4th International Forum on Tourism Statistics
Copenhagen, 17-19 June 1998

FORECASTING TOURISM DEMAND IN THE SHORT TERM: THE CASE OF ANDALUSIAN HOTEL ESTABLISHMENTS

José M. Otero and F. Trujillo

University of Málaga

Abstract

Policy makers and businessmen demand real-time information and short-term forecasts about bed-nights occupied and the occupancy rate in hotel establishments (the latter being the percentage ratio of the number of bed nights occupied to the number of bed-nights on offer in the same period, usually one month). So, when these forecasts are obtained using single equations time series models, implicit forecasts of bed nights on offer, i.e., bed capacity, can also be derived. These implicit forecasts of bed capacity, which we call indirect forecasts, yield a higher RMSE than the "direct" forecasts when a univariate time series model is used. To face this problem within the single equation time series modelling framework, the only one feasible due to sample size limitations, we propose obtaining forecasts by the following method. First, a set of "direct" forecasts are obtained for each one of the three variables involved (bed nights occupied, bed capacity, and occupancy rate) using transfer functions and ARIMA models, according to the case. When obtaining direct estimations, alternative models are considered taking into account aggregation problems and different leading indicators related to expectations of the business people. Second, an "indirect" forecast is obtained for each variable using the "direct" forecasts corresponding to the other two variables and taking into account the relationship which defines occupancy rate. Third, a set of "combined" forecasts, defined as the quadratic mean of direct and indirect forecasts, is obtained. This paper shows that: i) the combined forecasts of the bed-nights occupied and the occupancy rate perform better than the corresponding direct and indirect forecasts under the RMSE criteria, while the direct forecast is the best option for bed capacity; ii) the selected forecasts for the three variables, according to this criteria, fulfil the relationship established by the definition of occupancy rate.

FORECASTING TOURISM DEMAND IN THE SHORT TERM: THE CASE OF ANDALUSIAN HOTEL ESTABLISHMENTS[*]

José M. Otero and Francisco Trujillo[2]

University of Málaga
Facultad de Ciencias Económicas y Empresariales
El Ejido s/n
29015 MALAGA

1.INTRODUCTION.

A set of statistical and econometric models, aimed at obtaining short and mid-term forecasts of the main relative magnitudes regarding tourism in Andalusia, has been developed within the SAETA project[3].

The forecasts have, as is well-known, a marked interest for all managers in this sector, whether public or private. The short-term forecast, usually obtained by time series modelling, complements the available statistical information, which is always released after delays and so lags behind events, and allows us to carry out short-term analyses of the evolution of the sector that guides the managers' decisions. On the other hand, mid-term forecasts can be made via econometric models which quantify the effects on tourist demand of variables such as

the income level of tourist source countries, the exchange rate, and relative domestic prices and those of competitor countries. These allow us to predict, for example, the impact that a rise in prices or a change in the peseta exchange rate would have on demand[4].

This article presents the methodology developed within the SAETA project to design a short-term forecasting method for hotel occupancy rate and other related variables, such as

[*] This work has been partly supported with funds from DGICYT, reference PB96-0707.

[2] In collaboration with Juan Auzmendi.

[3] This project is supported by the Tourist Board of Andalusia (Consejería de Turismo y Deporte de la Junta de Andalucía). SAETA stands for Sistema de Análisis y Estadísticas del Turismo Andaluz (System of Analysis and Statistics of Andalusian Tourism).

[4] See Otero (1996) for mid-term forecasting with econometric models applied to Andalusia. For Spain, see Gonzalez and Moral (1995). An exhaustive survey on forecasting tourist demand is that of Witt and Witt (1995).

bed-capacity and bed-nights. This general objective raised a set of specific questions which we try to answer here.

On one hand, we have investigated whether the expectations of the business people (a variable about which statistical information exists) might serve as a leading indicator of occupancy rate and, if so, whether the corresponding time-lag differs from one Andalusian province to another. The same question holds for bed-nights and bed capacity, although in the latter case an additional question was raised. In effect, as the data available regarding Spanish and foreigner bed-nights is disaggregated, whose pattern of evolution over time does not usually coincide, we wonder if it would be better to predict each component separately and put the corresponding results together afterwards, or whether to directly obtain the forecasts for the aggregate via a model which would incorporate the expectations of business people as an explanatory variable.

