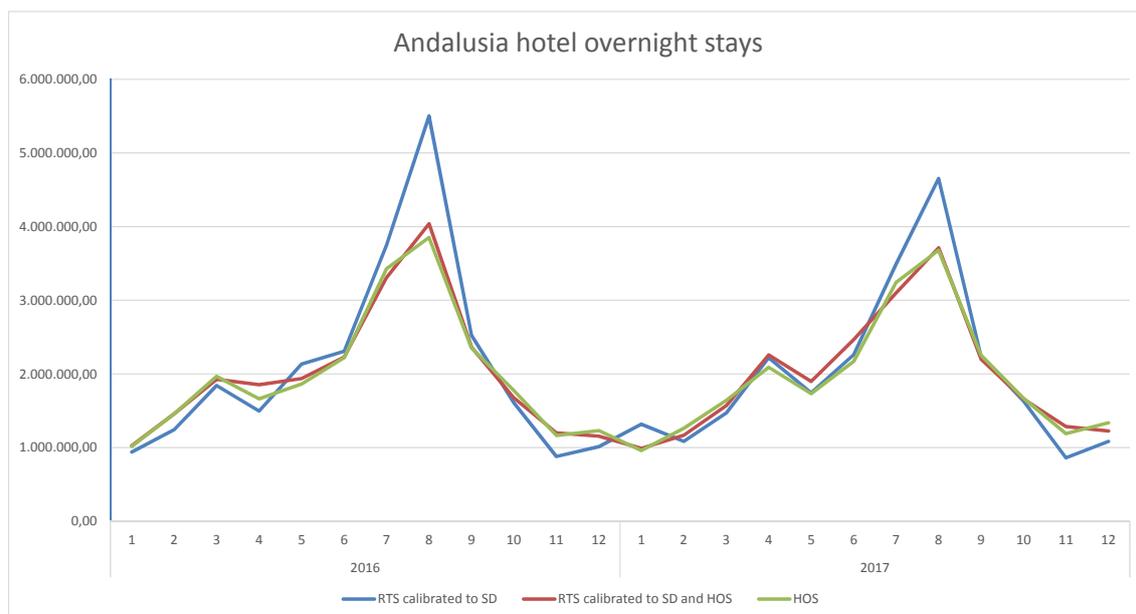
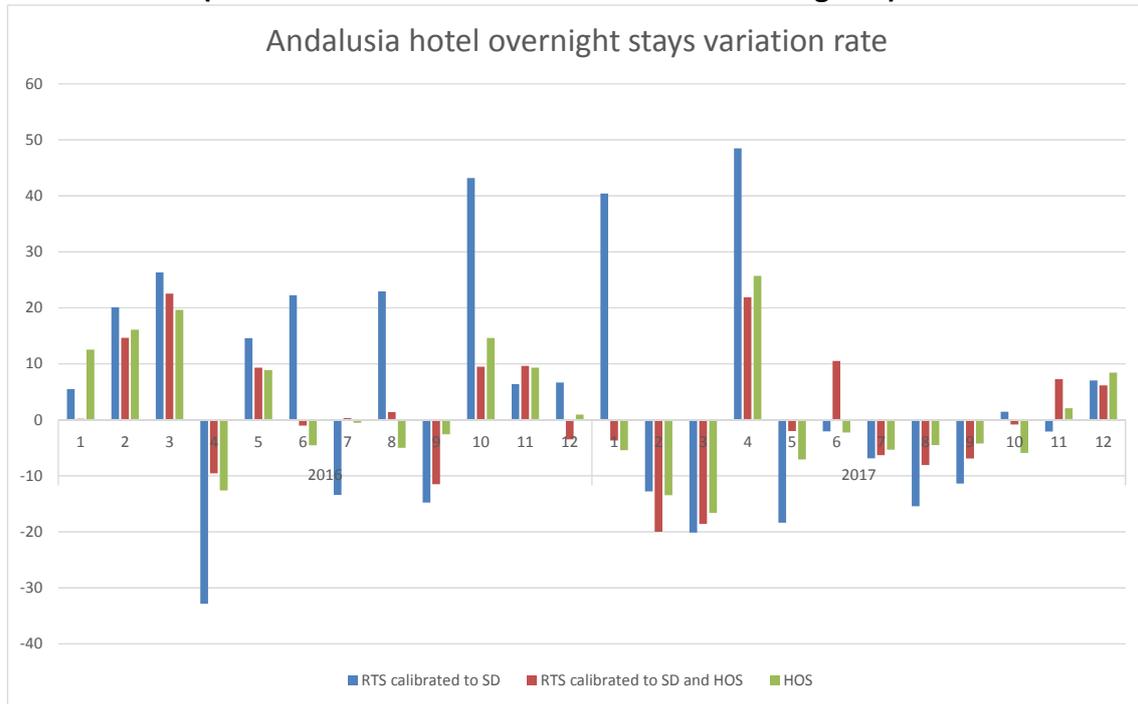
