Real-time forecasting regional tourism with business sentiment surveys

(Article begins on next page)

20 April 2024

Real-time Forecasting Regional Tourism with Business Sentiment Surveys

Guizzardi, Andrea*, Stacchini, Annalisa*

*Department of Statistical Science and Centre for Advanced Studies in Tourism - CAST – University of Bologna. Italy

Highlights

1. We include business sentiment indicators in naïve and STS models.

2. We evaluate forecasting accuracy with descriptive and inferential methods.

3. Business sentiment indicators improve goodness of fit and forecasting accuracy.

4. Extending business sentiment surveys to tourism promises great informative gain.

Abstract

This study provides evidence that supply-side soft information, retrieved from business surveys, is effective in real time forecasting of hotel arrivals at the regional level. We assess the effect of including business sentiment indicators in commonly used naïve specifications and structural time series models, using residuals and predictive diagnostics. We find that both the goodness-of-fit and the forecasting accuracy of the augmented models are superior to those of the baseline models. Whence the opportunity to extend to the tourism sector the surveys on the business sentiment currently realized by the provincial chambers of commerce for the manufacturing sector, allowing an effective and timely managing of local tourism market, where official information is likely to be either lacking or poor in quality.

1. **Introduction and hypothesis**

Indexes based on the opinions and expectations of either consumers or professionals are commonly used to model and anticipate the business cycle, industrial production, national economic turning points and price dynamics (see, among others: Anderson et al, 2011; Gelper et al. 2007; Koskinen & Öller, 2004; Taylor & McNabb, 2007).

Business indicators of production and price dynamics are widely used in modeling and forecasting sectoral conjunctural dynamics (see, among others, Jad, 2011; Pedersen, 2009; Maurin & Paries, 2008), due to the advantages of subjective information over quantitative data. Among others: the provision of reliable signals on the short-term evolution of the industry at the firm level, earlier availability, and less need for corrections and revisions (Darne, 2008, Mitchell et al. 2005). Thus business sentiment indicators (BSIs) are constructed and used by government institutions, chambers of commerce, business associations and institutes of high education throughout the world. Nonetheless, to the best of our knowledge, BSIs have never been employed in the tourism field. The great majority of the literature, concerning modelling and forecasting tourism demand, rely on quantitative data (see the surveys of Song & Li, 2008; Goodwin, 2008 and Goh & Law, 2011), even though a large body of evidence indicates that the use of exogenous 'hard' variables provides no forecasting improvement (Kulendran & Witt, 2001; Song & Witt, 2000, Athanasopoulos et al. 2010). In particular, Song & Witt (2006) underline the difficulties of correctly specifying econometric models, precisely distinguishing between exogenous and endogenous factors. Wong et al. (2006) highlight the problems in forecasting the values of the explanatory variables, not to mention that these variables are often not specifically connected with the tourism dynamics (e.g. the basket of goods consumed by tourists tends to be different from those considered in the consumer price indexes calculation; see Divisekera, 2003). Other studies (e.g. Fildes et al., 2011), although finding some improvement in the forecasting accuracy through the use of exogenous variables, conclude that pure time series approaches are more likely to be appropriate in the majority of cases. It should also be mentioned that, notwithstanding the growing interest in forecasting methods, the literature has not yet considered supply side information, which could affect market dynamics, other than prices and marketing expenditures (Li, et al. 2005).

Swarbrooke & Horner (2001) were among the first to propose to add soft data to the traditional exogenous tourism variables. They stated that business travellers' expectations about a foreign country's economic future and stage of development allow a better explanation of business travel flow. Following these assumptions, some researchers used the opinions of professionals as

business-services consumer sentiment indicators (Njegovan, 2005; Allen & Yap, 2009). However, to the best of our knowledge, no conjunctural information from the tourism supply side has ever been employed, in either forecasting or real time prediction of the tourism business cycle, although it has - at least - four advantages over quantitative measures. First, BSIs, are generally able to synthesize the effects of a wide variety of factors, even not economic in nature, on the dynamics of the accommodation production. Literature highlighted: the security conditions of the destination, its climate and natural environment, its demographic characteristics, social and technological factors and psychological and fashion-related issues (see, among others: Dwyer et al. 2009; Hamilton et al. 2005, Ryan, 2003). Second, BSIs may allow to measure in real time the impacts of unpredictable one-off events, such as: environmental disasters, wars, earthquakes or changes in the productive destination of a region, on the business cycle (see for example: Athanasopoulos et al. 2010; Huang & Min, 2002; Chang et al. 2008), while a more common 'dummy variables approach' (e.g. Claveria & Datzire, 2009), holds only if such events have already been observed, entering the information set. Third, as hotel keepers/managers are (partially) informed about their customers' satisfaction and expectations about local tourism products, BSIs can potentially summarise the impact of territorial policies (e.g. marketing initiatives and/or structural investments). Last but not least, in a sector where official data on tourism demand are provided by hotels keepers/managers, BSIs reflect exactly – in terms of definition - what is measured by the accommodation statistics (see UNWTO, 2008): the sum of the number of arrivals of both tourists and travellers within their usual environment (Govers et al. 2008; Guizzardi & Bernini, 2013). Therefore, hereafter we prefer to use the term 'accommodation production', to which we will also refer as 'tourism production', just for avoiding repetitions.

It should also be stressed that the business sentiment is a subjective, personal assessment of the environment, so the construction of indicators might be affected by strong measurement errors, due to the frequent ambiguity and intrinsic qualitative nature of surveys. Therefore, whether indicators derived from surveys to hotel keepers/managers can be informative about tourism business cycle dynamics is questionable, and investigating it is the aim of the present study.

Thus, the key question posed in this paper is whether and to what extent business sentiment indexes can be informative in modelling and forecasting the tourism production, especially in a small area, where the costs of surveys are low and – more important - official statistics are not able to provide indicators with sufficient territorial detail and timeliness, with respect to the policy makers' needs. If the answer is affirmative, this work will acquire a specific importance for local policy makers, by demonstrating that the subjective evaluations of hotel keepers and managers are

useful to make their decision-making process more effective, allowing a real time knowledge of the market dynamics. To address this issue, we compare and assess the effect of including BSIs in some of the naïve and time series specifications that are usually employed to model or forecast tourism flows. By using residuals and predictive diagnostics, we provide a comparative evaluation of the baseline models versus the "augmented" ones. Forecasts are assessed based on the results of three cost functions for the prediction errors and some inferential methods. Therefore, the present research makes a contribution to the literature by investigating the explanatory and predictive power of BSIs, within the field of tourism.

We focus on the number of arrivals in the hotels of the province of Rimini, a NUTS 3 leading Italian tourist destination, accounting for 3.1% of the national market in terms of sales of overnight accommodations, with 2,130 hotels (the 6.6% of the national market share) and 7% of the national capacity in term of hotel rooms in 2010 (ISTAT, 2011). The BSIs are retrieved from a four-monthly survey, covering the period from April 2000 to December 2012.

The paper is structured as follows. Section 2. presents the business survey, used to obtain BSIs, and the data. The model specifications and estimation results are described in detail in Section 3. Section 4 compares, in both descriptive and inferential terms, the forecasting performances of baseline models with those of the corresponding ones augmented with BSIs. Finally, Section 5 summarises the main findings of the study and draws some conclusions.

2. The survey and the data

2.1 The OSCAR business survey

We collected information on the sentiments of hotels keepers through a survey performed within the context of the "Osservatorio Statistico sulla Congiuntura Alberghiera Regionale" (OSCAR; in English: Statistical Observatory on the Regional Hotels Conjuncture) activities, a project of the University of Bologna. OSCAR has been performing periodic surveys to assess hotel keepers' (or managers') opinions and expectations, through telephonic interviews, since the pilot survey in August 1999. The results have been published each four-months by the local press. The questionnaire is composed of 4 questions on the respondents' opinions regarding:

- the observed level of their own hotels accommodation production (Ob_t);

- their own hotel's expected production for the next period (Ex_t);

- their own hotel's prices dynamic – net of the inflation – (P_t);

- the trend expected for the near future, with respect to the entire provincial market (Mkt_t).

The survey also requests a subjective evaluation of the effectiveness of public spending, which is routinely made available to local politicians.