On the other hand, given that the occupancy rate is defined as the ratio between bed-nights and bed-capacity available monthly, it has been necessary to design a system which carries out forecasts that take into account the relationship between these variables. Otherwise, if we limit ourselves to searching for optimum forecasts of bed-nights, bed capacity and occupancy rate via independent models, the corresponding results obtained separately will not necessarily satisfy the relationship existing between these variables.

All these models have taken into account the possible presence of outliers and the 'Easter effect', that is, the fact that Easter Week does not fall in the same month each year.

This article is organised as follows. In the first part the characteristics of the available data are described. Following this there is a description of the methodology used to establish a forecasting system which ensures coherence between bed-nights, bed capacity, and occupancy rate forecasts. Next, we present the main results relating to the methods and alternative models used for forecasting the three variables. Finally, we sum up with the main conclusions.

It must be borne in mind that the results obtained with the models considered here form the basis for putting into effect a coherent forecasting system for variables related to Andalusian hotel tourist flow. However, in order to obtain the forecasts the SAETA project is required to generate each month, extra information is used. The models analysed here allow us to obtain point and interval forecasts. Once the extra information becomes available, the final forecasts are placed within the corresponding interval forecasts but shifted from the centre of the interval towards the upper or lower limit by an amount which depends on the nature and characteristics of the available complementary information.

The complementary information used by SAETA is normally of two types: that systematically generated by AENA (the official Spanish airports agency) concerning runway reservations, and that normally offered by tour operators regarding their own expectations. The latter information has to be approached with caution, as the nature of the data is often not objective and can be influenced by the self-interest of the business.

2. DATA

The basic variables used in this research, which always relates to hotel establishments, are those referring to the following concepts:

- Bed-nights occupied in a given period;
- Bed capacity i.e., bed-nights on offer in the same period;
- Bed occupancy rate;
- Expectations of the business people.

Bed occupancy rate is the percentage ratio of the number of bed-nights occupied during a given period, normally a month, to the total bed capacity on offer during the same month. The performance of the hotel business is closely linked to this indicator, hence its practical interest.

The statistical data relative to these variables are those published by INE (The Spanish National Institute of Statistics) in their work regarding the survey *Movimiento de Viajeros en Establecimientos Hoteleros. (Tourist Flow in Hotel Establishments).*

The expectations published in the tourist flow statistics fall into two categories, depending on the type of questions asked in the questionnaires: subjective and objective. Subjective expectations were created from the replies the hotelier gave about the occupancy rate expected over the next six months – high, medium, or low – which were assigned values 1, 0, and –1 respectively. The figures in the tables are averages of individual replies.

Data regarding objective expectations was created by asking about the actual number of beds already booked for the third month after the month including the reference week. This data is expressed as a percentage of the number of existing beds according to the official directory of hotels.

The available statistical information consists of a monthly time series which began in January 1977 for bed-nights, in January 1979 for occupancy rate, and, much more recently, in January 1992 for expectations. In the latter case, we encounter a scarcity of information for identifying and suitably estimating the time series models.

An additional problem concerning data about expectations arises due to the fact that it is published several months after the other variables, so its practical use as a leading indicator for occupancy rate and bed-nights is considerably restricted.

3. METHODOLOGY

According to the above definition of bed occupancy rate, each of the three variables involved can be predicted in the three following alternative ways that we denominate as ‘direct’, ‘indirect’, and ‘combined’.

The direct forecast of occupancy rate is made in the light of historical information about this variable (including trends, cycles, and seasonal variations) as well as about other variables which influence it. In the latter category we could highlight, for example, the intervention variable aimed at reflecting the "Easter effect", and business peoples' expectations, which could be used as a leading indicator. In a similar way, "direct" forecasts for bed-nights and bed capacity can be obtained.

The "indirect" forecast for bed occupancy rate is obtained from the ratio between direct forecasts of bed-nights occupied and bed capacity, calculated using independent models for each of them.

The "combined" forecast is obtained by combining direct and indirect forecasts in some way. The copious literature about combining forecasts[5] shows us that, in general, a combination of two or more forecasts obtained from an average, for example, is usually better than any obtained separately. For reasons of internal coherence which will be addressed later, the geometric mean of direct and indirect forecasts as combined forecasts is used here.