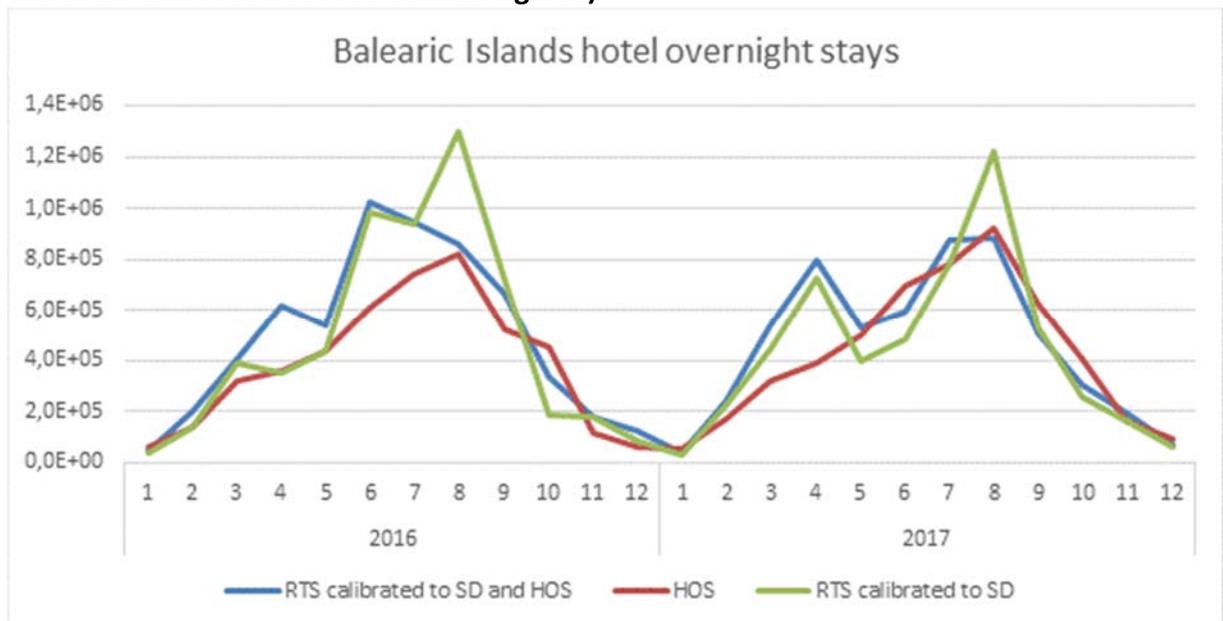


