Can tourism confidence index improve tourism demand forecasts?

Valeria Croce

Valeria Croce is based at European Travel Commission, Brussels, Belgium.

Abstract

Purpose – The link between confidence and economic decisions has been widely covered in the economic literature, yet it is still an unexplored field in tourism. The purpose of this paper is to address this gap, and investigate benefits in forecast accuracy that can be achieved by combining the UNWTO Tourism Confidence Index (TCI) with statistical forecasts.

Design/methodology/approach – Research is conducted in a real-life setting, using UNWTO unique data sets of tourism indicators. UNWTO TCI is pooled with statistical forecasts using three distinct approaches. Forecasts efficiency is assessed in terms of accuracy gains and capability to predict turning points in alternative scenarios, including one of the hardest crises the tourism sector ever experienced.

Findings – Results suggest that the TCI provides meaningful indications about the sign of future growth in international tourist arrivals, and point to an improvement of forecast accuracy, when the index is used in combination with statistical forecasts. Still, accuracy gains vary greatly across regions and can hardly be generalised. Findings provide meaningful directions to tourism practitioners on the use opportunity cost to produce short-term forecasts using both approaches.

Practical implications – Empirical evidence suggests that a confidence index should not be collected as input to improve their forecasts. It remains a valuable instrument to supplement official statistics, over which it has the advantage of being more frequently compiled and more rapidly accessible. It is also of particular importance to predict changes in the business climate and capture turning points in a timely fashion, which makes it an extremely valuable input for operational and strategic decisions.

Originality/value – The use of sentiment indexes as input to forecasting is an unexplored field in the tourism literature.

Keywords Confidence index, Forecast combination, Forecasts comparative accuracy, International tourism demand, Panel of experts, Short-term forecasts

Paper type Research paper

1. Introduction

Tourism forecasting is of great value for tourism practitioners, as it provides strategic input to the decision-making process of both governmental bodies and businesses. Forecasts on the long-term development of demand are a fundamental input when formulating and implementing strategies and policies. Short-term forecasts are equally relevant to operational decisions about capacity, pricing and all other factors that can vary in the short run. As a service sector, excess output of tourism service providers cannot be stocked, hence short-term adjustments can solidly contribute to optimise the use of existing resources. This is best exemplified by the aviation and hotel pricing models, where adjustments are currently made on a daily basis, if not in real-time.

A broad range of alternative approaches has been developed to forecast tourism demand over the past decades. Quantitative methods dominate by large the existing literature, and

a number of studies also report about the usefulness of qualitative approaches (for reviews, see Peng *et al.*, 2014; Goh and Law, 2011; Song and Li, 2008; Witt and Witt, 1995). What remains still a largely unexplored field is the use of sentiment indexes as input to forecasting (Croce *et al.*, forthcoming; Mihalic *et al.*, 2013; Kester and Croce, 2011).

A sentiment or confidence index measures the current and prospected situation based on the evaluations of a sample of individuals. Business surveys typically collect evaluations from a fixed panel of experts across firms representing the sector of interest, while consumer surveys typically target households. In the broader economic literature, sentiment indexes have been widely employed for the purpose of forecasting. The extent to which these indicators can forecast economic activity has also been exhaustively researched, with findings suggesting that they contain information that goes beyond economic fundamentals, suitable to improve forecasts (Eppright *et al.*, 1998; Bodo *et al.*, 2000). Consumer confidence indexes, for instance, have a degree of explanatory power in predicting changes of the economic output (Christiansen *et al.*, 2013; Taylor and McNabb, 2007; Easaw *et al.*, 2005; Bram and Ludvigson, 1998; Santero and Westerlund, 1996), and a good predictive power in identifying discrete turning points in the business cycle, while they prove less efficient in predicting changes in consumer expenditure, with appreciable differences across countries (Taylor and McNabb, 2007).

A confidence index for the tourism sector is measured by the World Tourism Organisation, the United Nations agency specialised in tourism (UNWTO). Since 2003, UNWTO has been surveying the level of confidence in the sector, and regularly disseminating Tourism Confidence Index (TCI) values in its "UNWTO World Tourism Barometer" publication. The Index is meant to assist governments, business leaders and other decision makers around the world in formulating tourism policies. Its usefulness was tested lately, during the 2008/2009 economic and financial crisis, when the UNWTO Tourism Confidence Index proved its ability to capture the impact on the tourism sector as the crisis unfolded. Ex-post analyses revealed a strong correlation between expert prospects and actual values, and renewed the interest in the use of a sentiment index as input to forecast tourism demand.

This study is a first attempt to examine whether the UNWTO TCI helps predicting changes in tourism demand. The focus is on the intra-year development of international tourism demand, worldwide and in world regions. UNWTO TCI is used in combination[1] with statistical forecasts, to test if sentiment brings benefits in forecast accuracy. If any, contributions brought by the index are measured in terms of accuracy gains and assessed in alternative scenarios, including one of the hardest crises the tourism sector ever experienced (Smeral, 2010). The usefulness of this indicator in predicting turning points is also reported. The outcome of this research is meant to provide operational guidance to tourism practitioners on the opportunity cost to produce short-term forecasts using alternative approaches.

2. Forecasting short-term tourism demand

Since 1990s, tourism demand modelling advanced significantly, with the adoption of up-to-date econometric methodologies, such as error correction models, time varying parameter models and structural equation models. Methods based on state-of-the-art technology have also been introduced recently, among which genetic algorithms, fuzzy theory and neural networks (for references see Peng *et al.*, 2014; Song and Li, 2008; Li *et al.*, 2005; Petropoulos *et al.*, 2006). However, findings are still contradicting and none of the proposed methods proved to its superiority in all circumstances (Song and Li, 2008; Wong *et al.*, 2007; Li *et al.*, 2006a, b; Witt and Witt, 1995).

Econometric models tend to outperform time-series methods when arrivals for a specific origin-destination pair have to be forecast, but often fail in various other areas (Smeral, 2010; Song *et al.*, 2008; Witt and Witt, 1995; Dawes *et al.*, 1994). Whenever the task involves large aggregates of international arrivals, as in this study, auto-regressive moving average methods proved to be the most efficient approach, as the effect of key exogenous variables – which determine demand at a country level – or the impact of special events, tend offset each other at that level (see Smeral,

2007; Papatheodorou and Song, 2005; du Preez and Witt, 2003). Under appropriate conditions, ARIMA models can be interpreted as the univariate, hence more parsimonious, time-series representation of structural models (Holden *et al.*, 1990), with the non-negligible benefit of low-marginal costs coupled with high-marginal benefits. Evidence from the tourism forecasting literature also points to auto-regressive models as to the best approach to intra-year forecasts.