Opinions are expressed as 'high', 'low' or 'same', with reference to those which the respondents consider the usual dynamics of the sector and their own enterprise. Quantitative variables for the *k*-th (*k=1,2,3,4)* issue are constructed using the net balance statistic, the most commonly used measure for the purpose of nowcasting (see, among others: Matheson et al. 2011; Aylmer & Gill, 2003). In particular, each BSI is calculated as the difference between the weighted proportion of firms reporting an increase and those reporting a decrease, rescaled to make it vary between -100 and 100: $BSI_{m,t} = \left[\%(high_{m,t}) - \%(low_{m,t}) \right] \cdot 100$. Weights are given by the total number of rooms at the surveyed hotels' disposal. Positive balances tend to be associated with growth in the variable of interest and vice versa.

The OSCAR surveys are regularly performed each April, August and December. Since April 2000, the survey has been based on stratified sampling, following two different strategies: one in the summer season and one in the remaining periods. The strata are constructed considering the hotel size (number of rooms), category (number of stars) and location (municipality). Further details on the sampling methodology will be provided upon request. The allotted financial resources were available for 50 (120) interviews performed in the non-summer (summer) surveys, and the sample size remained constant over time. Therefore the coverage rates varied with respect to the number of hotels in the reference population, being approximately 5.6% in the summer season of 2005 and 10.5% in the other 2 periods of the same year. The coverage rate tended to increase, because the overall hotel population followed a slowly decreasing trend over the duration of the project.

2.2 The Data

We chose the number of tourist arrivals as dependent variable (y_t), for which a monthly disaggregated homogeneous time series is available since 1987. Four-monthly data are computed by summing over the monthly observations, to ensure consistency with the frequency of BSIs.

The provisional figures on the tourism production for the year (*t-1*) are released once a year, in the year-end Provincial Statistics Report, published in April of the t-*th* year (see Regional Chamber of Commerce website: RCC, 2014). The same publication reports the ultimate data for

year (*t-2*). Therefore, policy makers and tourism managers obtain reliable information about the *past* accommodation production with 2 years of delay, while effective intervention or business planning, reporting and evaluation would require timely information about the *current* tourism production. The mentioned publication policy drives the way we conduct the "nowcasting" experiment in the ex-post framework, allowing the results of the present work to have direct management implications.

The number of arrivals displays a slight positive trend, with an average annual growth rate of 0.15% over the considered 25 years. This rate rises to 1.6% if the year 1989 is taken as starting point of the calculation, because in that year the mucilage crisis in the Adriatic sea caused a sharp decline in arrivals, and a structural change in the product mix offered by Rimini began, shifting production from the leisure to the business segment. An Augmented Dickey-Fuller (ADF) test, performed on the four-monthly series, shows that it is not stationary, also considering the maximum available time span (1987 – 2012). Conversely, the differentiated series ($y_t - y_{t-3}$) is stationary at any lag length, with a type-I error probability below 1%, suggesting the stochastic nature of the seasonality pattern. Considering the doubts of Hyndman & Khandakar (2008) regarding the biasedness of seasonal unit root tests (leading to over-differencing), we use the ratio-to-moving average multiplicative method (Shareef & McAleer, 2007) to verify the hypothesis of stochastic seasonality. Results are reported in table 1 below.

Table 1. Seasonal variation coefficients in tourist arrivals: 3-year average figures.

	'87-'90	'91-'93	'94-'96	'97-'99	2000-2002	'03-'05	'06-'08	'09-'12
Jan-Apr	40%	20%	28%	40%	47%	48%	50%	45%
May-Aug	213%	163%	215%	215%	202%	201%	197%	204%
Sep-Dec	47%	49%	48%	46%	51%	51%	53%	51%

These indexes show peaks in the summer months, when, in 6 out of 8 periods, more than the double of the annual average number of tourists is recorded. Over the same four-month period, the seasonal coefficients display a high variability and no trend. Moreover, the seasonal coefficients maintain the same rank across different periods, but their relative distance is highly variable, confirming the presence of a stochastic seasonality.

Descriptive statistics for BSIs are reported in table 2, showing the operators' essentially pessimistic attitude in reporting the industry performance: despite the long term market growth, the balance between optimists and pessimists is always negative, with the exception of the assessments

of the prices' dynamic. It is worth noting that the average level of BSIs tends to stabilize at lower values since 2008, in accordance with the strong worsening of the world economic crisis.

Table 2. BSIs' descriptive statistics. Average figures (years: 2000 to 2012).

BSI:	Ob_t	Ex_t	P_t	Mkt_t
Overall				
Mean (%)	-22	-17	28	-16
Variance	358	328	216	618
January-April				
Mean (%)	-21	-14	19	-9
Variance	377	311	152	799
May-August				
Mean (%)	-21	-14	41	-18
Variance	446	578	154	592
September-December				
Mean (%)	-23	-24	24	-21
Variance	311	92	102	482

The balance is even worse when the operators are asked about the performance of their own hotel, suggesting an understatement of their own business volume - probably driven by tax considerations - which can be partially corrected considering the annual differences of BSIs (also proved to be stationary) as explanatory variables.

In any four-month period, the market trend indicator displays the highest variance, likely due to both a higher frequency of extreme evaluations and a lower propensity to report extreme negative performances. This evidence strengthens the indication that the evaluations of hotel keepers/managers are less affected by exogenous factors (e.g. taxation) when referring to the dynamics of the entire market, than when referring to that of their own hotel.

Each BSI displays an ADF statistics that does not permit the hypothesis of the presence of a unit root to be rejected at the usual 5% significance level. The series, differentiated at the seasonal frequency, computed as ($BSI_t - BSI_{t-3}$), are stationary at any lag length, with a type-I error probability lower than 1%. However, the main reason to consider the annual variations in the BSIs as explanatory variables is to explore their usefulness in modelling short-term shocks (including one-off events), which are presumed to be known by the interviewed operators.

The plots of annual differences in arrivals against those in each BSI (see figure 1) show that the two series co-vary, as expected. All regression lines (straight lines in scatter plots) are positively oriented, except for P_t. Focusing on extreme values, only some of the main changes in tourism professionals' perception coincide with the highest percentage variations in arrivals. However it is to be considered that Ex_t and Mkt_t reflect expectations on the future production (they should be lagged), while Ob_t synthesizes the opinions of a sample of hotel keepers/managers on their own production, while data on arrivals refer to the whole population in the province. In the considered period, no sudden decrease in four-monthly accommodation production, eventually related to specific one-off events, is observed.

The final dataset is composed by 39 four-monthly observations, spanning from April 2000 to December 2012. For each period, the dataset includes the number of tourist arrivals (y_t), and the four BSIs previously described.
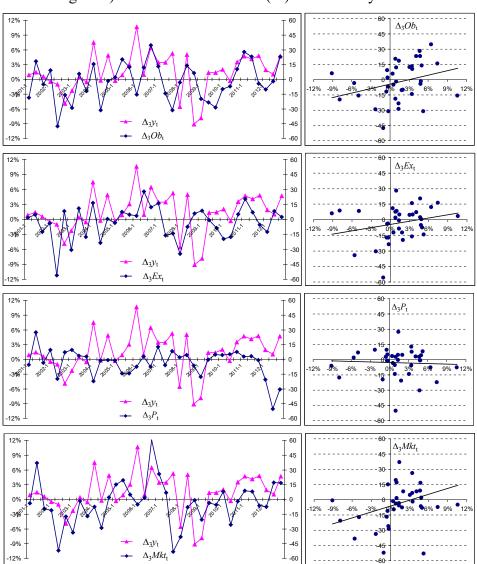
Figure 1) Differences in arrivals (%) and BSIs: dynamics and scatter

3. Methodology

To assess whether and to what extent BSIs may be useful to overcome delays in the publication of official data, about the realized accommodation production, (real time forecasting) and for supporting the regional policy decision-making process, we compare different baseline specifications with the corresponding models augmented with BSIs. We focus on naïve and time series models, both known in the tourism literature for their superior forecasting accuracy in comparison with models including socio-economic exogenous variables (Athanasopoulos, 2010). The latter are also excluded from the present analysis, as official measures of important production drivers are not timely provided - at the provincial (NUTS 3) level - with a four-monthly frequency, disaggregated at the hotel industry level.