Figure 6. Nights spent at hotels in Andalusia. Variation rate (%). HOS estimates and RTS estimates (before and after calibration to HOS derived figures)



However, if we observe one of the destinations with a smaller sample, which is included in the group 'others', we still find important differences (figure 7). The monthly procedure described ensures coherence at national level and for the five regions treated individually. In the next section a similar approach is proposed for those other regions, combining different reference periods.

Figure 7. Nights spent at hotels in Andalusia. HOS estimates and RTS estimates (before and after calibration to HOS derived figures)



2.4. Extended calibration: combining monthly and quarterly totals

For these autonomous communities with a smaller sample size, a new solution has been explored. The survey is designed to provide monthly estimates at national level so the information collected is treated on a monthly basis and therefore weights are calculated for each reference month.

However, results are only released on a quarterly basis to be able to obtain more rich information. Estimates of the number of trips made in one quarter are the result of adding up the figures of the three months of the quarter. Pooling monthly files into a quarterly one would increase the number of collected trips to each region of destination. The new approach consists of calculating in the same process the weights of the three months of the quarter, taking into account that:

- Three different samples are considered and each of them should be treated individually when dealing with population totals.
- The number of nights for the main regions of destination are adjusted separately for each one of the three months, as described in the previous sections.
- The number of nights spent in the regions included in 'others' during the monthly process, is adjusted for the whole quarter.

In practical terms, this means that a new quarterly variable representing the overnight stays in each region not calibrated individually is added to the process. To ensure the coherence with the monthly aggregate for this group of regions, new quarterly marginals are computed sharing the monthly marginal among all the regions grouped together, using their weight in the HOS estimates.

Figure 8 shows the quarterly series for the Balearic Islands hotel overnight stays when the quarterly totals obtained from the HOS figures are included in the calibration process. While RTS figures obtained applying only the monthly treatment systematically overestimate the number of nights spent at hotels in the region, the new approach achieves a better performance of RTS estimates.

Figure 8. Balearic Islands hotel overnight stays. HOS estimates and RTS estimates (monthly and quarterly calibrated)

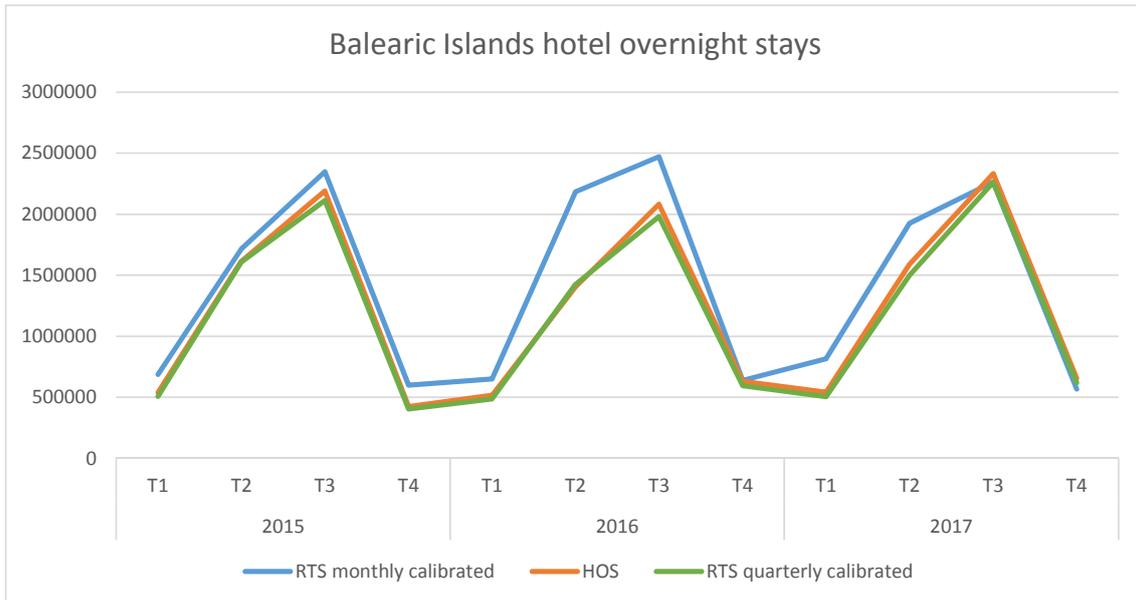
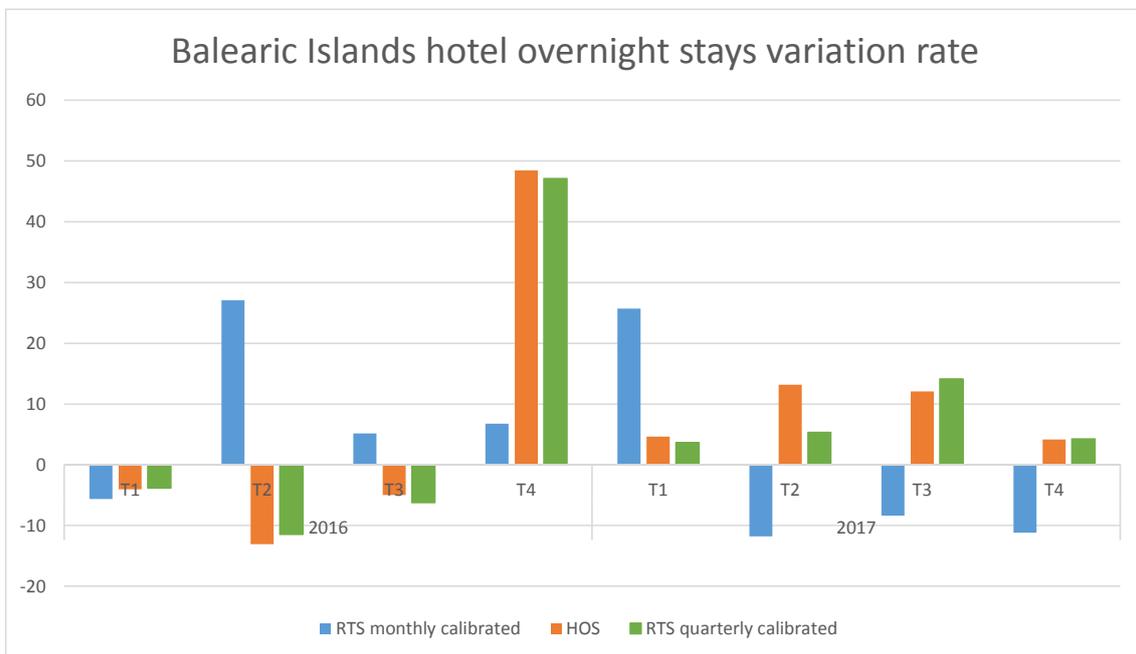