Further, accuracy improvements can be sought through combination. Judgment and statistical methods are typically conceived as alternative approaches to produce forecasts, though their characteristics suggest treating them as complementary approaches. Statistical models proved to be superior in detecting patterns objectively and in attributing an equal weight to past and recent observations (Makridakis *et al.*, 1998). Judgment-based forecasts, on the other end, incorporate "those economic and market phenomena that are known but not quantified" (Caniato *et al.*, 2011) and tend to more effectively adjust forecasts in a timely manner (Lawrence *et al.*, 2006). As qualitative and quantitative approaches also tend to be based on different types of predictors, theory suggests that their combination would lead to results that are more accurate than using each method alone (Bram and Ludvigson, 1998; Garner, 1991).

Forecast combination techniques acknowledge the complementarities of the two approaches, in the attempt to compensate for their deficiencies. Bates and Granger (1969) initiated this field of research, demonstrating that the combination of two unbiased model-based forecasts leads to a new, more accurate forecast. Progressively, studies on the combination of statistical forecasts with judgmental forecasts proved to produce, under appropriate conditions, more accurate results than each approach used separately (Fildes *et al.*, 2006, 2009; Goodwin, 2002; Goodwin and Fildes, 1999), although sometimes gains are only of modest magnitude (Holden *et al.*, 1990).

The many empirical studies available in the general forecasting literature help outlining principles for a meaningful combination of forecasts. Empirical evidence suggests that combination performs best when forecasts are based on different information, and the broadest possible range of techniques (see Fildes *et al.*, 2006, 2009; Goodwin, 2002; Flores and White, 1989 for a review of findings). Unbiased forecasts and constant covariance are good premises for the use of weighted approaches (Holden and Peel, 1986). When expert forecasts are combined, heterogeneity of respondents is another desirable characteristic (Figlewski, 1983). Combination works best when used on series of medium to low volatility. When dealing with volatile series, evidence suggests that combined forecasts tend to be consistently less distant, but hardly ever close to the actual value than individual forecasts (Wong *et al.*, 2007; Song and Li, 2008; Shen *et al.*, 2008).

In the tourism forecasting literature, studies on forecast combination are skewed towards quantitative models. Attempts of combining econometric models have been published, for example, by Li *et al.* (2006a, b). In general, the combination of forecasts produced with different methods is strongly recommended for the tourism sector (Song *et al.*, 2009), but very few tourism studies made an attempt to combine model-based and expert forecasts. Tideswell *et al.* (2001) combined quantitative forecasts with the results of a Delphi survey. Findings suggest that the forecast accuracy was quite high overall, for both international and domestic visitors, but judgmental forecasts based on a limited number of experts tended to be volatile. More recently, Song *et al.* (2011) proved that judgmental adjustments of statistical forecasts improve forecast accuracy.

This study suggests using prospects provided by members of UNWTO Panel of Tourism Experts as judgmental forecasts, to be used in combination with statistical forecasts. The rationale is to use information captured through the UNWTO survey to improve the accuracy of model-based forecasts. Accuracy improvements, if any, are also tested in two alternative scenarios, one associated with stability ("business-as-usual") and one with economic or political uncertainty ("crisis").

An original aspect of the proposed approach, compared to the existing literature on forecast combination, is the integration of statistical and judgmental forecasts, as opposed to their

combination. Most frequently, the combination of qualitative and quantitative forecasts follows a sequential use of each method: experts use statistical forecasts as input to produce their judgmental forecasts; alternatively, predictive models can be extrapolated from experts' data processing approach and used in quantitative models (Armstrong, 2001). Integration implies pooling qualitative and quantitative forecasts produced separately. Such sets of data are seldom available, and therefore object of few studies even in the broad forecasting literature (see for instance Garner, 1991; Bram and Ludvigson, 1998). In this respect, UNWTO expert forecasts and statistical database offer a unique set of data to test the proposed approach in a real-life setting and on a worldwide scale.

3. Research design

The major aim of this study is to test accuracy gains, if any, brought by the UNWTO TCI when used in combination with statistical forecasts. To achieve this goal the efficiency of three combinatory approaches is tested against individual forecasts, both in terms of accuracy gains and capability to predict turning points.

3.1 Statistical forecasts

Actual values of tourism demand are approximated by UNWTO series of international tourist arrivals (ITA), a selection of data included in the UNWTO database on world tourism statistics. This database contains a variety of series for over 200 countries and territories covering data for most countries. Monthly ITA since 1999 have been aggregated into four-month series and used to estimate model parameters for extrapolative forecasts. Each series of actual data consists of 45 observations. Figure 1 shows series of ITA, actual and forecast values, and series of the TCI for the six regional aggregates. Charts also highlight periods related to crises of various natures.

ARIMA models are used to forecast ITA. Series of ITA have been tested for unit root using the augmented Dickey-Fuller test. For each series, alternative ARIMA models have been estimated and tested using R "forecast" package, based on the analysis of each series' autocorrelation and partial autocorrelation functions. The best candidate model has been selected on the basis of various diagnostics and evaluation criteria (see Table I[2], more criteria – such as the results of the Dickey-Fuller test, AICc and BIC values – are available upon request).

3.2 Qualitative forecasts

Prospects of the UNWTO TCI are used as judgmental input. Prospects measure the perceived short-term development of the tourism sector on a five-step Likert scale. Prospects are collected by the means of regular e-mail surveys among members of UNWTO Panel of Tourism Experts, a heterogeneous group of tourism experts from different sector from all over the world. Since the second quarter of 2003, when the survey started, 1,127 experts participated in at least one of the 33 waves conducted to date, with an average of 300 participating experts per wave. The number of participating experts has regularly grown over time, but differences across regions remain noteworthy: a considerable number of experts provide, on average, estimates for the series Europe (119), Americas (72) and Asia and the Pacific (64), while the respondents' base for the series Africa (20) and Middle East (14) are typically smaller.

