QÜESTIIÓ, vol. 26, 1-2, p. 197-211, 2002

SMOOTHING THE CATALAN TOURISM MICRO-DATA TIME SERIES[⋆]

M. ARTÍS ORTUÑO[*]
J. L. CARRION I SILVESTRE[**]
À. COSTA SÁENZ DE SAN PEDRO[*]
J. SURIÑACH CARALT[*]

In this paper we propose a method for smoothing the Catalan tourism time series between 1997 and 2000. These time series, built upon a micro database drawn from a survey conducted by the Statistical Institute of Catalonia, are somewhat volatile due, it would seem, to the incomplete nature of the information. The application of a smoothing procedure based on the combination of classical techniques and weighted moving averages allows us to overcome the problems caused by this lack of information and to obtain time series that evolve smoothly over time.

Keywords: Smothing micro-data, tourism time series

AMS Classification (MSC 2000): 62P20

⋆ This paper is a joint undertaking between the Statistical Institute of Catalonia (Idescat) and the Anàlisi Quantitativa Regional Research Group of the University of Barcelona.
* Anàlisi Quantitativa Regional (AQR) Research Group Departament d'Econometria, Estadística i Economia Espanyola Universitat de Barcelona. Av. Diagonal, 690, 08034 Barcelona.
** Institut d'Estadística de Catalunya, Via Laietana, 58, 08003 Barcelona.
– Received October 2001.
– Accepted January 2002.

1. INTRODUCTION

Studies of the tourism sector in Catalonia have traditionally drawn on macro aggregates corresponding to each tourist season, such as private consumption and gross domestic product. In so doing they have tended to rely on one of the main data sources for this sector i.e. the survey of the supply of hotel accommodation in each of the Spanish regions. This survey, undertaken by the Spanish Statistical Institute (INE), provides information about hotel occupancy, in other words, information provided by the supply side of the tourism market.

In 1997, the Statistical Institute of Catalonia (Idescat) introduced a new survey of Catalan tourism, but in contrast to the survey described above this sought to obtain information from the demand side. This survey provides analysts with valuable information about visitors to Catalonia, whose point of origin is one of the other Spanish regions. Given its recent introduction, these statistics offer information for the most recent tourism seasons only and comparative data is only provided with the previous tourism seasons and, consequently, no time series is defined. It should, however, be noted that the time interval of each tourist season has changed since the introduction of the survey which hinders the definition of an appropriate time series. Thus, three tourism seasons were identified for 1997 and 1998: January to May, June to August and September to December; while for 1999 and 2000 four seasons were identified: January to April, May to June, July to August and September to December.

The recent introduction of the survey and the varying time intervals used in defining the tourist season hinder comparisons. Furthermore, although the survey was designed to embrace all the Spanish regions, only a few observations are eventually included within the database, and so the information describing individual characteristics tends to be highly heterogeneous. This heterogeneity becomes even more marked when the data are raised to the entire population.

Consequently using this database to calculate growth rates gives highly volatile time series. Therefore, the aim of this paper is to present a methodology for computing time series from the micro data (the survey) but, in contrast with the original (population-raised) time series, with a smoothed temporal pattern.

It is not, however, our aim to supply the analyst with a specific set of smoothed time series but rather to design a methodology that allows practitioners to obtain smoothed time series automatically whatever the concepts crossed in the database. The successful achievement of this goal depends on the application of simple smoothing methods that can be adequately employed in all cases.

This paper is organised as follows. In Section 2 we describe the database provided by the survey carried out by Idescat. We describe some of the characteristics of this database and define the time series that constitute the focus of this paper. Section 3 outlines the methodology that is applied in order to smooth these time series. This section contains three sub-sections that offer a detailed description of the transformations involved at each stage of our methodological proposal. Section 4 presents the results obtained. Finally, Section 5 concludes.

2. DESCRIPTION OF THE DATABASE

The availability of a database built upon the conducting of a survey at different points in time allows us to undertake the analysis at different time intervals. As a last resort, the information contained in the database can always be used to define a daily time series. However, problems arise owing to the absence of observations and distortions in the significance of the findings as the time frequency of the analysis increases.

For these two reasons, in this paper, the temporal reference is fixed at monthly intervals and the monthly time series is the basic information to which our methodology is applied. This specification allows us to use classical smoothing techniques including exponential smoothing and Holt-Winters smoothing procedures. In addition, we can obtain time series of varying temporal frequency (quarterly and annual) by aggregating monthly time series.

The micro database used here provides information about individual personal characteristics, including age, profession, marital status and region of residence. It also contains details about their holidays: number and characteristics of the other group members, destination, type of accommodation, amount of expenditure, number of days spent in Catalonia and the number of overnight stays, among others. However, here we focus on just two of these variables: (1) the number of overnight stays and (2) the number of tourists. Both variables are classified by type of accommodation (hotel, family and friends' households, other types of accommodation and total) and by destination (Barcelona, Costa Daurada, other destinations in Catalonia and all destinations in Catalonia). The different combinations give rise to forty time series.