Figure 9. Balearic Islands hotel overnight stays variation rate (%). HOS estimates and RTS estimates (monthly and quarterly calibrated)



A general overview of the results achieved for every region is shown in Table 3. Variation rates for the number of overnight stays spent at hotels and similar establishments in 2016 are provided in the table. First column refers to the RTS results after the monthly

procedure, which deals with the less popular regions in the same group. Second column shows the results after the quarterly process, much closer to those on column 3, obtained from HOS, and use as a reference during calibration.

Table 3. Nights spent at hotels and similar establishments. Variation rate. RTS and HOS estimates. 2016

	Variation rate (%)		
	RTS (monthly calibrated)	RTS (quarterly calibrated)	HOS
01 Andalucía	2,39%	2,39%	2,13%
02 Aragón	19,98%	11,03%	11,95%
03 Asturias, Principado de	-3,30%	-0,29%	1,92%
04 Balears, Illes	11,28%	-2,82%	-2,84%
05 Canarias	0,18%	0,18%	1,56%
06 Cantabria	12,29%	6,08%	7,29%
07 Castilla y León	12,26%	3,97%	4,92%
08 Castilla - La Mancha	-6,42%	8,15%	8,97%
09 Cataluña	5,65%	5,65%	5,90%
10 Comunitat Valenciana	2,18%	2,18%	3,75%
11 Extremadura	5,34%	6,84%	7,19%
12 Galicia	-4,68%	6,96%	7,48%
13 Madrid, Comunidad de	-2,38%	-2,38%	-1,28%
14 Murcia, Región de	31,65%	5,19%	7,82%
15 Navarra, Comunidad Foral de	-9,53%	9,54%	8,57%
16 País Vasco	-7,91%	1,44%	2,33%
17 Rioja, La	-24,87%	-0,04%	0,84%

Although differences due to methodological reasons remain, this new figures give the users a coherent message and eases the analysis at regional level. Besides, after introducing in the process auxiliary information with lower sampling errors and highly correlated with the RTS figures, the new results are more accurate.

3. Inbound tourism: adjustment of the demand survey to accommodation figures

3.1. Introduction

Official statistical figures on international tourism from the demand side are provided by the two following surveys carried by National Statistical Institute of Spain:

- FRONTUR: border survey whose main objective is to measure the number of international visitors (tourists and same-day visitors) that come to Spain monthly, classified by means of transport, country of residence, destination, accommodation.
- EGATUR: it is a survey only for a sub-sample of FRONTUR that analyses more details of each trip and collects information about the tourist expenses associated to those trips.