The empirical evidence regarding the complexity of the optimal time series approach is mixed. On one hand, there is general agreement that naïve models exhibit superior forecasting performance compared to complex ARIMA models (Li et al, 2005; Song & Witt, 2000). On the other hand, more recent literature displays good forecasting performances of more complex specifications, such as Vector Autoregressive (VAR) and Structural Time Series (STS) models (Li, et al. 2005; Fildes et al. 2011; Athanasopoulos, 2010). In their assessments of forecasting performances, Kulendran & Witt (2003) and Turner & Witt (2001) also conclude that the simplest form of the STS models produces good short-term forecasts. They also argue that the inclusion of economic variables, on the right hand side of STS models, does not appear to generate any improvement. Our argument in this study goes a step further, as we consider both naïve and more complex STS models, for evaluating the explicative power of BSIs.

3.1. The baseline naïve models

We select the following naïve time series models, in which trend and seasonality are modelled in the most often recurring forms in literature:

$$\Delta_3 \ln(y_t) = c + \beta_1(\Delta_3 \ln(y))_{t-1} + \varepsilon_t \qquad (1)$$

$$\Delta_3 \ln(y_t) = c + \varepsilon_t \qquad (2)$$

$$\Delta_3 \ln(y_t) = \beta_1(\Delta_3 \ln(y))_{t-1} + \varepsilon_t \qquad (3)$$

$$\ln(y_t) = c + \beta_1 \ln(y_{t-3}) + \varepsilon_t \qquad (4)$$

$$\ln(y_t) = c + \varepsilon_t \qquad (5)$$

$$\ln(y_t) = \beta_1 \ln(y_{t-3}) + \varepsilon_t \qquad\qquad (6)$$

$$\ln(y_t) = \sum_{j=1}^{3} \beta_j D_j + \beta_4 trend + \beta_5 \ln(y_{t-1}) + \varepsilon_t \qquad\qquad (7)$$

where y_t denotes the number of tourist arrivals, Δ_3 is the seasonal difference operator, D_j is a seasonal dummy variable (for four-monthly data), and *trend* is a linear trend component. Models (1) and (4) are auto-regressive specifications; they are also estimated by constraining to zero the beta parameters - equations (2) and (5) – and the intercepts - equations (3) and (6). Model (7) is an AR(1) specification, with a linear trend and seasonal dummy variables. The previous seven naïve specifications are chosen because of their frequent use as effective benchmarks in the literature (see, among others, Athanasopoulos et al, 2010; Darne, 2008; Song et al. 2003b). Parameters are estimated through Ordinary Least Squares (OLS) on a (sub)sample, running from 2000 to 2009 (26 to 30 observations, loosing up to 4 observations for differencing and delays). We refer to this sample as S_1. The 9 most recent data, from April 2010 to December 2012 (hereafter S_2), are removed from the estimation sample for subsequent evaluation of the models' forecasting accuracy. The full pattern of estimations and residuals statistics is available upon request from the authors.

In summary, all of the models appear to be substantially mis-specified, as the residuals tests do not allow to assume they to be normally distributed nor serially uncorrelated. Moreover, the coefficients are often not significant (even considering a 20% p-value threshold). We evaluate each model based on its Mean Absolute Percentage Error (MAPE) on a (sub)sample (hereafter $S_{1,1}$), containing six four-month observations, from April 2008 to December 2009. This choice stems from the hypothesis that ex-post forecasting accuracy (evaluated on S_2 in section 4) is better disclosed by the MAPE value on the most recent observations (not included in S_2).

The best performing model is the random walk with drift in equation 2 (hereafter Naïve_1), an expected result, given that the choice of seasonally differenced data has often been suggested as "optimal" among naïve or automatic time series predictors in the tourism field (Athanasopulos, 2010). By representing the average annual growth rate in each four-month, Naïve_1 reflects a market characterised by three independent stochastic processes, each one associated with a specific season (or product). Indeed, since the environmental crisis in the late '80s (mucilage in the Adriatic sea), the Rimini province has become a multi-supply destination, offering at least 3 macro-products: leisure (spring thermal and sport tourism), seaside, winter business and congress tourism, each with its own demand segment, supply typology and promotion strategy.

The estimation output of Naïve_1 model is shown in table 3, in the following paragraph, to allow a comparison with its augmented version, including BSIs.

3.2 The Augmented naïve model

We model the annual percentage variation of visitors arrivals in the Rimini province as a linear function of BSIs, both coincident and lagged. The augmented naïve model takes the following general form:

$$\Delta_3 \ln(y_t) = c + \sum_{k=1}^{K} \sum_{j=1}^{J} \beta_{k,j} \Delta_3 X_{k,t-j} + \varepsilon_t \qquad (8)$$

$X_{k,t}$ is the *k-th* BSI (*K=4*) and *J=3* the maximum lag. So we introduced BSIs on the right hand side of specification (2). The choice of the annual percentage change as dependent variable allows to perform a scale reduction, desirable because of the large absolute values of the observed arrivals and their scale difference compared to the explanatory variables. The latter are in fact annual differences in the judgements of the tourism operators, originally expressed in a range from -100 to 100.

For the model selection, in the absence of a theory to guide us in detecting the correct lag structure, and considering that the main goal of the model is to provide accurate forecasts, we follow a two-step specification search procedure. The first step takes an exploratory role. We estimate model (8) on sample S_1 (years 2000 – 2009) following a general-to-specific pattern. To avoid the exclusion (based on purely statistical rules) of relevant variables and delays, BSIs are excluded/included based on two criteria: (a) a fixed significance level of 20%, as threshold for dropping regressors, along with well-behaved residuals (in terms of serial non-correlation, homoskedasticity and normality), for assessing the correct specification of the model; (b) the meaningfulness of the coefficients' signs (positive coefficients for all of the BSIs, except for the price dynamics). Starting from the estimation of the 'full model' (including all BSIs at all lags), we check the restrictions to zero on the beta parameters one at a time, following two parallel selection procedures.

In the first one, we eliminate the parameter with the lowest *t*-statistic, following criterion (a), and re-estimate the model until all of the parameters have a significance level below the fixed threshold. Then, following criterion (b), we eliminate the parameter having the 'wrong' sign. The

resulting model is named $base_{1,0}$. In the second procedure, criterion (b) prevails, so we iteratively

eliminate the parameter with the highest *t*-statistic among those having the wrong sign. The

resulting model is named $base_{2,0}$. The two $base_{i,0}$ *(i=1,2)* models are then augmented by

introducing the excluded variables one at a time. Whenever a model satisfying criteria (a) and (b) -

with an Akaike Information Criterion (AIC) lower than those of the previous models - is identified,

it is recorded as $base_{i,j}$ model *(j=1,2,…)*. The order of inclusion of the variables is random, and the

experiment is iterated three times. This procedure is designed to partially compensate the problem

of few degrees of freedom (df) for the estimation of the initial 'full' model. In fact, the risk to

exclude significant sentiment indicators (based on biased t-statistics), is kept under control by

reintroducing each removed variable. In this 'forward phase', no model ever includes more than 4

regressors, so that df are always sufficient.

Afterward, we evaluate the forecasting performance of each $base_{i,j}$ model, based on its

MAPE on sample $S_{1,1}$, which was also included in the estimation sample, to maximize the degrees

of freedom. In this phase we do not use data for years 2010-2012, thus models augmented with

BSIs are not advantaged in the subsequent forecasting competition. No competing model is

favoured over the others (because all have been estimated on the same sample) and redundancy of

evaluation measures is avoided, as the MAPE loss function is different from that upon which AIC

relies. The final augmented specification is that with the minimum MAPE. Three BSIs are selected

as explanatory variables: $(\Delta_3 Ob)_t$, $(\Delta_3 Ex)_{t-2}$ and $(\Delta_3 Mkt)_{t-3}$. The BSI of the price trend was

excluded, because it never met criterion (a), although it always maintained the expected negative

sign. This result is unexpected: literature brings substantial evidence for the key role of prices in

determining the tourism market dynamics (Song & Li, 2007; Han et al. 2006; among others). We

surmise that the lack of statistical significance of this BSI may be due to a measurement problem:

respondents are asked to evaluate the prices dynamic already adjusted for inflation, which is likely

to be heterogeneously perceived among the interviewed tourism professionals. Furthermore, their

answers may be conditioned by diffidence in reporting strategic competitive information, such as

the prices dynamic of their own business.