The definition of these time series is strongly conditioned by the quality of the information comprising the tourism micro-database. Thus, firstly, although in aggregating the information we have tried to avoid missing or zero values, this has been unavoidable in certain periods for some time series. This might have a detrimental effect on the quality of the output following the application of the smoothing procedure. Secondly, graphic inspection of the time series indicates that there might be some outliers, the presence of which implies growth rates of doubtful validity. Finally, there would seem to be an Easter Week effect due to the fact it is a moveable feast and as such does not always occur in the same time period.

In this analysis these first two problems with the information are left for future consideration, particularly given that Idescat is planning to modify some of these anomalies. The third problem is discussed below.

3. METHODOLOGICAL PROPOSAL

In this section we present the methodology adopted in smoothing the time series described in the previous section. One of the reasons why these time series are apparently so erratic is that the survey loses precision as the geographical and conceptual range is increased. Our proposal tries to compensate for this absence of observations by increasing the amount of information used when estimating the micro data of one particular month.

The increase in the amount of information is achieved by the joint consideration of the micro-observations referring to the same month in two consecutive years. Thus, we compute the average number of tourists and overnight stays in the same month for two consecutive years and assign this mean value to these months. Hence, we take into account information that refers to two similar periods (month) and, as a consequence, we are able to reduce the volatility of the time series.

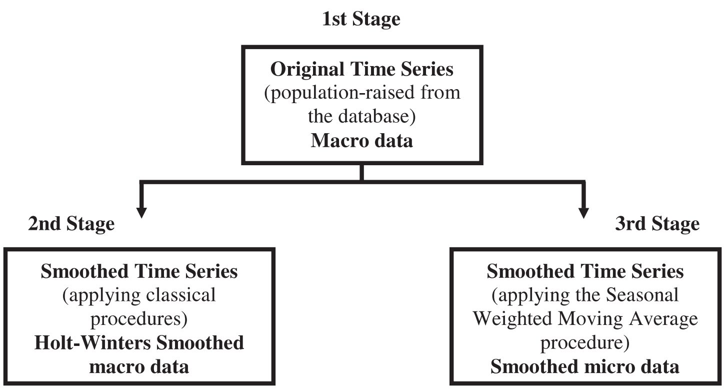


Figure 1. Brief description of the methodological proposal.

This simple method allows us to obtain time series that have a smoother pattern throughout the time period under consideration. The main problem arises, however, when deciding the weightings that should be applied when computing this mean value. One possibility is the specification of equal weights for each time period. Yet, it might be argued that a weighting system that gives greater weighting to more recent information is preferable to a system that attaches the same importance to the two sets of information.

If this is the case, the analyst needs to select these weightings. This is not, however, a straightforward decision, given that different weightings will result in different numbers of tourists and overnight stays. We try to overcome this drawback by suggesting a method by which the weightings can be estimated. The method comprises three stages.

3.1. First stage: The computation of the original time series

The approach relies on the definition of two sets of time series. The first set is the one defined by the original time series, that is, the time series that are derived from raising the data of the survey to the population. As mentioned in the previous section, the forty time series thus obtained are highly volatile over time, which is the problem that this paper seeks to rectify. The large number of observations available for the short period under analysis (forty-eight observations in just four years) means that the application of the stochastic approach to the modelling of these processes is not the most appropriate and that the classical approach should be the one to be adopted.

3.2. Second stage: Definition of the time series of reference

In the second stage of the analysis we obtain the set of forty time series following the application of a classical smoothing procedure to the original time series. This second set of smoothed time series serves as a *referent* for defining the system of weightings to be used when computing the average.

Before applying the classical methodological approach to the modelling of the time series we need to know the type of time series that is being dealt with. Here, the characterisation of the time series was performed using two test statistics. In order to decide the consideration of a time trend we applied the Daniel test, while for the seasonal component we applied the Kruskal-Wallis test. These tests indicated that in most cases the patterns of the time series are given by both components. Nevertheless, it should be noted that these results are not entirely reliable since these statistical tools are more suited to moderate or large sample sizes. Table A.1 in the Appendix shows the results of the application of both tests. The most appropriate smoothing procedure for these data is that of Holt-Winters since there are trend and seasonal components in most of the forty time series. Graphical inspection indicates that the additive model can provide a good fit, although this conclusion might need to be revised as further information comes available.