Although the aim of short-term analysis is forecasting bed occupancy rate and bed-nights, it is clear that when these two variables are forecasted, we implicitly obtain, a forecast of bed capacity due to the deterministic relationship existing between these three variables. For this reason, it is necessary to design a system which allows us to obtain the best forecasts for the three variables simultaneously.

The methodology used in this research is as follows. First, various alternative time series models were tried for predicting ("directly") each of the three variables involved in the analysis: bed-nights, bed capacity, and bed occupancy rate.

Second, "indirect" forecasts were obtained for each of the mentioned variables from the best forecasts of the other two, using to this effect the relationship existing between the three variables.

Finally, combined forecasts were obtained for each variable using the geometric mean of the corresponding direct and indirect forecasts.

In order to choose between alternative forecasts for each of the individual variables, a set of *ex post* forecasts was obtained to determine which of them produced the minimum root mean square error (RMSE) of the forecast.

One of the basic conclusions of this work is that the best forecasts of bed occupancy rate are obtained by combining direct and indirect forecasts.

---

[5] See the special issue of *International Economic Review*, vol. 5, num 4, 1989, dedicated to this subject and especially the excellent article by Clement (1989).

As mentioned, in the case of bed-nights it was necessary to take into account the fact that direct forecasting could be done in two alternative ways, depending on whether or not Spaniards and foreigners were disaggregated. In our case, the best forecasts were those corresponding to an aggregated model. On the other hand, as in the previous case, when it came to choose between direct, indirect, and combined forecasts, the latter turned out to be the best.

As regards bed capacity, direct forecasts were so much better than indirect ones that, in this case, the combination of both was worse than direct ones alone.

In order to ensure coherence in the results it was necessary to verify that the ratio between the combined forecast of bed-nights and the combined forecast of bed occupancy rate coincided with the direct forecast of bed capacity. As shown in the Appendix, this result is kept as long as combined forecasts are generated with the geometric mean of the corresponding direct and indirect forecasts.

4. DIRECT FORECASTS OF BED OCCUPANCY RATES AND BED CAPACITY

The aim of this section is to look into the possibility that expectations might serve as a leading indicator in forecasting occupancy rate. As explained previously, data of two types of expectations are available, subjective (EXPECTS), and objective (EXPECTO). Given the nature of the data, subjective expectations were are expected to have a better relationship to bed occupancy rate, as these are created on the basis of hoteliers' replies regarding expected occupancy rates during the next six months.

In an analysis carried out by González (1995), it was concluded that subjective expectations can be used as leading indicators of occupancy rate with a lead of three months. This result showed that, although the question referring to expectations concerned the next six months, the hoteliers' reply enabled a lookahead of only three months.

In the present work we undertake the same issue, but try to deal with it using statistical inference methods (transfer function models). We use AUTOBOX, a commercial expert program which automatically identifies the transfer function model that relates bed occupancy rate to expectations. It therefore automatically selects the lag best suited for the explanatory variable, considered as a stochastic variable, while taking into account other intervention variables, among which the one related to the "Easter effect" is mentioned.

The first conclusion we draw from this analysis of Andalusia as a whole, is that EXPECTS works optimally with a lag of two months.

Taking into account that the data about Andalusia is an aggregate of eight different areas (its provinces), and that the behaviour of each one would presumably be different depending whether they were coastal or inland provinces, we analysed each of them independently. The results were rather discouraging because, although there is evidence supporting a lag of three months in the case of Seville, the remaining provinces have yielded inconclusive results. The basic reason for these fuzzy results is the model's lack of robustness due to the fact that the sample size is too small for successfully identifying complex transfer functions, together with the Easter effect, and the presence of *outliers*. In fact, when tests were carried out in the absence of certain sample observations, fairly different results were obtained, which further indicates the model's lack of robustness.

On the other hand, we carried out a similar analysis with the EXPECTO variable. Although the question given to the hoteliers in this case was related to bed-nights with a three-month lookahead and not to occupancy rate, a certain degree of correlation is expected between these variables, and thus the possibility of EXPECTO working as a leading indicator for bed occupancy rate.

The main conclusion reached in this case is that EXPECTO is a leading indicator for occupancy rate in Andalusia as a whole, with a three month lag. The analysis by province yielded similar results for those in the interior, but not for those on the coast in which the statistically more acceptable lags were less than three months. This result is illogical and can be attributed again to the lack of robustness of the models due to the extremely small sample size.