Therefore, the functional form of the final model (hereafter Naïve_A) can be expressed as

follows:

$$\Delta_3 \ln(y_t) = \alpha_0 + \beta_1 (\Delta_3 Ex)_{t-2} + \beta_2 (\Delta_3 Mkt)_{t-3} + \beta_3 (\Delta_3 Ob)_t + \varepsilon_t \qquad (9)$$

We estimate equation (9) using OLS. Table 3 summarises the estimation output, allowing for a direct comparison with the Naïve_1 model. The juxtaposition demonstrates that the use of the selected BSIs produces an appreciable improvement in the goodness-of-fit, as indicated by AIC values. Moreover, the residuals of the Naïve_A model are normally distributed, homoskedastic and serially uncorrelated.

Table 3. Parameter estimates for the Naïve models:
Baseline (Naïve_1) and Augmented (Naïve_A)

Parameter	Naïve_A (24 Obs.)		Naïve_1 (24 Obs.)	
	Coeff.	*S.E.*	*Coeff.*	*S.E.*
Intercept	0.01951***	0.00844	0.00920	0.00962
$(\Delta_3 Ob)_t$	0.00055*	0.00042		
$(\Delta_3 Ex)_{t-2}$	0.00112***	0.00045		
$(\Delta_3 Mkt)_{t-3}$	0.00061**	0.00034		

Diagnostic

Adjusted R-squared.	0.340	0.000
Jarque Bera	0.071	0.271
ARCH LM (3)	1,753	1,605
Box&Pierce Q(20)	9.159	9.977
Breusch-Godfrey (20)	24.000	22.963
AIC	-152.998	-145.676
BIC	-76.999	-73.211

*** indicates a significance level lower than 0.05. ** indicates a significance level between 0.1 and 0.05. * indicates a significance level between 0.2 and 0.1.

All of the BSIs' coefficients are significant, although the type I error probability, associated with the current observed production $(\Delta_3 Ob)_t$ is nearly 20%. The exclusion of this variable would produce a slight increase in the AIC value. The decision to keep it is also justified by the bootstrap of standard errors and confidence intervals. Results (available upon request), confirm that the estimated coefficients are correct and significant.

Turning to parameters analysis, the higher significance of the lagged variations indicates that hotel managers/keepers are able to anticipate the future output through their opinions on both the overall market cycle-trend (*Mkt)* and their own business' expected dynamic (*Ext*). Not surprisingly,

their proficiency in anticipating the dynamic of the accommodation market production performs at best at one-year horizon. The current observed variation in tourist overnight stays, $(\Delta_3 Ob)_t$, is the least significant variable, confirming - together with the low significance of the price variable - that respondents are more accurate in evaluating future dynamics than their current levels of production or price. Contrary to what is stated in Lui (2011a) - and usually expected - here BSIs seem more informative for forecasting than for nowcasting. We owe this "low explanatory power puzzle" to the diffidence of tourism operators in disclosing explicit information about the current state of their business, which hotels keepers may fear will contrast with the figures that they will officially report for tax purposes (Guizzardi & Bernini, 2013). Although a detailed exploration of this issue is beyond the scope of the present paper, the evidence appears sufficiently strong to encourage the introduction of a new question - related to the evaluation of the provincial dynamic of current production - in the questionnaire, to allay the respondents' concern in expressing their judgements.

3.3 The baseline STS Models

In specifying the baseline STS model, we consider the accommodation production to be composed by a stochastic trend, a seasonal term, an irregular component and a cyclical one. The latter follows a given trigonometric specification, but the amplitude of the cycle is allowed to vary over time. The STS model is specified as:

$$\ln(y_t) = \mu_t + \phi_t + \psi_t + \varepsilon_t \qquad (10)$$

where y_t is the tourism production, μ_t is the local linear trend component, φ_t is the seasonal component, ψ_t represents the cycle and ε_t is the irregular component (stochastic error). This specification is the same used in previous studies (e.g. Vu & Turner, 2006; Guizzardi & Mazzocchi, 2010). The local linear trend is assumed to be a random walk with a stochastic random drift (δ_t):

$$\begin{cases} \mu_t = \mu_{t-1} + \delta_{t-1} + \upsilon_t \\ \delta_t = \delta_{t-1} + \zeta_t \end{cases} \qquad (11)$$

where υ_t and ζ_t are mutually uncorrelated white-noise Gaussian errors. The seasonal component is assumed to follow a stochastic dummy specification:

$$\phi_t = \sum_{j=1}^{s-1} \phi_{t-j} + \omega_t \tag{12}$$

where s depends on the data periodicity (for 4-monthly data, $s=3$) and ω_t is also a white-noise Gaussian error. Finally, the cyclical component is defined as in the trigonometric specification described in Harvey & Jaeger (1993):

$$\begin{cases} \psi_t = \rho \cos \lambda_c \psi_{t-1} + \rho \sin \lambda_c \psi_{t-1}^* + \kappa_t \\ \psi_t^* = -\rho \sin \lambda_c \psi_{t-1}^* + \rho \cos \lambda_c \psi_{t-1}^* + \kappa_{t-1}^* \end{cases} \tag{13}$$

where ρ is a damping factor between 0 and 1, λ_c is the frequency of the cycle (in radians) and the errors κ_t and κ_t^* are mutually uncorrelated white-noise Gaussian errors.

Maximum likelihood estimation of (10) is achieved by rewriting the model in the state-space form and applying the BHHH optimisation procedure of Berndt et al. (1974) (see Harvey, 1989 for technical details). The damping factor and the frequency of the cycle are the parameters to be estimated.

3.3.1 Estimation

We estimate two different STS models, both with the logarithm of the number of arrivals as dependent variable, on two different estimation samples. The first model is estimated using the longest homogeneous time series available (1987:1 – 2009:3); we name it STS_W. The second model (STS_C) is estimated on a left-censored interval, in order to keep the same sample available for the Naïve_A specification.

Both STS models are estimated using the STAMP software, for various starting values of the parameters, ρ (0, 0.1, 0.2, …, 0.9 and 0.95) and λ_c (6, 9, 14, 16, 18 and 21). The models appear to be robust to changes in the initial values, but return a cycle component that is significantly different from zero only for STS_W.

The adequacy of the relationship portrayed by the STS models can be assessed by examining the sample residuals diagnostics, including the usual goodness-of-fit statistics for STS models (R^2_D and R^2_S) and two information criteria (AIC and Bayesian Information Criteria, or BIC), which enable to compare the STS baseline models with the augmented ones. The R^2_D and R^2_S values compare the goodness-of-fit of the equations (Harvey, 1989, p. 268), to that of a simple random

15

walk plus drift (RWD) and a RWD model in first differences about the seasonal means, respectively. Finally, we report a test for serial correlation in the residuals, based on the Ljung-Box Q-statistic and a normality test, based on the Doornik & Hansen (1994) omnibus normality test. Models' estimation outputs are reported in table 4.

Table 4. Parameter estimates for the Structural Time Series (STS) models
('_C' stands for 'Cut sample', '_W' stands for 'Wide sample')

		Final state (third four month of 2009) Dependent variable: ln(arrivals)			
		STS_W (69 Obs.)		STS_C (24 Obs.)	
Parameter		*Coeff.*	*R.M.S.E.*	*Coeff.*	*R.M.S.E.*
μ	Intercept	13.528***	0.0251	13.501***	0.0161
δ	Drift	0.0084***	0.0025	0.00447	0.0042
ψ	Cycle 1	-0.0345***	0.0108		
$\psi*$	Cycle 2	-0.0234***	0.0109		
ϕ_1	Season 1	-0.4305***	0.0261	-0.4364***	0.016
ϕ_2	Season 2	0.9765***	0.0209	0.9768***	0.011
Seasonal χ^2 test		2610***		11370***	
Cycle Analysis					
ρ	Damping factor	1			
λc	Period (years)	6.378			
Cycle χ^2 test		15.84***			
Diagnostic					
R^2_D	Goodness-of-fit 1	0.9971		0.9986	
R^2_S	Goodness-of-fit 2	0.7925		0.4444	
DW	Durbin-Watson	1.988		2.102	
Q(12; 6)	Ljung-Box	13.48		2.857	
Normality	Doornik-Hansen	14***		1.58	
BIC		-4.759		-5.3705	
AIC		-5.0828		-5.714	

*** indicates a significance level lower than 0.05. ** indicates a significance level between 0.1 and 0.05. * indicates a significance level between 0.2 and 0.1.