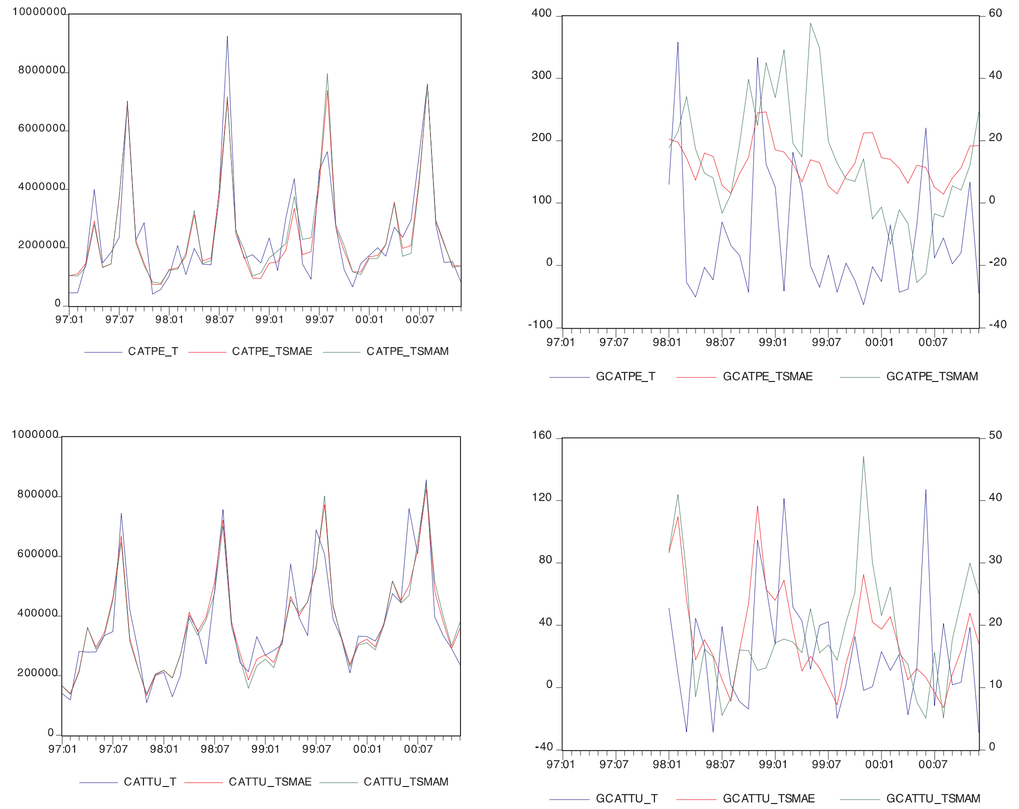
Before the Holt-Winters smoothing procedure can be applied to the time series under consideration, we need to analyse the effect of Easter Week on these time series. The only period that might have had an influence on the time series was in 1997. In this year Easter fell in the month of March while for the remaining years it fell in April. In order to avoid distortions that might affect the output of the smoothing procedure we decided to test for the presence of a 1997 Easter Week effect and, if there was found to be such an effect to correct the time series to take it into account.

This meant the estimation of a regression model that specifies the time series as a function of an independent term, a time trend, a seasonal dummies set, and an impulse dummy that captures the effect that can be assigned to March 1997. Only in three cases was this impulse dummy found to have a statistical significance of 10%, and the effect was corrected in each case. The three time series were the total numbers of overnight stays in Catalonia (CATPE_T), overnight stays with family or in a friend's household in Catalonia (CATPE_F) and overnight stays with family or in a friends' household using this minimisation criterion in Barcelona (BCNPE_F).

Table 1. Estimated coefficients for the Holt-Winters smoothing procedure.

	Overnight stays			*Tourists*		
	ALFA	BETA	GAMMA	ALFA	BETA	GAMMA
CAT_T	0	0	0	0	0	0
CAT_H	0.02	0.03	0	0.03	0	0
CAT_F	0	0	0	0.01	0	0
CAT_R	0	0	0	0	0	0
CAT_NH	0	0	0	0.01	0	0
BCN_T	0	0	0	0.02	0	0
BCN_H	0.01	0.09	0	0.26	0	0
BCN_F	0	0	0	0	0	0
BCN_R	0	0	0	0.01	0.18	0
BCN_NH	0	0	0	0.01	0.07	0
CD_T	0	0	0	0	0	0
CD_H	0.01	0.08	0	0	0	0
CD_F	0	0	0	0	0	0
CD_R	0	0	0	0	0	0
CD_NH	0	0	0	0	0	0
RD_T	0	0	0	0	0	0
RD_H	0	0	0	0	0	0
RD_F	0	0	0	0	0	0
RD_R	0	0	0	0	0	0
RD_NH	0.1	0	0	0	0	0

Note: *CAT_T* refers to all types of accommodation used in Catalonia (*CAT*). *CAT_H* indicates those people staying in a hotel. *CAT_F* those staying with family or in a friend's household. *CAT_R* denotes the other types of accommodation used. Finally, *CAT_NH* denotes those staying in accommodation other than a hotel. This notation is repeated for the territorial division considered here: *BCN*-Barcelona, *CD*-Costa Daurada and *RD*-remaining destinations.