The TCI is derived from the ratio between the positive and negative responses collected through this survey (see UNWTO, 2015). In order to be compared with actual values, the index values have been homologised to series of ITA, these latter expressed as percentage change on the previous period. The homologation process takes the form of a linear transformation as follows:

$$\overline{P}_t = a + bP_{t-1} \tag{1}$$
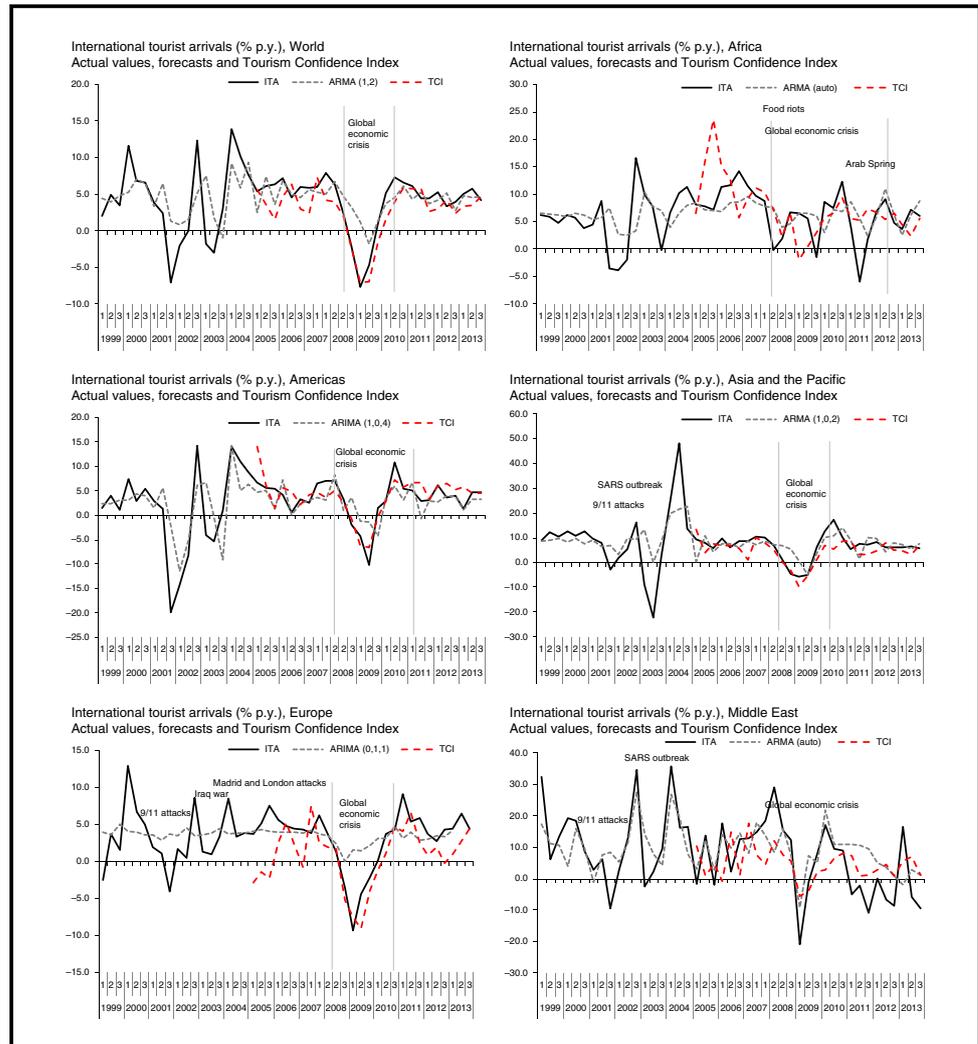
where P is the index value at time t. Intercept a and slope b are estimated through a linear regression between the index and the equivalent series of ITA to return estimated growth values $(\bar{P})$ at each time, hereafter referred to as homologised TCI. Series of homologised TCI values are used in combination with statistical forecasts.

3.3 Integration approaches

This paper tests the efficiency of three different approaches, which can be used to integrate forecasts. One method is the pioneering variance-covariance (VARCO) approach, proposed by Bates and Granger (1969) and based on a quadratic loss function to minimise error variance. An alternative approach is the Discounted Mean Square Forecast Error (DMSFE), which assumes that recent forecasts can better help predicting future values and weights them more heavily than distant ones (Diebold and Pauly, 1987; Winkler and Makridakis, 1983). A simple average combination is also computed, as baseline for comparisons (Armstrong, 1989).

The discounted MSFE method was introduced by Bates and Granger (1969) and generalised by Newbold and Granger (1974) for multiple forecast combination. Based on the outcome of

Table I Forecast models, evaluation criteria and specifications

Series	TCI r^2	Model	ARIMA r^2	AIC	Forecast periods Overall	n	Crisis	n	BaU	n
World	0.77**	ARIMA (1,0,2)	0.60**	160.2	2008/q1-2013/q3	16	2008/q1-2009/q3	6	2010/q1-2013/q3	10
Africa	0.12**	ARIMA (1,0,0)[a]	0.20	185.8	2008/q1-2013/q3	18	2008/q1-2012/q1	13	2012/q2-2013/q3	5
Americas	0.60**	ARIMA (1,0,4)	0.46**	189.6	2008/q1-2013/q3	18	2008/q1-2010/q3	9	2011/Q1-2013/q3	9
Asia Pacific	0.58**	ARIMA (1,0,2)	0.39**	207.1	2008/q1-2013/q3	18	2008/q1-2010/q2	8	2010/Q3-2013/q3	10
Europe	0.53**	ARIMA (0,1,1)	0.35**	145.5	2008/q1-2013/q3	18	2008/q1-2011/q1	10	2011/Q2-2013/q3	8
Middle East	0.15*	ARIMA (1,0,3)[a]	0.39**	272.1	2010/q2-2013/q3	11	–		–	

Notes: [a]Values refer to the initial model fitting in-sample data. *,**Significant at the $p < 0.05$ and $p < 0.01$ levels, respectively

forecast combination studies in tourism literature (Shen *et al.*, 2008; Song *et al.*, 2008; Wong *et al.*, 2007), weights for the ith forecast have the form as follows:

$$w_i = \sum_{t=1}^{T} \phi_{i,t}^{-1} / \sum_{i=1}^{n} \left(\sum_{t=1}^{T} \phi_{i,t}^{-1} \right) \qquad (2a)$$

where ϕ_{it} is the discounted sum of past square forecast errors:

$$\phi_{i,t} = \sum_{s=1}^{t-h} \alpha^{t-h-s} \left(y_s - f_{i,s} \right)^2 \qquad (2b)$$

y_s and $f_{i,s}$ are, respectively, the actual and forecast value at period s, h is the forecast horizon and α is a discounting factor terminated between 1 and 0. A high value of α implies that more weight is given to past information[3]. In this study the value of α is selected based on an iterative process that identifies the value that minimises the in-sample MAE value for the reference series.