Both models display a goodness-of-fit superior to that of the naïve benchmark models. The values of R^2_D and R^2_S are remarkably higher than those found in the tourism literature about STS

(compare: Kim & Moosa, 2005; Greenridge, 2001; Gonzales & Moral, 1996; Guizzardi & Mazzocchi, 2010), especially in the STS_W case. To allow a direct comparison with STS_W, we also estimated STS_C model with a cycle component. However it was never significant and worsened diagnostic (R^2_S=0.4422, BIC=-4.988, AIC=-5.479, in the best case), confirming the difficulty in estimating a cycle for such a short observation period.

The STS_W model displays a damping factor, ρ, equal to 1 and a cycle variance close to 0, hinting at a stationary deterministic cycle over the time span of the analysis. The model residuals are also well-behaved, with no evidence of serial correlation. The non-normal distribution of the residuals in the STS_W model may point at the presence of short-term shocks in the tourism production, misplacing the symmetry of the distribution (e.g. the mucilage crisis in the late 80s). Although dropping the normality assumption it is not guaranteed that the Kalman filter will provide exactly the conditional mean of the state vector, it is worth noting that it is still the best estimator within the class of all linear estimators, as it minimises the mean square error (Harvey, 1989).

3.4 The augmented STS model

We augmented the baseline STS_C model using those BSIs selected through the procedure described in section 3.2, to test their informative power within the STS functional form. The augmented specification (STS_A) is estimated based on sample S_1 (i.e. years 2000 to 2009) while sample S_2 (i.e. years 2010 to 2012), is kept out, to assess the forecasting performances. The results are reported in table 5 below, in comparison to the baseline STS_C model.

As in the Naïve_A case, the most significant indicators are lagged BSIs. However, the current observed variation in tourist overnight stays $(\Delta_3 Ob)_t$, here is more significant and improves the AIC value when included. The goodness-of-fit is improved by augmenting the STS_C specification with the selected set of explanatory variables, either basing the comparison on the R^2_D and R^2_S values or on information criteria. Moreover, the residuals display a better behaviour in the augmented model. Both these results lead to the conclusion that BSIs can be usefully exploited to explain the deviations from the trend-cycle and seasonality, due to short-term shocks.

Both the significance of the coefficients and the goodness-of-fit of STS_A model are superior to those of the Naive_A specification. The naïve approach seems therefore less suitable to model visitors arrivals dynamics than STS, when supply side soft information is available. As in Kauppi et al. (1996), this confirms that BSIs, as short-term indicators, can be more usefully exploited to explain the deviations from a modelled long term dynamic.

Table 5. Parameter estimates for the Structural Time Series models: Baseline (STS_C) and augmented (STS_A).

		Final state (third four month of 2009) Dependent variable: ln(arrivals)			
		STS_A (24 Obs.)		**STS_C (24 Obs.)**	
Parameter		*Coeffi*	*R.M.S. E.*	*Coeff*	*R.M.S. E.*
μ	Intercept	13.520***	0.0135	13.501***	0.0161
δ	Drift	0.0039	0.0032	0.00447	0.0042
ϕ_1	Season 1	-0.4469***	0.0105	-0.4364***	0.016
ϕ_2	Season 2	0.9637***	0.0075	0.9768***	0.011
β_1	$(\Delta_3 Ob)_t$	0.000477**	0.00025		
β_2	$(\Delta_3 Ex)_{t-2}$	0.000727***	0.00022		
β_3	$(\Delta_3 Mkt)_{t-3}$	0.000657***	0.00019		
Seasonal χ^2 test		46825***		11370***	

	Diagnostic		
R^2_D	Goodness-of-fit 1	0.9994	0.9986
R^2_S	Goodness-of-fit 2	0.7646	0.4444
DW	Durbin-Watson	2.077	2.102
Q(9; 6)	Ljung-Box	15	2.857
Normality	Doornik-Hansen	0.317	1.58
BIC		-5.832	-5.3705
AIC	.	-6.323	-5.714

*** indicates a significance level lower than 0.05. ** indicates a significance level between 0.1 and 0.05. * indicates a significance level between 0.2 and 0.1.

4. Evaluation of forecasting performances

While a good within-sample performance may be reassuring regarding the explanatory power of the BSIs, out-of-sample forecasting is a key evaluation criterion for the goal of the present paper. If a specification, augmented by BSIs, yields more accurate forecasts than the baseline model, then we have stronger evidence of the informative power of (cheap and easy to run) supply-side surveys in the field of tourism.

The issue of the relative forecasting performance of alternative specifications has been widely studied in the tourism-related literature. In their extensive review, Li et al. (2005) mention 23 studies targeted to forecasting accuracy comparisons (see also, Song & Witt, 2006; Goh & Law, 2011 and references therein). Although the comparisons are usually limited to complex models, there is empirical evidence indicating that basic time series models, or even more naïve predictive specifications, may often be preferred (see Kulendran & Witt, 2001 and 2003). There is therefore a pressing need to consider these simple models in assessing the trade-off between complexity and predictive power.

In almost all published work in the field of tourism, the forecasting performances are assessed using loss functions based on the ex-post predictions of alternative specifications. We consider the mean absolute percentage error (MAPE), the mean absolute error (MAE) and the mean squared error (MSE), the three cost functions most frequently appearing in tourism studies (see Li, et al. 2005). Alternative error measures include: the root mean square percentage error, Theil U statistic (see, among others, Goh & Law 2002; Cho, 2003) and mean and median absolute scaled errors (Athanasopoulos et al. 2010). However, the MAPE, MAE and MSE are sufficient to describe the magnitudes of the forecasting errors and, through comparisons, to signal whether one model outperforms another in forecasting extreme values. The MSE depends on the magnitude of the forecast variables and weights more than proportionally the largest forecasting errors (Clark & McCracken, 2011), while MAPE does not.

In addition, we employ inferential assessment methods. We first test the prediction biasedness, through an OLS regression of the model (M) prediction errors $(e_{M,h} = y_h - y_{M,h} \quad (h = 1,2,...,H))$ on a constant term. If the estimated regression coefficient significantly differs from 0, then the null hypothesis that the predictions are unbiased is rejected.

Moreover, we perform three tests to compare the relative forecasting performance of the (rival) baseline and augmented models. The first is an exact finite-sample non-parametric test,

equivalent to the Wilcoxon signed-ranks test for paired samples, implemented on the differences between the absolute errors of two rival models (M1 and M2):

$$TW = \sum_{h=1}^{H} I_- \cdot rank(|\delta_h|) \quad \text{where} \quad \delta_h = |e_{M1,h}| - |e_{M2,h}| \quad (14)$$

where I_- is an indicator function, equal to 1 when $\delta_h < 0$. The test statistics is the sum of ranks associated with negative differences. The null hypothesis being tested is that this sum equals the sum of ranks of the positive differences, in which case the forecasting performance of the two models being compared is similar.

If the one-step-ahead forecasting errors have zero mean and are normally distributed, it is possible to perform more powerful parametric tests. We first test whether the differences between the MSEs of the rival forecasts can be attributed to sampling variability, through the Harvey et al. (1997) version of the (Morgan) Granger and Newbold test (Granger and Newbold, 1977), which remains consistent in presence of non-normally distributed heteroskedastic forecasting errors. This Harvey Leybourne and Newbold (HLN) test takes the form:

$$e_{+,t} = \beta e_{-,t} + \varepsilon_t \quad \text{where} \quad e_{+,t} = e_{M1,t} + e_{M2,t} \quad e_{-,t} = e_{M1t} - e_{M2t} \quad (15)$$

Testing whether $\beta = 0$ allows to verify the null hypothesis of equivalent forecasting performances of the two predictors being considered. The test statistic is:

$$S_2^* = \frac{\hat{\beta}}{\sqrt{\dfrac{\sum e_{-,t}^2 \hat{\varepsilon}_t^2}{\left(\sum e_{-,t}^2\right)^2}}} \quad \text{where} \quad \hat{\beta} = \frac{\sum e_{-,t} e_{+,t}}{\sum e_{-,t}^2}$$

which is distributed approximately as a Student'T, with (n-1) df.