Note: CATPE_T denotes the raw time series of the total number of overnight stays in Catalonia, CATPE_TSMAE denotes the smoothed time series using the Holt-Winters procedure with the estimated coefficients and CATPE_TSMAM is the smoothed time series using the Holt-Winters procedure with a fixed value for the coefficients. GCATPE_T, GCATPE_TSMAE and GCATPE_SMAM are the corresponding growth rates. CATTU_T, CATTU_TSMAE and CATTU_TSMAM refer to the tourist series.

Figure 2. Overnight stays and tourists in Catalonia. Levels and growth rates of the original and smoothed time series.

Once the time series affected by the Easter Week had been modified, we applied the Holt-Winters smoothing procedure to estimate the value of the parameters of the independent term (α), the slope (β) and the seasonal parameter (γ) using the criteria of the minimisation of the sum of squared residuals. The estimated coefficients for each time series are presented in Table 1. Note that although it is possible to fix the value of these parameters, the estimation provides a better fit.

As can be seen from Table 1, in most cases the estimated parameters equal zero, indicating that the corresponding component —independent term, trend and seasonality— is stable, that is, it does not vary during the time period under analysis.

The monthly time series for the level and rate of growth for the number of tourists visiting Catalonia and the number of overnight stays following the application of the Holt-Winters smoothing procedure are given in Figure 2. We denote these time series as the smoothed-HW time series. Each figure contains information about the original time series, the smoothed-HW time series with manual selection of the parameters and the smoothed-HW time series with the estimated parameters obtained using the minimisation of the squared sum of errors' criteria.

A number of comments should be made. First of all, it can be seen that the smoothed time series built on the use of the estimated coefficients show a smoother behaviour than those in which the value of such coefficients is imposed (in this case the parameter values were fixed at 0.1). Second, these results indicate that the estimation of the initial values used in obtaining the smoothed-HW time series influences the computation of the growth rates. Thus, for instance, we encounter a contradiction for the time series of tourists coming to Catalonia in which the type of accommodation is not specified (CATTU_T). In this case the growth rates computed using the original time series are negative, while with the smoothed-HW time series they are positive. This is also the case for the time series of tourists coming to Catalonia and staying in other types of accommodation (CATTU_RD) and overnight stays in hotel accommodation in Catalonia (CATPE_T). Third, and in contrast to the smoothed-HW time series, for some time series and periods the original time series present null values which means the growth rates are discontinuous in these cases.

3.3. Third stage: Application of the Seasonal Weighted Moving Average (SWMA) smoothing procedure

In the third stage of the analysis we select the weightings that best fit the time series smoothed in the previous stage. The estimation of these weightings is carried out by specifying the criteria of minimisation of the sum of squared errors, where the errors are given by the difference between the smoothed-HW macro time series and the weighted average time series —hereafter smoothed-micro time series. This section describes the methodology adopted in this optimisation procedure.

Once the macro time series in question has been smoothed using the Holt-Winters procedure, employing an additive specification and by estimating the parameters of the model, we proceeded to select the set of weights used in the procedure applied in this paper to the micro series. This procedure can be understood as the computation of seasonal weighted moving averages (SWMA). For instance, in computing the smoothed time series of tourists for January 1998 using the SWMA procedure we need to take into account the information of the original time series of tourists that corresponds to January 1997 and January 1998. To compute the observation for February 1998 of the smoothed time series we need to look at the observations of the original time series referring to February 1997 and February 1998, and so on.

The important aspect of our proposal is the system of weightings applied in computing the average. As mentioned above, different smoothed time series are obtained depending on the set of weights used. The greater the weight given to more recent values in the time series, the more the smoothed-micro time series tends to resemble the original time series. In the first stage we smoothed the original time series by applying five sets of weights: 50/50, 40/60, 30/70, 20/80 and 10/90. Yet, in order to avoid being subjective when selecting the system of weightings, we estimated this parameter through the minimisation of the sum of squared error of the difference between the smoothed time series using the SWMA procedure —the smoothed-micro time series— and the smoothed-HW time series. This estimation can be outlined as follows.

If we denote the original time series by Y, the smoothed-HW time series by Y^* and the smoothed-micro time series using the SWMA procedure by $\hat{Y}$, the target function to be minimised is the function given by:

$$f(p) = \sum_{i=s+j}^{T} (y_i^* - \hat{y}_i)^2,$$

$j = \{1, 2, \ldots, s\}$, where s denotes the order of seasonality —here $s = 12$. The smoothed-micro time series is computed from:

$$\hat{y}_i = p y_i + (1-p) y_{i-s},$$

where p is the weight (parameter) to be estimated. Therefore, the optimisation program can be represented by:

$$\min \sum_{i=s+j}^{T} (y_i^* - (p y_i + (1-p) y_{i-s}))^2$$

$$\text{subject to} \quad 0 \leq p \leq 1.$$

The necessary condition establishes that:

$$\frac{\partial f(p)}{\partial p} = \sum_{i=s+j}^{T} 2 (y_i^* - p y_i - (1-p) y_{i-s}) (-y_i + y_{i-s}) = 0,$$

so that the optimal weights are given by:

$$\hat{p} = \frac{-\sum_{i=s+j}^{T} y_i^* \left(-y_i + y_{i-s}\right) + \sum_{i=s+j}^{T} y_{i-s}\left(-y_i + y_{i-s}\right)}{-\sum_{i=s+j}^{T} y_i \left(-y_i + y_{i-s}\right) + \sum_{i=s+j}^{T} y_{i-s}\left(-y_i + y_{i-s}\right)}. \tag{1}$$