A similar situation as the one explained between RTS and HOS, also happens between FRONTUR-EGATUR and HOS, but taking into account non-resident tourists. Moreover, not only guests and overnight stays in hotels are in the focus of this statistical exercise,

but also guests and overnight stays (of non-residents) in other market accommodation (as campsites or rural accommodation establishments) are included.

Looking for having a tourism statistical system with integrated data, some actions have to be developed in order to avoid situations as the ones shown in Table 4.

Table 4. Variation rate by month

	012018	022018	032018	042018	052018	062018	072018	082018
EGATUR variation rate (*)	18,10	18,44	27,55	10,95	14,61	6,91	10,38	5,89
HOS variation rate	0,75	1,24	3,3	-4,54	1,58	-1,77	-2,69	-2,54

(*) Figures before calibration process.

3.2. Methodology of FRONTUR-EGATUR

Some brief notes about the methodology of these surveys are going to be mentioned in the following paragraphs.

FRONTUR is a continuous on a monthly basis survey whose population scope is the set of people who go into Spain (whether or not they stayed overnight, as well as to those who pass through the country in transit). The sample is collected at the main access points to Spain, by road, airport, porta and railway, and the survey is conducted when respondent is leaving the country. FRONTUR questionnaire is quite simple, collecting information on means of transport, main destination in Spain, number of overnight stays, main accommodation, main reason of the trip, tourist package (if it was used or not).

In the elevation process, using as population framework people crossing the borders into our country (this framework is build up based on different administrative registers), number of tourist and same-day visitors are calculated.

Around a third part of FRONTUR sample has to answer a wider questionnaire (EGATUR), describing with more details the trip and declaring the expenses related to the trip that is ending in the moment of the interview.

EGATUR elevation process takes as population framework the number of visitors estimated by FRONTUR, broken down by point of entrance to Spain and country of residence.²

3.3. Calibration. Description

The calibration process in FRONTUR-EGATUR has a triple objective:

- To integrate information of overnight stays provided by the Tourist Accommodation Occupancy Surveys (hotel, tourist apartments, campsites and rural tourism accommodations- hereinafter TAOS), in order to make FRONTUR-EGATUR

² You can find the complete methodological document of FRONTUR-EGATUR in the following link:
http://www.ine.es/en/daco/daco42/frontur/frontur_egatur_metodologia_en.pdf

estimations more efficient, due to the fact that this exogenous information has a recognized quality, wide coverage and high accuracy.

- To present consistent results from different sources on the same target variables, in the context of an integrated tourism statistics system.
- To recalculate the initial sample weights of EGATUR to obtain the same estimations for tourists and same-day visitors in a set of cells, that the ones calculated with the FRONTUR sample.

This process is set up in three consecutive phases. In the first one, FRONTUR weights are calibrated being the marginal introduced in CALMAR algorithm the following ones:

- Number of visitors (FRONTUR)
- Number of visitors (residents and non-residents) by way of access (FRONTUR)
- Number of tourists (FRONTUR)
- Number of tourist by main destination (FRONTUR)
- Number of tourist in TAOS by main destinations³ (data from TAOS with specific adjustments to make variables comparable)
- Number of tourist in TAOS by country of residence⁴ (data from TAOS with specific adjustments to make variables comparable)

In the second phase, EGATUR weights are calibrated to a set of figures calculated with the first calibrated FRONTUR and to the overnight stays of TAOS.

The monthly marginals of this calibration process are the overnight stays of TAOS divided by twice crossing destination Autonomous Communities (Andalusia, the Balearic Islands, Catalonia, Valencian community, Madrid, Rest of the ACs) with country of residence (UK, Germany, France, Rest of World). Overnight stays at national level of the following countries of residence are also included: Belgium, Ireland, Italy, The Netherlands, Portugal, Switzerland, Russia, the Nordic countries, Rest of Europe, USA, and Rest of America.

The calibration marginals of overnight stays from TAOS cannot be used directly due to two main types of effects that must be adjusted:

1. Systematic differences in coverage (for example, permanent residents in hotels)
2. The effect due to the presence of overnight stays compiled in surveys corresponding to entries of the month before the month of reference, and those not compiled because the departure takes place in the month after the reference month. However, since the FRONTUR-EGATUR weights refer to visitor entries, whereas the sampling data of overnight stays is visitor departures, the latter is assumed to be largely corrected and therefore, hereinafter, only the former is taken into account.