3.4 Accuracy improvements

Combining forecasts should bring to gains that justify the use of a sophisticated approach. It is a commonly accepted wisdom that criteria to distinguish between good and poor forecasts need to be tailored to the specific forecasting task, as each measure synthesizes different aspects of forecasts error series (Armstrong and Fildes, 1995). In this study, accuracy is expressed in terms of Mean Absolute Errors (MAE), a measure that provides the most informative and direct information about forecast errors distribution. The MAE has been preferred, to accuracy measures based on percentages or squared errors, which more frequently used in forecasting studies, as these latter would lead to misleading results with series of growth rates, as they suffer of instability when actual values are equal or close to zero.

Accuracy gains brought by combining forecasts, if any, are assessed in terms of percentage increments compared to the best individual forecast. Accuracy gains are tested on out-of-sample values, which are further grouped into a "crisis period"[4] and a "business-as-usual" period (see Table I)[5]. The Diebold-Mariano (1995)[6] statistics is eventually computed on series of absolute errors, to compare the predictive accuracy of each combinatory approach against the best performing individual forecast. This statistics is particularly suited to compare the accuracy of model-free forecasts, as for instance with survey-based

forecasts. Furthermore, the Diebold-Mariano test accommodates for a number of series characteristics, among which the presence of serially correlated forecast errors (see Diebold and Mariano, 1995, p. 10), which characterises the benchmark of a compound forecast with one of its inputs (see Table II). Coupled with the short forecast horizons, forecast errors correlation can lead to particularly conservative results of the DM test, with the null hypothesis being rejected too often. This may explain the limited number of statistically significant results, and encourages the interpretation of significant ones as solid recommendations about the validity of the correspondent forecasting approach.

3.5 Turning points

Pattern of movements in series of ITA present a sequence of downturn and upturn regimes. A turning point is defined as the point when the regime shifts. Series co-movements are assessed using the concordance statistics originally proposed by Harding and Pagan (1999). This simple non-parametric statistics measures the proportion of times two series are in the same state. Series of actual values and forecasts are transformed into binary indicator series ($S_{i,t}$ and $S_{j,t}$), where a value of 0 corresponds to a contraction (i.e. growth rate is 0 or lower) and a value of 1 corresponds to an expansion. The degree of concordance between transformed series as follows:

$$C_{i,j} = T^{-1} \left\{ \sum_{t=1}^{T} \left(S_{i,t} S_{j,t} \right) + \left(1 - S_{i,t} \right) \left(1 - S_{j,t} \right) \right\} \tag{3}$$

where T is the sample size. The higher the concordance value, the closer the patterns of the actual and forecast value series is.

Table II	Forecast error series correlation (combination model against best performing forecast)		
Period	*SA*	*VARCO*	*DMSFE*
World			
Overall	0.15	−0.13	0.33
BaU	0.90	−0.13	0.44
Crisis	−0.55	−0.01	0.37
Africa			
Overall	0.87	0.97	0.93
BaU	−0.26	−0.08	−0.12
Crisis	0.88	0.98	0.94
America			
Overall	0.48	0.72	0.63
BaU	0.28	0.96	0.47
Crisis	0.45	0.62	0.63
Asia Pacific			
Overall	0.75	0.00	0.78
BaU	0.58	0.43	0.48
Crisis	0.78	−0.22	0.87
Europe			
Overall	0.87	0.70	0.97
BaU	0.95	0.71	0.99
Crisis	0.94	0.71	0.99
Middle East			
Overall	0.89	0.30	0.75
BaU	−	−	−
Crisis	0.92	0.36	0.84

3.6 Hypotheses

Based on the methodology described above, benefits in forecast accuracy, which can be achieved by combining the homologised TCI and ARIMA forecasts, are tested based on the following set of hypotheses.

For each series, for the whole set of data and for each of the two scenarios (crisis and business-and-usual), it can be expected that:

H1a. The predictive accuracy of a simple average combinatory approach is higher than that of the most accurate individual forecast.

H1b. The predictive accuracy of a VARCO combinatory approach is higher than that of the most accurate individual forecast.

H1c. The predictive accuracy of a DMSFE combinatory approach is higher than that of the most accurate individual forecast.

The following hypotheses concerning individual forecasting approaches are also tested for each series:

H2. The predictive accuracy of the homologised TCI is higher than that of the corresponding statistical model.

4. Empirical results

The TCI and ARIMA forecasts have been regressed[7] against the corresponding series of ITA, to assess their predictive power. Results (Table I) suggest that, for most series, the homologised TCI explains a larger part of variation in ITA than statistical forecasts. This is coherent with previous research findings, showing that judgmental forecasts' main value is in predicting directional change. In both cases, forecasts predictive power is considerably reduced for those two series marked by high volatility and a small base of experts, namely the Middle East. Still, for this latter region the TCI returns a most accurate forecast than any individual or combined forecast (Table III).

Table III Mean absolute error, concordance statistics and accuracy tests – overall sample

| | Mean absolute error | | | | | | | |
| | Individual forecasts | | Combined forecasts | | | | | |
	TCI	ARIMA	Simple av.	VARCO	DMSFE[a]	Simple av. (%)	VARCO (%)	DMSFE[a] (%)
World	1.54	2.02	1.59	1.78	1.80	3	16	17
Concordance statistics	0.94	0.88	0.88	0.88	0.94	–	–	–
Africa	3.47	3.18	3.08	3.00*	3.03	–11	–14	–13
Concordance statistics	0.82	0.88	0.88	0.88	0.88	–	–	–
Americas	1.66	2.62	1.72	1.85	1.63	3	11	–2
Concordance statistics	0.94	0.82	0.88	0.88	0.88	–	–	–
Asia Pacific	3.16	3.30	2.39	3.59	2.20	–24	14	–30
Concordance statistics	100	0.88	0.94	0.88	1.00	–	–	–
Europe	2.29	2.76	1.91	4.05	1.78	–17	77	–22
Concordance statistics	0.83	0.72	0.94	0.78	0.89	–	–	–
Middle East	8.07*	10.45	9.05	19.59	10.03	12	143	24
Concordance statistics	0.53	0.41	0.47	0.53	0.41	–	–	–

Notes: [a]α values: World = 0.04; Africa = 0.05; Americas = 0.04; Asia Pacific = 0.25; Europe = 1.00; Middle East = 1.00. *,**,***Significant at the $p < 0.1$; $p < 0.05$ and $p < 0.01$

4.1 Accuracy gains and directional change of combined forecasts

Table III reports out-of-sample MAE and concordance statistics of individual and combined forecasts: an italics font denotes the best performing forecast, while asterisks denote for which series the hypothesis of equal accuracy with the best individual forecast could be rejected, based on Diebold-Mariano test values.