A second test is developed from the most general formulation used for encompassing tests (Clements & Harvey, 2010):

$$y_t = \alpha_0 + \alpha_1 y_{M1,t} + \alpha_2 y_{M2,t} + \varepsilon_t \quad (16)$$

where $y_{Mj,t}$ denotes the one-step-ahead forecast for y_t obtained with the *j*-th model (*j*=1, 2). Instead of testing for $\alpha_2 = 0$ (to assess whether the first forecast incorporates all the useful predictive information contained in the second one), we test for $\alpha_1 = \alpha_2$, to determine whether both forecasts contain an equal amount of predictive information. This way, provided that the rival models are nested, we are able to test whether the inclusion of a set of variables, in a baseline specification, is useful for prediction. To reduce multi-collinearity in eq.12, we apply the usual orthogonalizing transformation, estimating the following model:

$$y_t = \alpha_0 + \beta_1 y_{-,t} + \beta_2 y_{+,t} + \varepsilon_t \quad \text{where} \quad y_{-,t} = y_{M1,t} - y_{M2,t} \quad \text{and} \quad y_{+,t} = y_{M1,t} + y_{M2,t} \quad (17)$$

By construction, $\beta_1 = 0.5(\alpha_1 - \alpha_2)$, and $\beta_2 = 0.5(\alpha_1 + \alpha_2)$; therefore, if the null hypothesis $\beta_1 = 0$ is rejected, then $\alpha_1 \neq \alpha_2$. Moreover, the inclusion of additional information is useful for prediction if $\beta_1 > 0$, provided that M1 is the model with the highest number of variables.

Thus, we test the forecasting performances of the considered models on an *H=9* four-month horizon (3 years), from January 2010 to December 2012 (S_2). Many previous studies, focusing on quarterly data, consider an out of sample width up to 2 years, namely: 6 (Song et al. 2013) or 8 (Athanasopoulos et al. 2010; Song et al. 2011) periods. For yearly data Li et al. (2006) and Athanasopoulos et al. (2010) consider a 4 year horizon. No extant work deals with four-monthly observations. As this frequency is intermediate between those previously quoted, a forecast sample of 3 years is here considered to be consistent with literature.

We apply a yearly rolling scheme, estimating the models three times, as required by the data publication practices of the Rimini Province. Our one-year-ahead four-monthly real time forecasts, based on the predictions of the surveyed operators in April, August and December, allow for a 12 month maximum anticipation period. We found no systematic nor significant difference in accuracy between forecasts obtained considering an expanding window and those obtained by setting a fixed estimation window, so we simplify the presentation reporting only results obtained using the second estimation methodology (see table 6).

Table 6. Forecasting errors and descriptive measures of forecasting accuracy
('_A' stands for 'Augmented', '_C' stands for 'Cut sample', '_W' stands for 'Wide sample')

	RW	Naive1	NaiveA	STSW	STSC	STSA
1st four month 2010	-2,0%	-1,1%	-1,5%	-1,6%	-0,1%	2,2%
2nd four month 2010	0,3%	1,2%	1,1%	1,1%	1,9%	1,2%
3rd four month 2010	-3,5%	-2,6%	-3,2%	-1,7%	-2,2%	-2,8%
1st four month 2011	-4,6%	-3,0%	-2,7%	-0,5%	-3,8%	-1,1%
2nd four month 2011	-4,0%	-2,5%	-0,7%	0,7%	-2,8%	-1,3%
3rd four month 2011	-4,7%	-3,1%	0,6%	0,3%	-3,8%	-0,9%
1st four month 2012	-1,9%	0,1%	1,1%	4,0%	0,5%	0,7%
2nd four month 2012	-1,0%	0,9%	0,8%	4,3%	1,3%	0,7%
3rd four month 2012	-4,6%	-2,7%	-2,9%	0,1%	-2,9%	-3,6%
MAPE	2.94%	1.91%	1.62%	1.59%	2.15%	1.60%
MAE	23,297	16,999	12,214	17,765	20,407	13,035
MSE (millions)	998	461	184	977	682	232
MaxAPE	4.7%	3.1%	3.2%	4.3%	3.8%	3.6%

The addition of the BSIs improves the forecasting accuracy of both STS and naïve specifications. Based on these accuracy measures, Naive_A is the optimal predictor, followed by STS_A, which slightly outperforms Naïve_A, only in terms of MAPE. The substantial predictive information content of BSIs is also demonstrated by the fact that STS_A is the augmented version of STS_C, the weakest predictor after RW.

Among the baseline specifications, Naïve_1 outperforms the more complex time-series models, except when the database is large enough to permit modelling the business cycle; in this case STS_W becomes the best predictor, but only in terms of MAPE.

Focusing on the latter accuracy measures, all the baseline specifications are also markedly inferior to both models with BSIs. Given the strong seasonality of the phenomenon under consideration, the ranking, in terms of different loss functions, demonstrates that BSIs are highly informative, especially about the production pattern in the 'peak season' (the most relevant season for policy makers, when the tourist flow is maximum). This result can be interpreted as consistent with the more general thesis of Lui et al. (2011b), who stated that professionals' predictions are more accurate than benchmark (auto-regressive) models in periods characterized by high volatility. Another result of interest is the poor performance of the no-change model, which outperforms more complex time series or regression models in many extant studies (see, among others: Kulendran & Witt, 2001; Song et al, 2003).

Finally, it is worthy of note that the inclusion of a cycle component (though not significant) worsens the out of sample performance of STS_C, while its augmented version (STS_A) displays a similar forecasting accuracy. This evidence strengthens the conclusion that BSIs allow to capture

the effects of an existing cyclical dynamic, even if the lack of empirical evidence does not allow to correctly parameterize it. Table 7 presents selected results from the inferential comparison. The bias test does not allow to consider all the forecasts unbiased. Bias is particularly severe for the no change model, while the errors of Naïve_1 are unbiased, thanks to a constant correction. The introduction of BSIs always lowers the error bias.

The HLN test demonstrates that both BSIs-augmented models are always more accurate than the baseline STS and no change specifications, at significance level below 5%. The same results apply in case of Naïve_1, with a higher p-value. No significant difference in accuracy is found between the two augmented models, though the positive signs of the beta coefficients confirm that Naive_A is slightly more accurate (in terms of squared loss). Thus, the choice of the functional form of the augmented model does not affect its forecasting accuracy.

Consistently, the test of equal predictive information, based on regression (17), shows that the inclusion of BSIs significantly improves the forecasting performance, regardless of the complexity of the adopted specification. However, when the estimation sample is large enough to allow modelling the business cycle, the differences in forecasting accuracy tend to reduce.

Table 7. Inferential comparisons on the biasedness and forecasting accuracy
('_A' stands for 'Augmented', '_C' stands for 'Cut sample', '_W' stands for 'Wide sample')

TEST	Naive_1	No change	Naive_A	STS_W	STS_C	STS_A
Bias test (intercept values)	-7,620	-22,075***	-2,082	13,870*	-6,124	-2,221
HLN (eq. 11): Values for β						
vs. STS_A	-2.112**	-1.258***	0.838	-0.827***	-2.386***	
vs.Naive_A	-1.560**	-1.049***		-1.082***	-1.725***	-0.838
BSI's Predictive information (eq. 13): Values for β1						
STS_A vs.				0.238	1.176***	
Naive_A vs.	0.942***	0.905***				
TW (eq.10)						
If TW>22.5: models in rows outperform rivals in columns						
vs. STS_A	31*	37***	20	16	35**	
vs.Naive_A	28	38***		17	34**	25

*** indicates a significance level lower than 0.05. ** indicates a significance level between 0.1 and 0.05. * indicates a significance level between 0.2 and 0.1.
Quantiles retrieved from: http://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2appendixc.pdf.

The usefulness of considering BSIs is also highlighted in a non-parametric framework. In fact, the exact finite-sample TW test, allows to reject the null hypothesis of equal sum of ranks in absolute errors differences, between augmented models and the worst baseline, regardless of the complexity of the adopted specifications. With reference to STS models, these two latter tests confirm the relevance, for the forecasting accuracy, of the sample width used for the estimation of trend-cycle components. In fact, STS_A specification exhibits forecasting performances significantly superior to those of STS_C only, while its forecasting capability is not (inferentially) distinguishable from that of STS_W model.