After adjusting the marginals, they were used as calibration base in the base year (October 2014-September 2015), obtaining for subsequent years the estimation achieved by updating them with the annual coefficient of variation in the direct data

³ 6 most important destinations at NUTS 2 level from international tourism point of view

⁴ 15 most important countries or groups of countries from international tourism point of view

provided by the external source TAOS. The calibration always take into account the overnight stays by stage in EGATUR.

The marginal vector to which EGATUR is calibrated includes the following variables:

- Number of tourists by way of access (FRONTUR)
- Number of tourists by accommodation (FRONTUR)
- Number of tourists by main reason of trip (FRONTUR)
- Number of tourists with tourist package (FRONTUR)
- Number of tourists by main destination (FRONTUR)
- Number of tourists by country of residence (FRONTUR)
- Number of same day visitors by way of access (FRONTUR)
- Number of same day visitors in transit (FRONTUR)
- Number of same day visitors by road and country (FRONTUR)
- Number of same day visitors by airport and country (FRONTUR)
- Number of same day visitors by port and country (FRONTUR)
- Number of same day visitors by airport and transit (FRONTUR)
- Number of same day visitors by port and transit (FRONTUR)
- Overnight stays crossing destination and country of residence (TAOS-adjusted)
- Overnight stays by country of residence (TAOS-adjusted)

The third and last phase in FRONTUR-EGATUR calibration process is to recalculate FRONTUR weights in order to include the improvements introduced in EGATUR in the prior phase for tourists. The external marginal variables come from calibrated EGATUR:

- Number of tourists by way of access (EGATUR)
- Number of tourists by accommodation (EGATUR)
- Number of tourists by main reason of trip (EGATUR)
- Number of tourists with tourist package (EGATUR)
- Number of tourists by main destination (EGATUR)
- Number of tourists crossing destination and country of residence (EGATUR)
- Number of tourists by country of residence (EGATUR)

The final set of margin vectors and the breakdown of the variables were established after a testing period where many different and more ambitious scenarios were got into the CALMAR algorithm. As it is a procedure that is repeated monthly, the exogenous detailed variables have to be coherent with the sample size of FRONTUR and EGATUR to ensure that the algorithm converges.

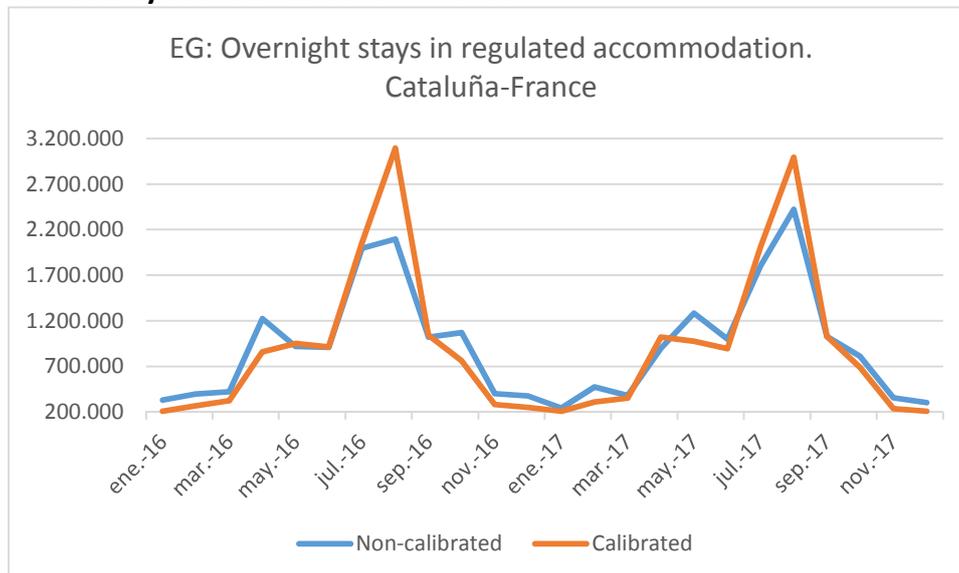
3.4. Calibration. Results

In this last section, the focus is put in analysing the impact of calibration in overnight stays in regulated accommodation in EGATUR (phase 2), considering that it is the most relevant step of the ones describe before, due to the fact that the rest are based on the evolution of this variable in TAOS.