Overall results confirm that forecasts combination leads to consistent gains in forecast accuracy, which can be generalised only under specific conditions. In four of the six series in exam, at least one of the combinatory approaches returns forecasts, which are more accurate than the best individual forecast. Combination is particularly appropriate for the series Africa, where both constituent forecasts perform poorly. For ARIMA forecasts, this can be partly explained by the high volatility of ITA to the region, frequently caused by geo-political unrest. The recent food riots, the uprisings linked to the Arab Spring and the break out of the Syrian conflict are the most recent examples of events that negatively affected tourism flows in Africa. The low number of African experts participating in UNWTO survey may explain instead the sub-performance of the TCI. In such a situation, combination successfully pools complementary information provided by each of the constituent forecasts, as each of the combinatory approaches returns more accurate results than their inputs. The Diebold-Mariano test points to a significantly more accurate forecast when the VARCO approach is used, suggesting that controlling for error variance is relevant under such circumstances.

The largest though not significant accuracy gains are achieved for the series Asia Pacific (−30 per cent) and Europe (−22 per cent). Compared to the best performing constituent forecast, DMSFE combination brings the MAE down by nearly 1 percentage point, which is an appreciable result. Both series are marked by a comparatively stable growth pattern, which is altered by the impact of large events such as the 2008/2009 economic crisis for the period in exam. DMSFE weights correctly calibrate the contribution of each input according to the operating environment, meaning that the TCI input – more sensitive to changes – is assigned a higher weight during periods of instability, while statistical forecasts – better in extrapolating long-term trends – are comparatively more relevant in periods of stability. α values suggest that recent values are more relevant to produce accurate forecasts for the series Europe, while a longer memory is crucial to determine appropriate weights for the series Asia and the Pacific. This can be explained by the magnitude of the impact of crises, as opposed to the frequency of their occurrences: the Asia and the Pacific region is more vulnerable to crises than Europe, due to the relatively recent development of its tourism sector and its comparatively higher dependency on visitors from outside the region.

Combination performs poorly for the highest level of aggregation, the World, as well as for the Middle East region. For the earlier, results hint that a large group of experts best captures intra-year variations that characterise ITA's flows, being in possess of recent information that can alter values which would be expected at a given time of the year. Results for the Middle East region may be better explained by series characteristics, as well as the concentration of international tourism flows in a few destinations. As for Africa, the Middle East region is marked by a comparatively recent development of tourism, high dependency from extra-regional source markets and frequent unrest, resulting in highly volatile international tourism flows. Volatility, coupled with a comparatively poor availability of comparable statistics, leads to sub-optimal conditions for the use of extrapolative forecasting methods in both regions. The number of Middle Eastern experts contributing to the UNWTO survey is the lowest of all world regions, yet they return the most accurate forecasts among those in exam. This may be explained by the fact that, different from Africa, some 70 per cent of international tourism to the Middle East is directed to only three countries, namely, Saudi Arabia, UAE and Egypt. In this setting, collecting opinions from experts from the largest receiving countries in the region seems to be sufficient to produce better forecasts than each alternative approach.

In line with previous findings (see for instance Croushore, 2005; Howrey, 2001; Bram and Ludvigson, 1998), forecast combination also brings benefits in terms of directional change

accuracy. Compared to statistical forecasts, concordance statistics of combined forecasts are at least as good as individual inputs, if not better. Still, they hardly outperform the homologised TCI in terms of concordance statistics. The only exception is the series Europe, for which a simple average combination returns a pattern closer to that of actual values series than any other forecast.

Results also confirm that forecast combination tends to work best for series of medium to low volatility. In periods of high volatility (Table IV), associated with crises of various natures, a combinatory approach brings accuracy gains to only three out of five series. For Europe, the region where the impact of the crisis lasted long, a method weighting recent forecasts more heavily than distant ones (DMSFE) correctly assigns increasingly larger weights to the TCI, an indicator that rapidly adjusts to directional changes. This method improves forecast accuracy by 13 percentage points. In Asia Pacific, the region that recovered fastest from the crisis, a method considering the historical performance of the series (VARCO) proves appropriate, and leads to a MAE which is nearly the half of the best individual forecast. Still, in a crisis scenario, the hypothesis of equal accuracy cannot be rejected for any of the combinatory approaches.

The benefit brought by combinatory approaches is best proved in a "business-as-usual" scenario, with series moving more regularly around trend values. Different combination approaches bring generous improvements in MAE values to three out of five series (Table V). For the series Americas, the predictive accuracy of a simple average is significantly higher,

| Table IV | Mean absolute error, concordance statistics and accuracy tests – crisis scenario |

| | Mean absolute error | | | | | | | |
| | Individual forecasts | | Combined forecasts | | | | | |
	TCI	ARIMA	Simple av.	VARCO	DMSFE[a]	Simple av. (%)	VARCO (%)	DMSFE[a] (%)
World	*1.20*	4.03	1.96	2.52	2.99	63	109	148
Africa	3.91	3.81	3.78	*3.71*	3.74	−1	−3	−2
Americas	*1.71**	3.85	2.37	2.40	1.94	39	41	14
Asia Pacific	3.96	4.24	3.39	*2.25*	2.99	−14	−43	−24
Europe	2.11	4.01	2.05	4.95	*1.85*	−3	134	−13
Middle East	–	–	–	–	–	–	–	–

Notes: [a]α values: World = 1.00; Africa = 0.00; Americas = 1.00; Asia Pacific = 0.25; Europe = 1.00; Middle East = na. *,**,***Significant at the $p < 0.1$; $p < 0.05$ and $p < 0.01$ levels, respectively

| Table V | Mean absolute error, concordance statistics and accuracy tests – business-as-usual scenario |

| | Mean absolute error | | | | | | | |
| | Individual forecasts | | Combined forecasts | | | | | |
	TCI	ARIMA	Simple av.	VARCO	DMSFE[a]	Simple av. (%)	VARCO (%)	DMSFE[a] (%)
World	1.68	*1.19*	1.43	1.48	1.30	21	25	10
Africa	2.33	1.52	1.27	*1.15*	1.16	−16	−24	−24
Americas	1.62	1.38	*1.07**	1.30	1.31	−22	−6	−5
Asia Pacific	2.52	2.56	1.58	4.66	*1.56*	−38	82	−39
Europe	2.52	*1.18**	1.73	2.92	1.70	47	148	44
Middle East	–	–	–	–	–	–	–	–

Notes: [a]α values: World = 0.85; Africa = 0.05; Americas = 0.50; Asia Pacific = 0.25; Europe = 0.40; Middle East = 1.00. *,**,***Significant at the $p < 0.1$; $p < 0.05$; $p < 0.01$ levels, respectively

in a statistical sense, than each constituent forecast. This suggests that the homologised TCI can efficiently adjust patterns extrapolated with statistical models when small-sized events occur.