5. Results, discussion and conclusions

This study provides innovative insights into the modelling of accommodation production through business sentiment indicators (BSIs), by showing that they allow to improve both the goodness-of-fit and the real time forecasting performance of widely used time series models. In this respect, this is a pioneering study, as - to the best of our knowledge - up to now no previous work on tourism production forecasting has ever used subjective supply-side information, although a large body of evidence established the difficulty of improving the forecasting accuracy using exogenous 'hard' variables, which are often nearly impossible to forecast (e.g. the GDP at the regional level) or only weakly correlated with the tourism dynamics (e.g. 'general' price index). The rationale for the proposed approach is that hotel managers/keepers are able to evaluate the regional production dynamics, considering the impact of macroeconomic factors, non-economic variables, strategic changes in progress within the sector and also the effects of one-off events, that cannot be efficiently or fully captured by quantitative data.

To prove this assumption, we first estimated naïve and STS models, using data on tourist arrivals in the Rimini province. These baseline specifications were then augmented by a set of BSIs, selected considering different lag structures. Finally, we assessed the relative performances of the rival specifications in terms of their goodness-of-fit, residual diagnostics and forecasting performance within an inferential framework.The results led to some main conclusions.

(1) The informative content, retrievable from business surveys, improves both the goodness-of-fit and the forecasting accuracy for the accommodation production in the Rimini province, regardless of the complexity of the specification (naïve or STS).

(2) The augmented naïve specification (Naïve_A) is the most accurate among all of the tested ones, even if the augmented STS specification (STS_A) immediately follows; the difference in accuracy is not significant.

(3) BSIs are especially informative on the high-season production pattern (the most relevant season for policy makers), as evidenced by comparing the relative forecasting performances of the baseline and augmented models, in terms of MAPE, MAE and MSE.

(4) The inclusion of a (though non-significant) cycle component in STS_C and in its augmented version (STS_A), worsens mainly the forecasting performance of the baseline specification, allowing to conclude that BSIs account for the existing cyclical component of the accommodation production.

(5) Among the baseline models, the simplest specification (Naïve_1) displays a forecasting accuracy much higher than that of the complex STS_C model, estimated on the same information set. However, widening the sample size, the STS_W specification outperforms the Naïve_1 in terms of MAPE and reduces the distances in terms of the other measures of performance. This last finding partially reconciles the conflicting positions of previous papers, concerning the superiority of naïve models over more complex ones in the tourism field, attributing the differences in forecasting performances to the possibility to correctly estimate the trend-cycle component.

The above results have also important managerial and policy implications. In fact, the possibility to obtain accurate real-time forecasts at the local level, at reasonable cost, is extremely useful for policy makers in countries like Italy, where, since 1993, the tourism sector has undergone a governance devolution. Therefore, timely and reliable statistical information on accommodation production at the local level has become especially valuable for business and intervention planning, reporting and evaluation, in the entire economic sphere that revolves around tourism production. Unfortunately, while the responsibility of local policy makers increased, the competence for tourism statistics remains in the hands of the central statistical office, which is not able to provide local details of most of the information collected at the national level. Moreover, the central statistical office authorizes the publication of official (regional) data only after all peripheral branches have completed their data collection, i.e. after an average delay of 8 months for provisional information, and of 16 months for the final one. Local policy makers have repeatedly underlined this information

gap, claiming more timely and locally accurate statistical information; their involvement made the realisation of the present work possible.

The fact that in the considered area, there are more than 2,100 active hotels of any category, offering a wide range of leisure and business tourism products, strengthens this conclusion beyond the observed area. Namely, the fact that BSIs were found effective for describing and forecasting the dynamics of the accommodation business cycle, on the regional scale, suggests the opportunity to extend to the tourism sector the surveys on the business sentiment, currently realized by the provincial chambers of commerce for the manufacturing sector. Even in tourism, business surveys are computationally more manageable than consumer ones, as the reference population is usually less numerous, better trained to take part to surveys and more precisely identified. This way it is possible to effectively and timely monitor one of the most important sources of GDP in many Italian provinces, where official information is likely to be either lacking or poor in quality.

REFERENCES

- Allen, D. & Yap, G. (2009). *Investigating other leading indicators influencing Australian domestic tourism demand*. 18th World IMACS / MODSIM Congress, Cairns (Australia) 13-17 July 2009, http://mssanz.org.au/modsim09. Accessed on 20 January 2012.
- Anderson, S. T., Kellogg, R., Sallee, J. M., Curtin, R. T. (2011). Forecasting Gasoline Prices Using Consumer Survey*s. The American Economic Review*, 101, 110-114.
- Athanasopoulos, G., Hyndman, R. J., Song, H. & Wu, D. C. (2010). The tourism forecasting competition. *International Journal of Forecasting*, 27, 822-844.
- Aylmer, C. & Gill, T. (2003). Business Surveys and Economic Activity. Research Discussion Paper. Economic Analysis Department, Reserve Bank of Australia, January 2003.
- Berndt, E.K., Hall, B.H., Hall, R.E., & Hausman, J.A. (1974). Estimation and Inference in Nonlinear Structural Models. *Annals of Economic and Social Measurement*, 3, 653-665.
- Chang, C-L., Chen, C-C., McAleer, M., Huang, B-W. & Kuo, H-I. (2008). An Econometric Analysis of SARS and Avian Flu on International Tourist Arrivals to Asia. Available at Science Direct SSRN: http://ssrcom/abstract=1355109.
- Cho, V. (2003). A comparison of three different approaches to tourist arrival forecasting. *Tourism Management*. 24, 323-330.
- Clark, T. E. & McCracken M. W. (2011). Advances in forecast evaluation. *Federal Reserve Bank of St. Louis Working Paper Series*, http://research.stlouisfed.org/wp/2011/2011-025.pdf. Accessed on 5 September 2012.
- Claveria, O. & Datzire, J. (2009). Forecasting tourism demand using consumer expectations, *Tourism Review*, 85, 18-36.
- Clements, M. P. & Harvey, D. I. (2010). Forecast encompassing tests and probability forecasts. *Journal of Applied Econometrics*, 25, 1028–1062.

- Darne, O. (2008). Using business survey in industrial & services sector to nowcast GDP growth: The French case. *Economics Bulletin*, Paris.
- Doornik, J. A. & Hansen, H. (1994). An Omnibus Test for Univariate and Multivariate Normality. *Nuffield Economics Working Papers*, W4&91.
- Driver C, Urga G. (2004). Transforming qualitative survey data: performance comparisons for the UK. *Oxford Bulletin of Economics and Statistics, 66:* 71–89.
- Dwyer, L., E., Mistilis, D., & Scott, R. C. (2009). Destination & enterprise management for a tourism future. *Tourism Management*, 30, 63–74.
- Fildes, R., Wei, Y., & Ismail, S. (2011). Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *International Journal of Forecasting*, 27, 902–922.
- Forex Ltd. (2012). Trading the News EUR/USD- Trading the U.S. Consumer Confidence Survey. By David Song.
- Gelper, S., Lemmens, A. & Croux, C. (2007). Consumer sentiment & consumer spending: Decomposing the Granger causality relationship in the time domain. *Applied Economics*, 39, 1-11.
- Goh, C. & Law, R. (2002). Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention. *Tourism Management*, 23, 499-510.
- Goh, C. & Law, R. (2011). The Methodological Progress of Tourism Demand Forecasting: A Review of Related Literature. *Journal of Travel & Tourism Marketing*, 28, 296-317.
- Gonzalez, P. & Moral, P. (1996). Analysis of tourism trends in Spain. *Annals of Tourism Research*, 23, 739-754.
- Goodwin, P. (2008). A quick tour of tourism forecasting,. *FORESIGHT The International Journal of Applied Forecasting*, 10, 35-37.
- Govers H., Van Hecke, E. & Cabus, P. (2008). DELINEATING TOURISM: Defining the Usual Environment. *Annals of Tourism Research, 35,* 1053-1073.
- Granger, C.W.J. & Newbold, P. (1977). *Forecasting Economic Time Series*. Orlando, Florida: Academic Press.
- Greenidge, K. (2001). Forecasting tourism demand - An STM approach. *Annals of Tourism Research*, 28, 98-112.
- Guizzardi, A. & Bernini, C. (2013). Measuring underreporting in accommodation statistics: evidence from Italy. *Current Issues in Tourism*. 15, 597 - 602.
- Guizzardi, A. & Mazzocchi, M. (2010). Tourism demand for Italy and the business cycle. *Tourism Management*. 31, 367–377.
- Hamilton, J. M., Maddison, D. J. & Tol, R. S. J. (2005). Effects of climate change on international tourism. *Climate Research*, 29, 245–254.
- Han, Z., Durbarry, R., & Sinclair, M. T. (2006). Modelling US tourism demand for European destinations. *Tourism Management, 27,* 1–10.
- Harvey, D., Leybourne, S. & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281-291.
- Harvey, A.C. & Jaeger, A. (1993). Detrending, Stylized Facts and the Business-Cycle. *Journal of Applied Econometrics*, 8, 231-247.
- Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter,* Cambridge, UK, Cambridge University Press.
- Huang, J. H. & Min, J. C. H. (2002). Earthquake devastation and recovery in tourism: the Taiwan case. *Tourism Management*, 23, 145-154.
- Hyndman, R. & Khandakar, Y. (2008), Automatic time series forecasting: The forecast package for R, *Journal of Statistical Software, 26, 3,*