It only remains to mention a final effort aimed at selecting the best model to obtain direct forecasts. That is, to know which of the two leading indicators, EXPECTS and EXPECTO, is the most suitable to include in our forecast model. To answer this question we formulated two models and made use of the automatic identification option of the transfer function in the AUTOBOX program, then obtained the corresponding *ex post* forecasts for the last twelve months and calculated their RMSE which are shown in Table 1.

Table 1
FORECASTING PERFORMANCE
OF OCCUPANCY RATE MODELS

Model	**RMSE**
Model 1 (without expectations)	5.18
Model 2 (with objective exp.)	5.24
Model 3 (with subjective exp.)	4.94

In this table, model 1 corresponds to an ARIMA occupancy rate model, taking into account the Easter effect and the existence of *outliers*, but without considering expectation variables as leading indicators. Model 2 includes EXPECTO as a leading indicator, and model 3 uses EXPECTS.

In order to automatically identify and estimate the models, 55 observations corresponding to the period January 1992 to July 1996 have been used. The data of August 1996 to July 1997 was used to calculate the RMSE of the *ex post* forecasts.

The results in Table 1 show that as the model with subjective expectations is a better predictor than the other two, so it was selected to carry out direct forecasts of occupancy rate.

It can also be observed that the model including objective expectations as a leading indicator is the worst predictor of the three which led us to discard it as a predictor. This result could once again be attributed to the lack of robustness of the transfer function models due to the small sample size.

Finally, we would like to mention the tests carried out to model bed capacity. Although in principle we did not expect this variable to be influenced by expectations in the same way as bed occupancy rate, we did not want to discard this possibility without carrying out appropriate tests. Thus, similar to occupancy rate, we have used three alternative models to forecast bed capacity, namely: i) an ARIMA model; ii) a transfer function model with bed capacity depending on subjective expectations (EXPECTS); and iii) a transfer function model with the number of beds depending on objective expectations (EXPECTO).

The results of this analysis, which we omit here, is that the best forecasts are those obtained with an ARIMA model using an intervention variable to pick up the Easter effect. This is used later to obtain the corresponding direct forecasts. Once again, in this case, problems were encountered regarding the lack of robustness in the transfer function models which included expectation variables.

5. BED-NIGHT FORECAST

Given the available statistical information, the following two questions can be raised regarding the way to forecast total bed-nights in Andalusia:

i) Would it be possible to forecast bed-nights more accurately if they were previously disaggregated into bed-nights by Spaniards and by foreigners?

ii) Do hoteliers' expectations contribute to obtaining more accurate forecasts of bed-nights?

The answer to both of these questions constitutes the content of the rest of the present work.

The total monthly bed-nights series (TB), for Spaniards (SB), and foreigners (FB) in Andalusia covers January 1977 to August 1997. Using the program TRAMO[6], we automatically identified the corresponding models for each of these three time series for the period January 1977 to August 1996, the remaining sample period serving to evaluate their forecasting performance.

In these models we have taken into account the three following factors: i) the possible existence of the above mentioned Easter effect; and ii) the presence of *outliers* in the series analysed[7].

For the subperiod September 1996 to August 1997, forecasts with both a one month lookahead period and twelve month lookahead multiperiods were obtained for the values of TB, SB, FB[8]. We tried to analyse whether, on the average, more accurate forecasts are obtained when the total bed-nights series (TB) is directly forecast, or when TB is obtained as the sum of SB and FB forecasts. In addition, it is possible to find out whether the result of the comparison depends on the forecast period or not.

Table 2 shows the RMSE and mean absolute error (MAE) for TB univariate forecasts. The term *aggregated* refers to total bed-nights forecasts and *disaggregated* to those obtained as the sum of bed-night forecasts for Spaniards and foreigners separately.

It can be observed from the results in Table 2 that, on average, *aggregated* forecasts are slightly more accurate than *disaggregated* when predicting with a one month lookahead. On the other hand, in multiperiod forecasts, the RMSE in *disaggregated* forecasts is slightly lower than in *aggregated* forecasts, but the MAE is considerably lower in the latter. So, in this case, there also seems to exist evidence in favour of *aggregated* forecasts.

Thus, as regards the first question raised, the answer is that no evidence exists that it is better to forecast total bed-nights as the sum of the forecasts of Spanish and foreigner bed-nights than to do so directly.