4.2 Predictive accuracy of individual forecasts

The Diebold-Mariano test is also used to identify statistically significant differences in the predictive accuracy of each individual forecast.

Overall, the homologised TCI is typically more accurate than ARIMA forecasts, but the hypothesis of equal accuracy can be rejected only for the most volatile series, the Middle East ($p < 0.1$). During the global financial and economic crisis, the index proves significantly more accurate than the ARIMA model for the series Americas, where the model, extrapolating patterns from previous crisis of shorter duration, changes sign too early. On the other end, in a "business-as-usual" scenario, the index is frequently outperformed by its model-based counterpart. Results for the region Europe are particularly noteworthy, as the ARIMA model returns significantly more accurate forecasts than the index, despite the large number and variety of European experts participating in the UNWTO survey.

5. Caveats and conclusions

The link between confidence and economic decisions has been widely covered in the economic literature, yet it is still an unexplored field in tourism. This study addresses this gap, and demonstrates that a TCI can provide meaningful indications about the sign of future growth in ITA, and that it can also significantly contributes to improve forecast accuracy in specific occasions.

Empirical results prove that the UNWTO TCI well captures changes in tourism demand generated by external shocks, but also by short-term, systematic factors. UNWTO TCI proves efficient in identifying turning points and, combined with statistical forecasts, it also contributes to improve the number of correctly signed observations. Directional change is an important aspect in tourism forecasting research (Liu, 1988; Kim and Moosa, 2005), with high-practical value, as tourism practitioners are keen to know the timing of change in tourism growth. Limited research has been conducted in forecasting directional change and turning points so far. Improving this aspect can contribute to the effectiveness of both business planning in the private sector and macroeconomic policy making in the public sector (Song and Li, 2008).

The capability to predict variations in international tourism demand flows is particularly relevant to assist policy formulation at the moment of a shock, as well demonstrated by the 2008/2009 economic and financial crisis. As the crisis unfolded, a number of challenges arose and increased uncertainty as to the depth and extent of its impact on tourism. Most advanced economies experienced a sharp decline in their economic activities, coupled with rising unemployment – a major source of uncertainty – and concern about recovery opportunities. In Europe, fears of a sovereign debt crisis progressively spread among investors. In this climate, leading indicators' capability to predict future developments of international tourism demand weakened, even over a short-term horizon. The crisis forced major economic institutions to constantly revise their short-term forecasts to keep pace with events, with a trickle-down effect on model-based tourism forecasts. In this scenario, the UNWTO TCI offered effective support to organisations and policy makers, as it estimated the upcoming impacts of the crisis independently from other economic inputs. Ex-post analyses revealed a stronger resilience of tourism compared to other sectors of the economy, a factor that could not have been captured by model-based forecasts.

Empirical evidence suggests that a sentiment index is a valuable instrument to supplement official statistics, over which it has the advantage of being more frequently compiled and more rapidly accessible. Forecast combination proves particularly appropriate for regions where statistical information is scarce or hardly comparable, as for some African and Middle Eastern countries. This factor is also relevant to smaller geographical aggregates, such as regions or

cities, where the availability of future-oriented indicators is typically scarce. Expert forecasting is a cost-effective method to obtain indications about the future evolution of a phenomenon, even beyond the short-term horizon. Due to the comparatively low cost of collecting expert opinions, panels of expert could be more widely adopted to compensate for the scarcity of tourism indicators (Croce *et al.*, forthcoming).

Results also suggest that the combination of a sentiment index with quantitative data can be a cost-effective solution to deal with volatile series. Modelling the stochastic component of volatility in ARIMA models implies the development of "asymmetric" volatility models, as negative shocks impact more on tourism demand than positive shocks would. ARIMA-volatility methods tend to sub-perform as stand-alone forecasts, and are best used in combination with other types of forecasts (Coshall, 2009). The approach proposed in this paper may offer a simpler and similarly efficient solution to embed volatility in extrapolative models, whose implementation does not require particular statistical skills.

This study confirms that a sentiment index can efficiently capture the sector's dynamics, and bring these changes into a combined forecast. Yet, the lack of significant results for most series is certainly a major limitation of this study. This can be partly justified by the conceptual discrepancy between the index and actual series: while ITA measure inbound travel flows, the TCI captures changes in the overall tourism sector, including domestic demand flows. This index therefore returns a measurement of the overall business climate rather than just its demand component, which may explain part of the error magnitude.

Used in combination with statistical forecasts, UNWTO TCI tends to improve the forecast accuracy, but results vary greatly across regions and can hardly be generalised. A combination approach is to be preferred when both constituent forecasts perform poorly, as it is the case for the region Africa. The lack of significant results suggests that the combination of forecasts produced separately should be preferred if the goal of the analysis is to avoid the risk of selecting the worst performing model, in line with previous findings (Song *et al.*, 2009).

Recommendations on which combinatory approach to select are also difficult to draw, as performance tends to vary across series and scenarios. Results advice against the use of a simple average combination, especially with volatile series. Weighted approaches, and especially the DMSFE approach, bring appreciable accuracy gains in most cases, although never to a statistically significant level. The VARCO approach seems to perform best with events whose impact is limited in time.

This study is to be seen as a preliminary step in the assessment of the TCI predictive power. This paper focuses on four-month prospects as a substitute for judgmental forecasts on international tourism volumes. Further research is needed to sees the index predictive power against different indicators of tourism performance, both from the demand and supply, to better understand its value in tourism forecasting. Research based on full year prospects may instead be directed to assess the statistical significance of the TCI against other predictors typically used on model-based forecasts, such as GDP and cost of travel. Improvements brought by the use of changes in the confidence index, as opposed to levels, could also be tested.

Notes

1. The terms "combination", "integration" and "aggregation" refer to the process of synthesising different forecast values into a single value. "Combination" is the term occurring most frequently in tourism studies. In the general literature, when the aggregated forecast results from a staticised process, the use of the term "aggregation" seems to be preferred, while "integration" seems to be preferred when qualitative and quantitative are produced separately, and then pooled into a combined forecast. In the revised literature, the three terms seem to be frequency used as synonyms, and the same applies to this paper.

2. Due to the high volatility of the series, data for the Middle East region don't allow for a distinction between crisis and non-crisis scenario.

3. This variant ignores the effect of the covariance on weights, which is instead considered in the VARCO approach. This variant of the DMSFE has been selected to avoid the risk of obtaining identical results to a VARCO approach for $\alpha = 1$.