The second derivative indicates that the solution corresponds to a minimum:

$$\begin{aligned}\frac{\partial^2 f(p)}{\partial p^2} &= \sum_{i=s+j}^{T} \left(-2y_i\left(-y_i + y_{i-s}\right) + 2y_{i-s}\left(-y_i + y_{i-s}\right)\right) \\ &= 2\sum_{i=s+j}^{T} \left(-y_i + y_{i-s}\right)^2 > 0.\end{aligned}$$

Table 2. Optimal weight (p) for the SWMA.

	Overnight stays	*Tourists CAT_F 0.53*
CAT_F	0.53	0.53
CAT_H	0.52	0.47
CAT_NH	0.51	0.51
CAT_R	0.49	0.51
CAT_T	0.53	0.57
BCN_F	0.48	0.48
BCN_H	0.53	0.57
BCN_NH	0.50	0.47
BCN_R	0.68	0.48
BCN_T	0.52	0.59
CD_F	0.56	0.51
CD_H	0.51	0.42
CD_NH	0.44	0.46
CD_R	0.43	0.48
CD_T	0.54	0.43
RD_F	0.52	0.49
RD_H	0.47	0.44
RD_NH	0.49	0.50
RD_R	0.49	0.51
RD_T	0.50	0.47

Note: CAT_T refers to all types of accommodation used in Catalonia (CAT). CAT_H indicates those people staying in a hotel. CAT_F those staying with family or in a friend's household. CAT_R denotes the other types of accommodation used. Finally, CAT_NH denotes those staying in accommodation other than a hotel. This notation is repeated for the territorial division considered here: BCN-Barcelona, CD-Costa Daurada and RD-remaining destinations.

As can be shown, the weight can be easily computed from (1) using the original and the smoothed-HW time series. The application of this procedure is undertaken in the next section.

4. RESULTS

The weights computed for the time series in question are shown in Table 2.

These weights were applied after fixing the data corresponding to the year 2000 as the base. Therefore, the values of the original and the smoothed time series using the SWMA procedure for the months of 2000 are the same. The decision to choose this year as the point of reference was made following advice received from experts at Idescat. Tables A.2 and A.3 present the original (Y), the smoothed-HW time series (Y^*) and smoothed-micro quarterly time series ($\hat{Y}$) and their growth rate, for the time series of overnight stays and tourists.

A comparison of the original and the smoothed time series (SWMA) reveals certain discrepancies. These differences are presented in Table 3 for Catalonia.

Table 3. Overnight stays and tourists in Catalonia. Original and smoothed time series.

	Overnight stays	*Overnight stays (smoothed)*	*Tourist*	*Tourist (smoothed)*
Annual time series				
1997	25,148,935	27,583,868	3,538,870	3,773,680
1998	29,746,499	29,491,193	3,947,711	4,388,195
1999	30,991,276	32,897,103	5,045,885	5,451,129
2000	32,862,961	32,862,961	5,420,252	5,420,252
Differences				
1997		−2,434,933		−234,811
1998		255,306		−440,484
1999		−1,905,827		−405,244
2000		0		0

Although these differences are not great, they can be proportionally distributed within each period in order to obtain the same number of overnight stays and tourists for the original and smoothed time series.

Table A.1. Daniel and Kruskal-Wallis Tests.