---

[6] TRAMO and SEATS (Beta version: September 1996), Victor Gómez and Agustín Maravall. Banco de España-Servicio de Estudios. Working Document number 9628.

[7] The same model , i.e., ARIMA $(0,1,1)x(0,1,1)_{12,}$ is identified for TB, SB and FB. In TB and SB the Easter Effect was present.

[8] In the one month lookahead period forecasts, the model is fed with observed variable values up to the month before the one to be forecast. For multiperiod forecasts, the last value observed and fed into the model is the month corresponding to the beginning of the forecasting period, in this case August 1996. Forecasts with a one month lookahead would refer to short-term forecasts and follow-ups carried out each month by the relevant services as soon as the information on the previous month became available. 12 month lookahead forecasts would be done at the beginning of the year or a given season and its aim would be to provide a global view of the period as a whole. Therefore, it is reasonable to evaluate the forecasting capacity of the models for both kinds of forecasts.

Table 2
FORECASTING PERFORMANCE OF BED-NIGHTS FORECASTS

	MULTIPERIOD		**ONE PERIOD**	
	AGGREGATED	*DISAGGREGATED*	*AGGREGATED*	*DISAGGREGATED*
RMSE	176,285 (8.44)	174,926 (8.37)	118,082 (5.65)	121,531 (5.82)
MAE	101,194 (4.84)	115,421 (5.53)	84,428 (4.04)	89,624 (4.29)

Note: percentage on average TB is shown in brackets.

On the other hand, as explained previously, two series of hotel managers' expectations are available (EXPECTS) and (EXPECTO), both starting in January 1992, with the most recent data corresponding to July 1997.

Various models have been estimated using EXPECTS or EXPECTO with 1, 2, 3 and 4 lags, trying in this way to determine whether either or both of the two series available could actually be considered as leading indicators of total bed-nights. It is clear that these results are limited by the small size of the available data, given that the origin of the sample period began in January 1992 and ended in August 1996, using the remaining period, till July 1997, for evaluating the forecasting performance of the different models.

TRAMO software was also used in this case, for the automatic identification and estimation of models for total bed-nights with different lags for EXPECTS and EXPECTO. AUTOBOX software, with its automatic identification option of the transfer function, led us to far from robust models, both for Andalusia as a whole as well as for its individual provinces, probably due to the scarcity of sample data. For this reason we decided to continue the analysis with TRAMO, specifying alternative models with various lags for exogenous variables (expectations) and choosing among them those which produced the best fit[9].

The analysis shows that EXPECTS can only be considered as a 3 or 4 month leading indicator for total bed-nights, since in the remaining cases the coefficient of the

[9] When the indicator lag (exogenous variable) is imposed, the program identifies the model for residuals and estimates all the parameters together. The small size of the sample available leads TRAMO to automatically choose a default model which is of the same kind as the one chosen when it carried out an univariate analysis for Total Bed-night series, i.e., ARIMA $(0,1,1)x(0,1,1)_{12}$. The possible presence of an Easter effect and atypical values was also established.

corresponding lag for EXPECTS does not significantly differ from zero. Even in the case of a three-month lag the model suffers from serious technical deficiencies,[10] thus, strictly speaking, we should only consider the four-month lag. Concerning EXPECTO, the conclusions are even more discouraging since in no case is the indicator's coefficient significant at 5%. However, in order not to take a decision based on a single statistical criterion, the results of the models with three and four month lags have been included in the analysis, although their respective *t*-ratios only reach 1.6 and 1.7.

Table 3
FORECASTING PERFORMANCE OF BED-NIGHT MODELS

	MULTIPERIOD				
	UNIVARIATE	SUBJECTIVE EXPECTATIONS		OBJECTIVE EXPECTATIONS	
		3 months	4 months	3 months	4 months
RMSE	176,285 (8.44)	202,495 (9.69)	188,815 (9.04)	191,876 (9.19)	180,734 (8.65)
MAE	101,194 (4.84)	168,378 (8.06)	123,102 (5.89)	121,917 (5.84)	119,862 (5.74)
	ONE PERIOD				
	UNIVARIATE	SUBJECTIVE EXPECTATIONS		OBJECTIVE EXPECTATIONS	
		3 months	4 months	3 months	4 months
RMSE	118,082 (5.65)	200,155 (9.58)	186,925 (8.95)	190,857 (9.14)	183,065 (8.76)
MAE	84,428 (4.04)	166,780 (7.98)	132,106 (6.32)	140,678 (6.74)	136,498 (6.53)

Note: the corresponding percentages regarding the average TB are shown in brackets.