4. For each series, the crisis period starts with the first period after a turning point that precedes one (or more) negative peak(s) and ends at the turning point associated with the first positive peak after the crisis. This definition entails anomalous growth values, both positive and negative, which can be attributed to the impact of that crisis. On an operational level, this definition grants a sufficient number of observations to valuate accuracy gains.

5. For the Middle East series, accuracy gains are tested only for the crisis period due to the instability of actual values during the period for which prospects are available.

6. The DM test has been chosen as measure of significance due to the non-zero mean and serially correlated nature of forecast error series. Empirical applications of the test suggest that on small samples the test can have the wrong size and reject the null hypothesis too often. For this purpose, confidence levels start at 0.1.

7. Cubic models best fit series of prospects and arrivals, with the exception of the series World and Americas, which are modelled with a linear and quadratic regression, respectively.

References

Armstrong, J.S. (2001), "Selecting forecasting methods", in Armstrong, J.S. (Ed.), *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Kluwer Academic Publishers, Norwell, MA, pp. 365-386.

Armstrong, S.J. (1989), "Combining forecasts: the end of the beginning or the beginning of the end?", *International Journal of Forecasting*, Vol. 5 No. 4, pp. 585-8.

Armstrong, S.J. and Fildes, R. (1995), "On the selection of error measures for comparison among forecasting methods", *Journal of Forecasting*, Vol. 14 No. 1, pp. 67-71.

Bates, J.M. and Granger, C.W.J. (1969), "The combination of forecasts", *OR*, Vol. 20 No. 4, pp. 451-68.

Bodo, G., Golinelli, R. and Parigi, G. (2000), "Forecasting industrial production in the Euro area", *Empirical Economics*, Vol. 25 No. 4, pp. 541-61.

Bram, J. and Ludvigson, S. (1998), "Does consumer confidence forecast household expenditure? A sentiment index horse race", *Economic Policy Review of the Federal Reserve Bank of New York*, Vol. 4 No. 2, pp. 59-78.

Caniato, F., Kalchschmidt, M. and Ronchi, S. (2011), "Integrating quantitative and qualitative forecasting approaches: organizational learning in action research case", *Journal of Operational Research Society*, Vol. 62 No. 3, pp. 413-24.

Christiansen, C., Eriksen, J.N. and Møller, S.V. (2013), "Forecasting US recessions: the role of sentiments", CREATES Research Papers 2013-2014, available at: ftp://ftp.econ.au.dk/creates/rp/13/rp13_14.pdf (accessed 2 May 2014).

Coshall, J.T. (2009), "Combining volatility and smoothing forecasts of UK demand for international tourism", *Tourism Management*, Vol. 30 No. 4, pp. 495-511.

Croce, V., Wöber, K. and Kester, J. (forthcoming), "Expert identification and calibration for collective forecasting tasks", *Tourism Economics*, IP Publishing Ltd, available at: http://www.ingentaconnect.com/content/ip/tec/pre-prints/content-teft183

Croushore, D. (2005), "Do consumer-confidence indexes help forecast consumer spending in real time?", *The North American Journal of Economics and Finance*, Vol. 16 No. 3, pp. 435-50.

Dawes, R., Fildes, R., Lawrence, M. and Ord, K. (1994), "The past and future of forecasting research", *International Journal of Forecasting*, Vol. 10 No. 1, pp. 151-9.

Diebold, F.X. and Mariano, R.S. (1995), "Comparing predictive accuracy", *Journal of Business and Economic Statistics*, Vol. 13 No. 3, pp. 253-63.

Diebold, F.X. and Pauly, P. (1987), "Structural change and the combination of forecasts", *Journal of Forecasting*, Vol. 6 No. 1, pp. 21-40.

du Preez, J. and Witt, S.F. (2003), "Univariate versus multivariate time series forecasting: an application to international tourism demand", *International Journal of Forecasting*, Vol. 19 No. 3, pp. 435-51.

Easaw, J.Z., Garratt, D. and Heravi, S.M. (2005), "Does consumer sentiment accurately forecast UK household consumption? Are there any comparisons to be made with the US?", *Journal of Macroeconomics*, Vol. 27 No. 3, pp. 517-32.

Eppright, D., Arguea, N. and Huth, W. (1998), "Aggregate consumer expectation indexes as indicators of future consumer expectations", *Journal of Economic Psychology*, Vol. 19 No. 2, pp. 215-35.

Figlewski, S. (1983), "Optimal price forecasting using survey data", *Review of Economics and Statistics*, Vol. 65 No. 1, pp. 13-21.

Fildes, R., Goodwin, P. and Lawrence, M. (2006), "The design features of forecasting support systems and their effectiveness", *Decision Support Systems*, Vol. 42 No. 1, pp. 351-61.

Fildes, R., Goodwin, P., Lawrence, M. and Nikolopoulos, K. (2009), "Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning", *International Journal of Forecasting*, Vol. 25 No. 1, pp. 3-23.

Flores, B.E. and White, E.M. (1989), "Combining forecasts: why, when and how", *Journal of Business Forecasting Methods & Systems*, Vol. 8 No. 3, pp. 2-5.

Garner, A.C. (1991), "Forecasting consumer spending: should economists pay attention to consumer confidence surveys?", *Economic Review of the Federal Reserve Bank of Kansas City*, Vol. 76 No. 3, pp. 57-70.

Goh, C. and Law, R. (2011), "The methodological progress of tourism demand forecasting: a review of related literature", *Journal of Travel and Tourism Marketing*, Vol. 28 No. 3, pp. 296-317.

Goodwin, P. (2002), "Integrating management judgment and statistical methods to improve short-term forecasts", *Omega*, Vol. 30 No. 2, pp. 127-35.

Goodwin, P. and Fildes, R. (1999), "Judgmental forecasts of time series affected by special events: does providing a statistical forecast improve accuracy?", *Journal of Behavioral Decision Making*, Vol. 12 No. 1, pp. 37-53.

Harding, D. and Pagan, A. (1999), *Knowing the Cycle*, University of Melbourne, Melbourne.

Holden, K. and Peel, D.A. (1986), "An empirical investigation of combinations of economic forecasts", *Journal of Forecasting*, Vol. 5 No. 4, pp. 229-42.

Holden, K., Peel, D.A. and Thompson, J.L. (1990), *Economic Forecasting: An Introduction*, University of Cambridge, Cambridge.

Howrey, E.P. (2001), "The predictive power of the index of consumer sentiment", *Brookings Papers on Economic Activity*, Vol. 2001 No. 1, pp. 175-216.