	Test of Daniel	*Probability*	*Test of Krusal-Wallis*	*Probability*		
CATPE_F	1.56	0.94	20.73	0.96	Trend	Seasonality
CATPE_H	2.22	0.99	30.42	1.00	Trend	Seasonality
CATPE_NH	1.01	0.84	25.46	0.99	Trend	Seasonality
CATPE_R	1.08	0.86	30.25	1.00	Trend	Seasonality
CATPE_T	1.64	0.95	26.70	0.99	Trend	Seasonality
CATTU_F	3.16	1.00	19.09	0.94	Trend	Seasonality
CATTU_H	3.07	1.00	22.34	0.98	Trend	Seasonality
CATTU_NH	2.79	1.00	29.13	1.00	Trend	Seasonality
CATTU_R	1.50	0.93	28.67	1.00	Trend	Seasonality
CATTU_T	3.22	1.00	29.45	1.00	Trend	Seasonality
BCNPE_F	1.47	0.93	20.69	0.96	Trend	Seasonality
BCNPE_H	1.88	0.97	19.27	0.94	Trend	Seasonality
BCNPE_NH	0.65	0.74	18.17	0.92	Trend	Seasonality
BCNPE_R	−0.59	0.28	17.72	0.91	Trend	Seasonality
BCNPE_T	1.55	0.94	21.04	0.97	Trend	Seasonality
BCNTU_F	2.89	1.00	17.30	0.90	Trend	Seasonality
BCNTU_H	3.08	1.00	9.88	0.46	Trend	Seasonality
BCNTU_NH	2.85	1.00	18.88	0.94	Trend	Seasonality
BCNTU_R	−0.23	0.41	11.48	0.60	Trend	Seasonality
BCNTU_T	4.00	1.00	13.00	0.71	Trend	Seasonality
CDPE_F	0.64	0.74	25.35	0.99	Trend	Seasonality
CDPE_H	2.59	1.00	29.41	1.00	Trend	Seasonality
CDPE_NH	0.94	0.83	32.58	1.00	Trend	Seasonality
CDPE_R	0.82	0.79	33.99	1.00	Trend	Seasonality
CDPE_T	1.72	0.96	33.89	1.00	Trend	Seasonality
CDTU_F	0.50	0.69	22.23	0.98	Trend	Seasonality
CDTU_H	2.30	0.99	33.53	1.00	Trend	Seasonality
CDTU_NH	0.78	0.78	35.50	1.00	Trend	Seasonality
CDTU_R	0.99	0.84	34.91	1.00	Trend	Seasonality
CDTU_T	1.87	0.97	36.42	1.00	Trend	Seasonality
RDPE_F	1.34	0.91	13.67	0.75	Trend	Seasonality
RDPE_H	1.49	0.93	24.90	0.99	Trend	Seasonality
RDPE_NH	1.63	0.95	21.46	0.97	Trend	Seasonality
RDPE_R	1.61	0.95	19.01	0.94	Trend	Seasonality
RDPE_T	1.85	0.97	22.12	0.98	Trend	Seasonality
RDTU_F	2.41	0.99	11.90	0.63	Trend	Seasonality
RDTU_H	2.20	0.99	23.06	0.98	Trend	Seasonality
RDTU_NH	2.81	1.00	17.61	0.91	Trend	Seasonality
RDTU_R	1.75	0.96	19.26	0.94	Trend	Seasonality
RDTU_T	2.77	1.00	23.91	0.99	Trend	Seasonality

Note: CAT_T refers to all types of accommodation used in Catalonia (CAT). CAT_H indicates those people staying in a hotel. CAT_F those staying with family or in a friend's household. CAT_R denotes the other types of accommodation used. Finally, CAT_NH denotes those staying in accommodation other than a hotel. This notation is repeated for the territorial division considered here: BCN-Barcelona, CD-Costa Daurada and RD-remaining destinations.

Table A.2. Overnight stays in Catalonia Level and growth rate of the quarterly time series.