- ISTAT. (2011). Capacità degli esercizi ricettivi, retrieved the 5th of September from http://www.istat.it/it/archivio/37399.
- Jad, S. S. (2011). The use of surveys to measure sentiment and expected behavior of key sectors in the economy: evidence from the business survey conducted by the Central Bank of Lebanon. *IFC Bulletin*, 34, 248-277.
- Kim, J.H. & Moosa, I.A. (2005). Forecasting international tourist flows to Australia: a comparison between the direct and indirect methods. *Tourism Management*, 26, 69-78.
- Kauppi E., Lassila J. & T. Teräsvirta. (1996). Short-term forecasting of industrial production with business survey data: experience from Finland's great depression 1990-1993. *International Journal of Forecasting,* 12, 373-381.
- Koskinen, L. & Öller, L.-E. (2004). A classifying procedure for signalling turning points. *Journal of Forecasting*, 23, 197– 214.
- Kulendran, N. & Witt, S. F. (2001). Cointegration versus least squares regression. *Annals of Tourism Research*, 28, 291–311.
- Kulendran, N. & Witt, S. F. (2003). Leading indicator tourism forecasts. *Tourism Management*, 24, 503–510.
- Li, G., Song, H. & Witt, S. F. (2005). Recent Developments in Econometric Modeling and Forecasting. *Journal of Travel Research*, 44, 82-99.
- Li, G., Song, H. & Witt, S. F. (2006). Time varying parameter and fixed parameter linear AIDS: An application to tourism demand forecasting. *International Journal of Forecasting,* 22, 57-71.
- Lui, S., Mitchell, J. & Weale, M. (2011a). The utility of expectational data: Firm-level evidence using matched qualitative-quantitative UK surveys. *International Journal of Forecasting*, 27, 1128–1146.
- Lui, S., Mitchell, J. &Weale, M. (2011b). Qualitative business surveys: signal or noise? *Journal of the Royal Statistical Society*. Series A (Statistics in Society), 174, 327–348.
- Matheson, T. D., Mitchell, J., Silverstone, B. (2010). Nowcasting and Predicting Data Revisions Using Panel Survey Data. *Journal of Forecasting*. 29, 313–330.
- Maurin, L. & Paries, M. D. (2008). *The Role of Country-Specific Trade and Survey Data in Forecasting Euro Area Manufacturing Production: Perspective from Large Panel Factor Models*. ECB Working Paper No. 894.
- Mitchell, J., Smith, R. J. & Weale, M. R. (2005). Forecasting Manufacturing Output Growth Using Firm-Level Survey Data. *The Manchester School*, 73, 479–499.
- Njegovan, N. (2005). A leading indicator approach to predicting short-term shifts in demand for business travel by air to and from the UK. *Journal of Forecasting*, 2, 421-432.
- Pagan, A. & Harding, D. (2001). Extracting, Using and Analysing Cyclical Information, MPRA Paper, 15, University Library of Munich.
- Pedersen, M. (2009). *Use of Chilean business surveys in conjunctural assessment and short-term forecasting*. OECD workshop on business and consumer opinion surveys. Brussels, 12-13 October 2009.
- RCC, Regional Chambre of Commerce of Emilia Romagna web site: http://www.rer.camcom.it/studi-ricerche/banche-dati/bd/turismo/mensili
- Ryan, C. (2003). *Recreational Tourism: demand and impacts*. Clevedon: Channel View Publications.
- Shareef, R. & McAleer, M. (2007). Modelling the uncertainty in monthly international tourist arrivals to the Maldives. *Tourism Management*, 28, 23–45.

- Song, H., Gao, B. Z. & Lin, S. V. (2013). Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. *International Journal of Forecasting*, 29, 295-310.
- Song, H. & Li, G. (2008). Tourism demand modelling and forecasting - A review of Recent research. *Tourism Management*, 29, 203-220.
- Song, H., Li, G., Witt, S. F. & Athanasopoulos, G. (2011). Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*, 27, 855-869.
- Song, H. & Witt, S. F. (2000). Tourism demand modelling and forecasting: modern econometric approaches. Oxford: Pergamon.
- Song, H.Y. & Witt, S.F. (2006). Forecasting international tourist flows to Macau. *Tourism Management*. 27, 214-224.
- Song, H., Wong, K. K. F. & Chon, K. K. S. (2003b). Modelling and forecasting the demand for Hong Kong tourism. *Hospitality Management*, 22, 435–451.
- Swarbrooke, J. & Horner, S. (2001). *Business travel and tourism*. Oxford: Butterworth-Heinemann.
- Taylor, K. & McNabb, R. (2007*). Business Cycles and the Role of Confidence: Evidence for Europe. *Oxford Bulletin of Economics and Statistics*, 69, 185-208.
- Turner, L. W., & Witt, S. F. (2001). Factors influencing demand for international tourism: Tourism demand analysis using structural equation modelling, revisited. *Tourism Economics*, 7, 21–38.
- Vu, J.C. & Turner, L.W. (2006). Regional Data Forecasting Accuracy: The Case of Thailand. *Journal of Travel Research*, 45, 186-193.
- Wong, K. K. F., Song, H. & Chon, K. S. (2006). Bayesian models for tourism demand forecasting. *Tourism Management*, 27, 773–780.

Appendix: the dynamic of the series (in levels)

BSIs have been extensively employed, especially in the manufacturing sector, for real time forecasting the periodic and irregular up-and-down movements of economic activities, which constitute the business cycles. Our focus is investigating whether BSIs can be informative about the dynamic of the tourism business cycle.

Therefore, in the paper we report only the plots of the series in seasonal differences, leaving in this appendix the graphical representation of the series in levels. Figure 2 benefits those readers interested in visualizing how the measured business sentiment indicators look like, along with the trend and the seasonal components, the statistical analysis of which is in section 2.2.
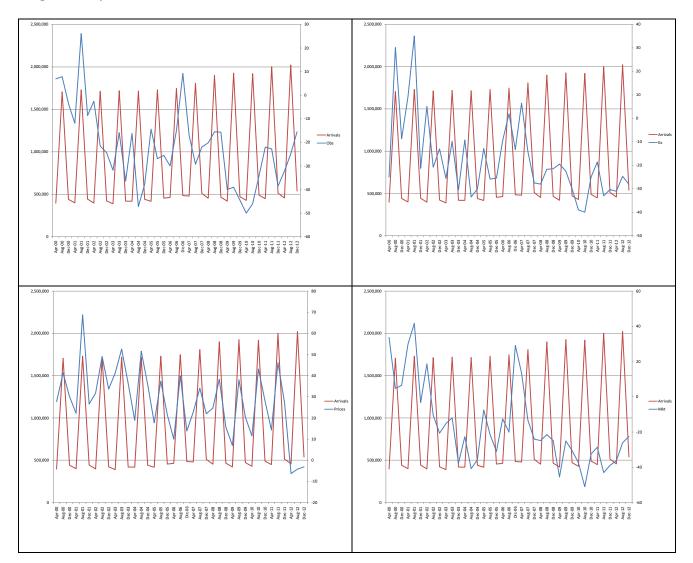
Figure 2: Dynamic of the series in levels.