Kester, J. and Croce, V. (2011), "Tourism development in advanced and emerging economies: what does the travel & tourism competitiveness index tell us?", in Blanke, J. and Chiesa, T. (Eds), *Travel and Tourism Competitiveness Report 2011*, World Economic Forum, Geneva, pp. 45-52.

Kim, J.H. and Moosa, I.A. (2005), "Forecasting international tourist flows to Australia: a comparison between the direct and indirect methods", *Tourism Management*, Vol. 26 No. 1, pp. 69-78.

Lawrence, M., Goodwin, P., O'Connor, M. and Önkal, D. (2006), "Judgmental forecasting: a review of progress over the last 25 years", *International Journal of Forecasting*, Vol. 22 No. 3, pp. 493-518.

Li, G., Song, H. and Witt, S.F. (2005), "Recent developments in econometric modelling and forecasting", *Journal of Travel Research*, Vol. 44 No. 1, pp. 82-99.

Li, G., Song, H. and Witt, S.F. (2006a), "Time varying parameter and fixed parameter linear AIDS: an application to tourism demand forecasting", *International Journal of Forecasting*, Vol. 22 No. 1, pp. 57-71.

Li, G., Wong, K.K.F., Song, H. and Witt, S.F. (2006b), "Tourism demand forecasting: a time varying parameter error correction model", *Journal of Travel Research*, Vol. 45 No. 2, pp. 175-85.

Liu, J.C. (1988), "Hawaii tourism to the year 2000: a Delphi forecast", *Tourism Management*, Vol. 9 No. 4, pp. 279-90.

Makridakis, S., Wheelright, S.C. and Hyndman, R.J. (1998), *Forecasting – Methods and Applications*, Wiley, New York, NY.

Mihalic, T., Kester, J. and Dwyer, L. (2013), "Impacts of the global financial crisis on African tourism: a tourism confidence index analysis", in Visser, V. and Ferreira, S. (Eds), *Tourism and Crisis*, Routledge, London and New York, NY, pp. 94-112.

Newbold, P. and Granger, C.W.J. (1974), "Experience with forecasting univariate time series and the combination of forecasts", *Journal of the Royal Statistical Society*, Series A, Vol. 137 No. 2, pp. 131-46.

Papatheodorou, A. and Song, H. (2005), "International tourism forecasts: time-series analysis of world and regional data", *Tourism Economics*, Vol. 11 No. 1, pp. 11-23.

Peng, B., Song, H. and Crouch, G.I. (2014), "A meta-analysis of international tourism demand forecasting and implications for practice", *Tourism Management*, Vol. 45 No. 12, pp. 181-93.

Petropoulos, C., Nikolopoulos, K., Patelis, A., Assimakopoulos, V. and Askounis, D. (2006), "Tourism technical analysis system", *Tourism Economics*, Vol. 12 No. 4, pp. 543-63.

Santero, T. and Westerlund, N. (1996), "Confidence indicators and their relationship to changes in economic activity", OECD Economics Department Working Papers No. 170, OECD Publishing, Paris.

Shen, S., Li, G. and Song, H. (2008), "An assessment of combining tourism demand forecasts over different time horizons", *Journal of Travel Research*, Vol. 47 No. 2, pp. 197-207.

Smeral, E. (2007), "World tourism forecasting-keep it quick, simple and dirty", *Tourism Economics*, Vol. 13 No. 2, pp. 309-17.

Smeral, E. (2010), "Impacts of the world recession and economic crisis on tourism: forecasts and potential risks", *Journal of Travel Research*, Vol. 49 No. 1, pp. 31-8.

Song, H. and Li, G. (2008), "Tourism demand modelling and forecasting – a review of recent research", *Tourism Management*, Vol. 29 No. 2, pp. 203-20.

Song, H., Gao, B.Z. and Li, V.S. (2011), "Combining statistical and judgmental forecasts via a web-based tourism demand forecast system", *International Journal of Forecasting*, Vol. 29 No. 2, pp. 295-310.

Song, H., Smeral, E., Li, G. and Chen, J.L. (2008), "Tourism forecasting: accuracy of alternative econometric models revisited", available at: www.wifo.ac.at/wwa/jsp/index.jsp?fid=23923&id=33239&typeid=8&display_mode=2 (accessed 26 July 2014).

Song, H., Witt, S.F., Wong, K.F. and Wu, D.C. (2009), "An empirical study of forecast combination in tourism", *Journal of Hospitality and Tourism Research*, Vol. 33 No. 1, pp. 3-29.

Taylor, K. and McNabb, R. (2007), "Business cycles and the role of confidence: evidence for Europe", *Oxford Bulletin of Economics and Statistics*, Vol. 69 No. 2, pp. 185-208.

Tideswell, C., Mules, T. and Faulkner, B. (2001), "An integrative approach to tourism forecasting: a glance in the rear view mirror", *Journal of Travel Research*, Vol. 40 No. 2, pp. 162-71.

UNWTO (2015), *UNWTO World Tourism Barometer*, Vol. 13, UNWTO, Madrid.

Winkler, R.L. and Makridakis, S. (1983), "The combination of forecasts", *Journal of the Royal Statistical Society*, Series A, Vol. 146 No. 2, pp. 150-57.

Witt, S.F. and Witt, C. (1995), "Forecasting tourism demand: a review of empirical research", *International Journal of Forecasting*, Vol. 11 No. 3, pp. 447-75.

Wong, K.K.F., Song, H., Witt, S.F. and Wu, D.C. (2007), "Tourism forecasting: to combine or not to combine?", *Tourism Management*, Vol. 28 No. 4, pp. 1068-78.

Further reading

Croce, V. and Wöber, K.W. (2011), "Judgemental forecasting support systems in tourism", *Tourism Economics*, Vol. 17 No. 4, pp. 709-24.

Kim, J.H. and King, M.L. (2006), "Forecasting international tourist flows to Australia: a comparison between the direct and indirect methods", *Tourism Management*, Vol. 26 No. 1, pp. 69-78.

About the author

Valeria Croce has worked as a Research and Development Manager at the European Travel Commission (ETC) since 2012. In her current position she is responsible for devising and implementing the ETC research programme, which comprises trends watch activities, studies of quantitative and qualitative nature and the dissemination of results to governmental organisations and the public at large. Market intelligence has been central to her education and professional experience over the past 12 years. As an Analyst in the public and private sector, she has gained substantial knowledge about tourism and operational experience with data analysis. As a Researcher and Lecturer, she has refined her knowledge of tourism statistics and quantitative methods of analysis. Through collaboration with international organisations, among which are UNWTO, institutions and expert groups, she has gained solid experience with tourism policy making and management. Valeria Croce can be contacted at: valeria.croce@icloud.com