	Total	*Total (SWMA)*	*Total (HW)*	*Hotel*	*Hotel (SWMA)*	*Hotel (HW)*	*Family*	*Family (SWMA)*	*Family (HW)*	*Other*	*Other (SWMA)*	*Other (HW)*	*No Hotel*	*No Hotel (SWMA)*	*No Hotel (HW)*
Jan-April97	6,371,284	6,258,097	6,532,288	1,348,147	1,211,671	1,218,876	4,250,680	4,378,042	4,100,855	688,708	636,305	1,199,149	5,006,671	5,040,412	5,316,825
May-Aug97	12,695,217	14,497,545	13,429,469	3,077,329	3,712,300	3,532,048	4,367,604	5,230,961	4,705,140	5,304,046	5,547,960	5,213,216	9,616,045	10,719,327	9,904,455
Sep-Dec97	6,082,434	6,828,226	5,187,177	1,716,735	1,887,964	1,395,263	2,272,053	3,242,448	2,084,342	2,123,632	1,737,762	1,704,021	4,384,007	4,920,523	3,785,444
Jan-April98	6,157,567	8,684,303	7,389,402	1,084,306	1,404,618	1,722,879	4,490,759	5,481,308	4,471,478	582,502	1,709,680	1,157,341	5,073,261	7,176,905	5,645,640
May-Aug98	16,098,317	14,079,534	14,286,583	4,304,880	3,854,960	3,996,419	5,995,048	5,512,106	5,075,763	5,798,388	4,795,953	5,171,408	11,793,437	10,302,730	10,233,270
Sep-Dec98	7,490,615	6,727,355	6,044,291	2,047,762	1,908,446	1,900,333	4,101,267	3,308,045	2,454,965	1,341,586	1,501,159	1,662,214	5,442,853	4,849,087	4,114,259
Jan-April99	10,928,471	9,455,646	8,246,516	1,703,545	2,365,957	2,221,540	6,357,964	5,478,901	4,842,101	2,866,962	1,680,558	1,115,534	9,224,926	7,166,946	5,974,455
May-Aug99	12,286,515	15,372,581	15,143,698	3,435,076	4,531,996	4,467,985	5,084,693	5,978,805	5,446,386	3,766,746	4,763,921	5,129,600	8,851,439	10,729,468	10,562,085
Sep-Dec99	6,049,452	6,342,039	6,901,405	1,778,431	2,077,740	2,258,853	2,606,028	2,685,014	2,825,588	1,664,993	1,578,610	1,620,406	4,271,021	4,257,754	4,443,075
Jan-April00	8,147,529	8,147,529	9,103,630	2,984,146	2,984,146	2,568,869	4,700,913	4,700,913	5,212,724	462,470	462,470	1,073,726	5,163,382	5,163,382	6,303,270
May-Aug00	18,113,527	18,113,527	16,000,812	5,555,687	5,555,687	4,912,999	6,770,112	6,770,112	5,817,009	5,787,728	5,787,728	5,087,792	12,557,840	12,557,840	10,890,900
Sep-Dec00	6,601,905	6,601,905	7,758,519	2,357,068	2,357,068	2,807,356	2,754,919	2,754,919	3,196,211	1,489,919	1,489,919	1,578,598	4,244,838	4,244,838	4,771,890
							INTER-ANNUAL GROWTH RATES								
Jan-April98	−3.35%	38.77%	13.12%	−19.57%	15.92%	41.35%	5.65%	25.20%	9.04%	−15.42%	168.69%	−3.49%	1.33%	42.39%	6.18%
May-Aug98	26.81%	−2.88%	6.38%	39.89%	3.84%	13.15%	37.26%	5.37%	7.88%	9.32%	−13.55%	−0.80%	22.64%	−3.89%	3.32%
Sep-Dec98	23.15%	−1.48%	16.52%	19.28%	1.08%	36.20%	80.51%	2.02%	17.78%	−36.83%	−13.62%	−2.45%	24.15%	−1.45%	8.69%
Jan-April99	77.48%	8.88%	11.60%	57.11%	68.44%	28.94%	41.58%	−0.04%	8.29%	392.18%	−1.70%	−3.61%	81.83%	−0.14%	5.82%
May-Aug99	−23.68%	9.18%	6.00%	−20.21%	17.56%	11.80%	−15.19%	8.47%	7.30%	−35.04%	−0.67%	−0.81%	−24.95%	4.14%	3.21%
Sep-Dec99	−19.24%	−5.73%	14.18%	−13.15%	8.87%	18.87%	−36.46%	−18.83%	15.10%	24.11%	5.16%	−2.52%	−21.53%	−12.19%	7.99%
Jan-April00	−25.45%	−13.83%	10.39%	75.17%	26.13%	15.63%	−26.06%	−14.20%	7.65%	−83.87%	−72.48%	−3.75%	−44.03%	−27.96%	5.50%
May-Aug00	47.43%	17.83%	5.66%	61.73%	22.59%	9.96%	33.15%	13.24%	6.80%	53.65%	21.49%	−0.82%	41.87%	17.04%	3.11%
Sep-Dec00	9.13%	4.10%	12.42%	32.54%	13.44%	24.28%	5.71%	2.60%	13.12%	−10.52%	−5.62%	−2.58%	−0.61%	−0.30%	7.40%

Note: Total indicates all types of accommodation. Hotel indicates those people staying in a hotel. Family denotes those staying with family or in a friend's household. Other denotes all other types of accommodation. Finally, No hotel denotes those that do not use hotel accommodation. (SWMA) indicates that the time series has been smoothed using the SWMA procedure whereas (HW) indicates that the time series has been smoothed using the Holt-Winters procedure.

Table A.3. Tourists in Catalonia Level and growth rate of the quarterly time series.