EGATUR estimation of overnight stays in regulated accommodation in Catalonia made by tourists from France is taken as example. Comparing both series, before and after the

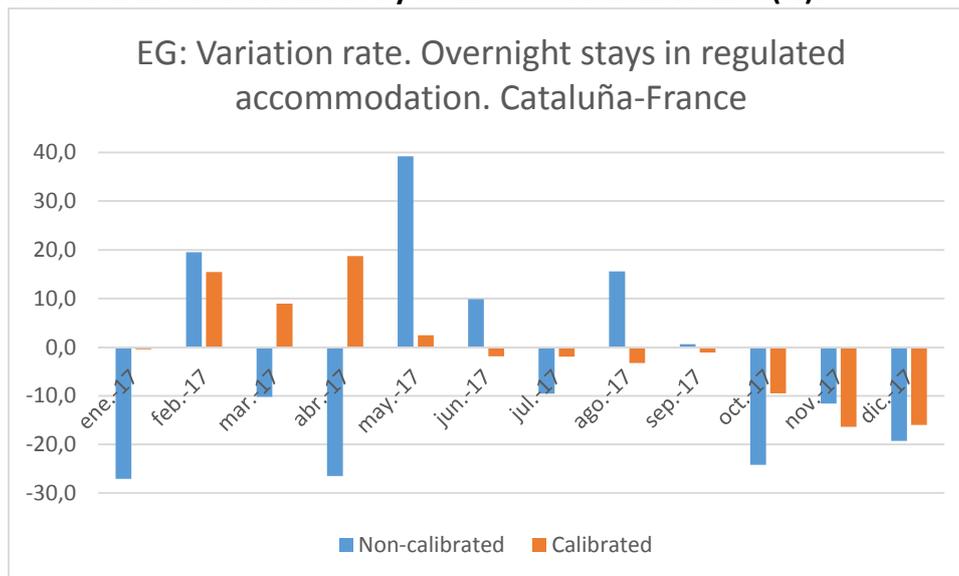
second phase of calibration, we can check that EGATUR and TAOS have similar profiles, but EGATUR underestimates the months of reference with more overnight stays.

Figure 10. EGATUR. Number of overnight stays in regulated accommodation in Catalonia made by tourists resident in France.



Analysing the behaviour of variation rates of that variable, the main impact of calibration is the general smoothing of rates, eliminating volatility, and providing accuracy to the series

Figure 11. EGATUR. Variation rate of number of overnight stays in regulated accommodation in Cataluña made by tourists resident in France (%).



To underline that the differences between EGATUR and TAOS are not always in the same direction, another example is shown, overnight stays in regulated accommodation,

crossing Canary Islands as destination and United Kingdom as country of residence. The behaviour of both sources shows that EGATUR provides over estimation of this variable. With the calibration process we ensure the integration and coherence of these surveys, increasing the quality and, as it was mentioned in the previous example reducing the variability of estimations.

Figure 12. EGATUR. Number of overnight stays in regulated accommodation in Canary Islands made by tourists resident in United Kingdom.