With the aim of evaluating the forecasting performance of the models, for the subperiod September 1996 Ð August 1997, multiperiod forecasts were obtained for a twelve-month lookahead as well as for one-month lookahead forecasts.

---

[10] In more concrete terms, residual show evidence of autocorrelation and the seasonal moving average coefficient is not significantly different from zero.

Table 3 shows the root mean square error (RMSE) and the mean absolute error (MAE) of the univariate TB forecasts and those obtained by considering EXPECTS and EXPECTO as leading indicators for 3 and 4 months.

In the light of the results appearing in Table 3 we can affirm that neither of the two series of available expectations contribute to obtaining, on average, more accurate forecasts of total bed-nights than those obtained via a univariate model.

Thus, bearing in mind all the limitations referred to, the answer to the second question raised is that the expectations, whether subjective or objective, do not constitute valid leading indicators since they do not improve univariate forecasts, and this is so independently of the forecast lookahead being one month or twelve.

Some tests were made using data referring to individual provinces and the conclusions are the same as those referring to Andalusia as a whole.

6. COMPARISON OF FORECASTING METHODS

Once the best models for carrying out 'direct' forecasts of occupancy rate, bed-capacity, and bed-nights have been designed, it is possible to obtain 'indirect' forecasts of these variables via the relationship existing between them.

On the other hand, starting from the direct and indirect forecasts of each variable, a combined forecast can be obtained using the geometric mean.

Table 4
COMPARISON OF FORECASTING METHODS

Variable	Type of forecast	RMSE
	direct	2.717
Occupancy rate	indirect	3.269
	combined	**2.422**
	direct	130,563
Bed-nights	indirect	112,476
	combined	**105,660**
Bed-capacity	**direct**	**3,985**
	indirect	9,312
	combined	5,015

Note: The size of the error depends on the measurement units of each variable

The question we consider in this section is which of the three forecasting methods (direct, indirect, or combined) is the one that works best for each variable. To answer this we have obtained three types of *ex post* forecast (direct, indirect and combined) for each of the variables under analysis (occupancy rate, bed-capacity and bed-nights) during the period January to July 1997. Table 4 shows the RMSE corresponding to each set of forecasts.

From the results shown in Table 4, the best method for forecasting occupancy rate is a combination of direct and indirect forecasts. This is also the case for bed-nights, but for bed-capacity the indirect forecast is so poor that better forecasts can only be obtained by using a direct method. These results are coherent as we already mentioned in Section 3 and it is proved in the Appendix.

7. SUMMARY AND CONCLUSIONS.

Short-term analysis of the tourist sector in Andalusia have to be based on information provided by the statistics of the previously mentioned Survey of Tourist Flow in Hotel Establishments. These statistics give monthly information about three key variables for the different provinces of Andalusia: one relating to demand, (the number of bed-nights), another referring to offer (bed-capacity), and a third relating them together (bed occupancy rate). As the latter is directly related to the performance of the hotel industry, it has special importance for business people in this sector.

Forecasts of bed-nights and bed occupancy rates play a key role in the short-term analysis of this sector, and thus it is important to find the most suitable methods for obtaining the best forecasts possible.

While bearing in mind certain aspects relating to the calendar (the ‘Easter effect’) in order to predict occupancy rate we have to take into account the possibility of using two candidates as leading indicators. First, the objective expectations of the business people (bed-places effectively booked for the next three months), and second, their subjective expectations (occupancy rate expected during the next six months). The information about these variables is made available by the IEA (the Andalusian Institute of Statistics) in charge of dealing with the The Survey of Tourist Flow in Hotel Establishments.

In this work it is shown that the best forecasts of occupancy rate in Andalusia are obtained when using subjective expectations with a time-lag of two months (which in some cases may be three months when the analysis is made for provinces separately). However, this conclusion should be approached with some caution due to the brief period covered by the information available on expectations (from January 1992), which causes a certain lack of robustness in the models.

Currently, however, it is not very useful to use expectations as leading indicators for occupancy rate due to the delay in releasing the data of this survey. The results presented here should serve to spur the IEA into accelerating the process of working with and releasing the statistical information concerning expectations.