	Total	*Total (SWMA)*	*Total (HW)*	*Hotel*	*Hotel (SWMA)*	*Hotel (HW)*	*Family*	*Family (SWMA)*	*Family (HW)*	*Other*	*Other (SWMA)*	*Other (HW)*	*No Hotel*	*No Hotel (SWMA)*	*No Hotel (HW)*
Jan-April97	817,174	889,129	882,127	305,527	306,255	336,181	361,723	448,344	390,952	145,427	127,831	152,941	508,722	573,268	544,303
May-Aug97	1,707,462	1,780,283	1,770,938	624,757	639,031	683,393	532,890	599,846	593,934	546,361	533,261	486,036	1,074,600	1,127,987	1,078,180
Sep-Dec97	1,014,234	1,104,269	885,804	479,450	458,087	383,322	428,460	484,085	334,009	114,275	159,325	167,086	545,814	643,047	503,136
Jan-April98	942,458	1,224,179	1,091,170	307,064	358,904	447,206	524,798	623,715	481,041	110,596	209,049	157,426	635,394	829,592	637,263
May-Aug98	1,834,254	1,944,690	1,979,981	654,881	706,447	782,052	658,944	770,472	684,604	520,429	456,523	490,521	1,179,373	1,221,901	1,172,151
Sep-Dec98	1,170,999	1,219,326	1,094,847	434,364	495,208	467,121	533,182	490,147	424,160	203,453	222,022	171,571	736,635	714,082	597,347
Jan-April99	1,432,977	1,466,439	1,300,213	416,467	586,352	529,390	711,022	651,277	572,472	305,488	210,624	161,911	1,016,510	863,509	737,902
May-Aug99	2,026,540	2,396,579	2,189,023	763,706	954,751	864,356	868,911	913,612	784,305	393,923	473,211	495,006	1,262,834	1,385,692	1,284,163
Sep-Dec99	1,255,144	1,256,887	1,303,889	562,770	543,022	556,159	452,162	509,035	524,688	240,212	208,709	176,056	692,374	715,156	703,535
Jan-April00	1,491,239	1,491,239	1,509,255	774,994	774,994	634,810	598,545	598,545	664,605	117,700	117,700	166,396	716,245	716,245	835,520
May-Aug00	2,670,835	2,670,835	2,398,066	1,166,891	1,166,891	1,007,511	953,067	953,067	867,382	550,877	550,877	499,491	1,503,944	1,503,944	1,369,178
Sep-Dec00	1,258,178	1,258,178	1,512,932	521,094	521,094	703,585	559,234	559,234	609,888	177,851	177,851	180,541	737,084	737,084	795,424
							INTER-ANNUAL GROWTH RATES								
Jan-April98	15.33%	37.68%	23.70%	0.50%	17.19%	33.03%	45.08%	39.12%	23.04%	−23.95%	63.53%	2.93%	24.90%	44.71%	17.08%
May-Aug98	7.43%	9.23%	11.80%	4.82%	10.55%	14.44%	23.65%	28.45%	15.27%	−4.75%	−14.39%	0.92%	9.75%	8.33%	8.72%
Sep-Dec98	15.46%	10.42%	23.60%	−9.40%	8.10%	21.86%	24.44%	1.25%	26.99%	78.04%	39.35%	2.68%	34.96%	11.05%	18.72%
Jan-April99	52.05%	19.79%	19.16%	35.63%	63.37%	18.38%	35.48%	4.42%	19.01%	176.22%	0.75%	2.85%	59.98%	4.09%	15.79%
May-Aug99	10.48%	23.24%	10.56%	16.62%	35.15%	10.52%	31.86%	18.58%	14.56%	−24.31%	3.66%	0.91%	7.08%	13.40%	9.56%
Sep-Dec99	7.19%	3.08%	19.09%	29.56%	9.66%	19.06%	−15.20%	3.85%	23.70%	18.07%	−6.00%	2.61%	−6.01%	0.15%	17.78%
Jan-April00	4.07%	1.69%	16.08%	86.09%	32.17%	19.91%	−15.82%	−8.10%	16.09%	−61.47%	−44.12%	2.77%	−29.54%	−17.05%	13.23%
May-Aug00	31.79%	11.44%	9.55%	52.79%	22.22%	16.56%	9.69%	4.32%	10.59%	39.84%	16.41%	0.91%	19.09%	8.53%	6.62%
Sep-Dec00	0.24%	0.10%	16.03%	−7.41%	−4.04%	26.51%	23.68%	9.86%	16.24%	−25.96%	−14.79%	2.55%	6.46%	3.07%	13.06%

Note: Total indicates all types of accommodation. Hotel indicates those people staying in a hotel. Family denotes those staying with family or in a friend's household. Other denotes all other types of accommodation. Finally, No hotel denotes those that do not use hotel accommodation. (SWMA) indicates that the time series has been smoothed using the SWMA procedure whereas (HW) indicates that the time series has been smoothed using the Holt-Winters procedure.

5. CONCLUSIONS

In this paper we propose a new approach to smoothing the Catalan tourism time series. In short, this procedure increases the information for one tourism season by taking into account information obtained by survey for that specific season and information referring to the same season but in the previous year. The application of the Seasonal Weighted Moving Average (SWMA) smoothing method means that, when determining the micro-data for one specific time series, we need to take into account the information for two similar time periods. As a result of this, the smoothed time series are smoother than the original series. The application of this procedure to the Catalan tourism micro-data time series reduces the volatility present in the original time series. As Costa *et al*. (2001) demonstrate elsewhere in this issue, it allows researchers to analyse the evolution between seasons without the distortions caused by an absence of information.

The main advantage of this approach is that once the weighting system has been defined, it can be applied to any micro-data time series that can be built from the database. Moreover, our proposal is not limited to just this database as it can also be applied to other micro databases that contain information of considerable interest to economists. These include the quarterly Active Population Survey (EPA) of the Spanish and Catalan economies, the monthly Overnight Stays in Hotels Survey that is conducted among the Hotels of the different Spanish regions and the quarterly Household Budget Continuous Survey, to name just a few.