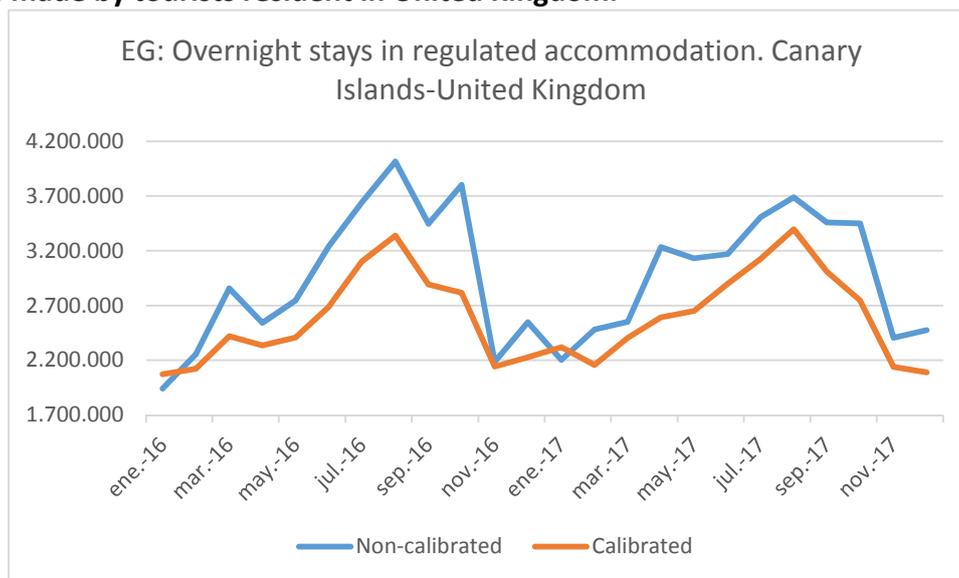
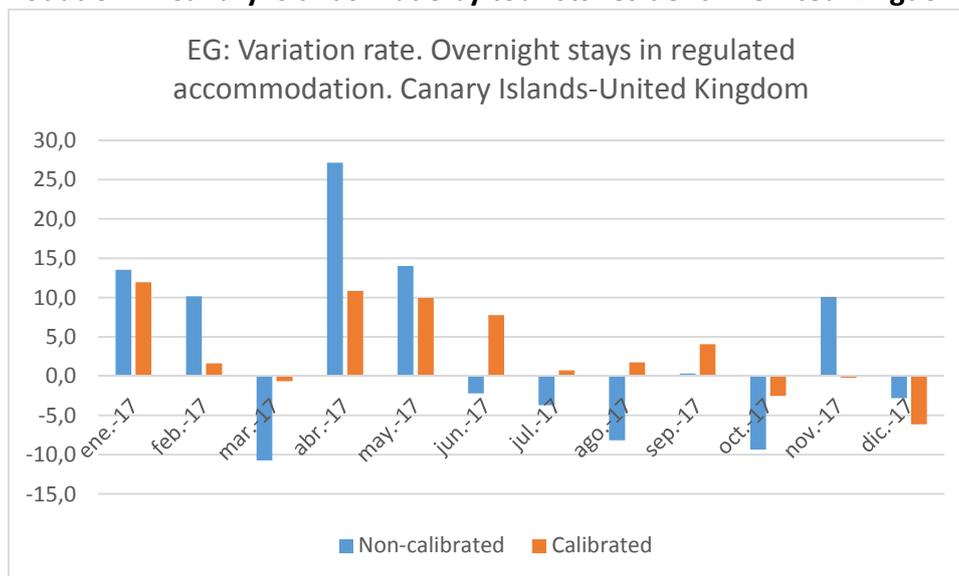


Figure 13. EGATUR. Variation rate of number of overnight stays in regulated accommodation in Canary Islands made by tourists resident in United Kingdom



4. Conclusions

These studies show how auxiliary information can be used to improve the regional estimates obtained from sample surveys.

During the weighting process, calibrating the demand surveys to the number overnight stays obtained from accommodation statistics allows to get more accurate hotel overnight stays estimates and to increase the coherence between the different sources making the analysis easier for users, especially at regional level.

Calibration is flexible enough to use external information with different periods of reference, so it is possible to accumulate files of different months and therefore increase the number of observations.

New sources enabling a more granular geographical detail such as data from mobile phone positioning are being gradually explored by tourism statistics producers. The described approach could serve as inspiration on how external information could be integrated in sample surveys.

5. References

Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, pp. 376-382

Eurostat (2014). Methodological manual for tourism statistics, version 3.1. Available at <http://ec.europa.eu/eurostat/documents/3859598/6454997/KS-GQ-14-013-EN-N.pdf/166605aa-c990-40c4-b9f7-59c297154277>

Mc Cormack, Kevin, CSO, Ireland. The calibration software CALMAR- What is it? Available at <https://es.scribd.com/document/344968317/CalmarEngDoc> (Accessed 18 May 2018)

Regulation (EU) No 692/2011 of the European Parliament and of the Council of 6 July 2011 concerning European statistics on tourism. Available at <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:192:0017:0032:EN:PDF>

United Nations, World Tourism Organization, (2008), International Recommendations for Tourism Statistics. Available at https://unstats.un.org/unsd/publication/Seriesm/SeriesM_83rev1s.pdf