Regarding forecasts of bed-nights, the main issue is whether or not it is appropriate to disaggregate bed-nights by Spaniards and by foreigners given that both series follow different time-paths. The greater stability of the aggregated series leads us to conclude that, after having carried out the appropriate tests, the total bed-nights in Andalusia are better predicted when these two series are not disaggregated. On the other hand, neither of the two expectation variables considered improves the results obtained with a simple ARIMA model, which leads us to conclude that the expectations of managers do not function well as leading indicators of demand. In this sense, we must stress the provisional nature of the results until such a time as the size of the temporal series available increases over the coming years.

Once occupancy rate and bed-nights forecasts have been obtained, the two variables of greatest interest in the short-term analysis of this sector, forecasts of bed-capacity are implicitly obtained (the ratio between daily average bed-nights and occupancy rate). Although this variable is not so interesting in itself, we could investigate whether 'indirect' forecasts thus obtained are appropriate or not. The conclusion we have reached in this work is that these bed-capacity indirect forecasts are much worse than those obtained using 'the best direct' model. It turns out that such a model is, as in the case of bed-nights, an ARIMA model which does not include as leading indicators any of the expectation variables.

The remaining problem is to solve the lack of internal coherence between the 'best' forecasts of the three variables studied: occupancy rate, bed-nights and bed-capacity.

One of the main conclusions arrived in the extensive existing bibliography as regards combining forecasts is that, if we combine two forecasts of the same variable obtained via different methods, using for example, the arithmetic mean, the resulting forecast is better than the two original forecasts. It is not really known why this is the case, but there seems to be a compensation of errors that gives more robustness to the combined forecasts.

In the light of this information, we decided to obtain 'combined forecasts' from two types of forecast, 'direct' and 'indirect'. The direct forecasts of each variable are those which have been chosen as the best in the preceding stage; the indirect forecasts of each variable are those obtained from the direct forecasts of the other two variables according to the following relationship: occupancy rate = bed-nights/bed-capacity.

Comparing the direct, indirect, and combined forecasts of each variable led us to the following conclusions. The best forecast for occupancy rate is the combined one; the same result holds for bed-nights, but for bed-capacity the best forecast is obtained with the direct method.

In order to guarantee the internal coherence of the results, that is, that the optimum forecasts fulfil the existing relationship between the three variables, the combined forecasts have to be obtained using the geometric mean of direct and indirect forecasts.

Appendix

Notations:

BN = bed-nights.
OR = occupancy rate.
BP = bed-places.

Subindex:

d = direct forecast.
i = indirect forecast.
c = combined forecast.

By defining the 'combined' forecast as the geometric mean of the direct and indirect forecasts, the relationship between the combined forecasts is

$$\frac{BN_c}{OR_c} = \sqrt{\frac{BN_d \cdot BN_i}{OR_d \cdot OR_i}} = \sqrt{\frac{BN_d \cdot (OR_d \cdot BP_d)}{OR_d \cdot \frac{BN_d}{BP_d}}} = BP_d ,$$

where the second part of the expression is obtained by substituting the direct forecasts of each variable according to the defined relationship.

The result obtained shows that the combined forecasts of bed-nights and occupancy rate implicitly provides a forecast of bed-places which coincides with the forecast yielded by the direct method. This ensures the internal coherence of the forecasting methodology proposed in this work.

REFERENCES

Clement, R. T. (1989). Combining Forecasts: A review and annotated bibliography. *International Journal of Forecasting.* 5, pp. 559–583.

González, A.(1995). Indicador Adelantado del Grado de Ocupación Hotelera. Congreso Nacional de Estadística e Investigación Operativa. Sevilla, 1995.

Gómez, V. and A. Maravall (1996): "TRAMO and SEATS" (Beta version: September 1996). Banco de España. Servicio de Estudios. Documento de Trabajo nº 9628.

González, P. and Moral, P. (1995) "An analysis of international tourism demand in Spain." *International Journal of Forecasting*, 11.233-251.

Otero, J. M. (1996). Principales determinantes del flujo de pasajeros extranjeros desembarcados en el aeropuerto de Málaga. *Estudios de Economía Aplicada*, núm 5, Junio.

Witt, S. F. and Witt, C. A. (1995). Forecasting tourism demand: a review of empirical research. *International Journal of Forecasting*, 11. 447-475.