Forecasting Tourism in Dubai

Economic Note No. 8

Authors
Dr. Nasser Saidi
Dr. Fabio Scacciavillani
Fahad Ali

DIFC
Dubai International Financial Centre

Abstract

During the past decade, Dubai has become a global tourism hub. This paper forecasts guest nights demand in Dubai hotels and hotel apartments as a proxy to measure the performance of the tourist sector in the Emirate. Although tourism represents only 2.5% of recorded GDP, it has a profound impact on retail sales and transportation, which together represent some 44% of recorded GDP. The empirical analysis uses traditional empirical methods, but in addition to the data compiled by Dubai authorities, it tests the explanatory power of seasonal factors and dummy variables capturing erratic effects such as the Holy Month of Ramadan, the school calendar and major sporting events. As a final element in the forecasting exercise, the analysis includes the number of searches conducted on Google using keywords related to travel to Dubai.[1]

Keywords: Dubai tourism, Google Trends, Guest Nights forecast.

1 We would like to thank Jaylyn Garcia from the Dubai Department of Tourism and Commerce Marketing for providing us with statistical data, the Dubai Statistics Centre for their kind assistance and Hal Varian for his precious advice on the use of Google Trends for forecasting purposes. All remaining errors are our own.

Disclaimer:

Disclaimer: This Economic Note should not be reported as representing the official position of the DIFC. The views here expressed are those of the authors and might not coincide with those of the DIFC or any other agency or body of the Dubai Government. Economic Notes focus on research activities and are published to elicit comments and to further debate.

DIFC Economics:

- Dr. Nasser Saidi
  E-mail: nasser.saidi@difc.ae
  Tel: +971 4 362 2550

- Dr. Fabio Scacciavillani
  E-mail: fabio.scacciavillani@difc.ae
  Tel: +971 4 362 2554

- Fahad Ali
  E-mail: fahad.ali@difc.ae
  Tel: +971 4 362 2421

Table of Contents

1. Introduction

Dubai has become an important destination for regional and global tourism, rising to prominence as a top location for shopping, leisure, sporting events, international conferences and media events. In addition to excellent services in the hotel sector, Dubai has built state-of-the-art facilities and tourism infrastructure encompassing convention centers, stadiums, malls, and urban transportation. The tourism sector is a key driver of the Dubai economy, not only for its direct contribution to GDP, which according to official data is a paltry 2.5%, but for its indirect effect on retail sales and transportation (including logistics), which together represent some 44% of GDP. In other words, tourism generates a powerful multiplier effect on the economy, which is depicted synthetically in Fig. 1.

The number of tourists at a global level is also forecast to grow strongly for the foreseeable future, driven by an ageing population in mature countries and higher income levels in emerging markets which will have positive repercussion on Dubai.

Given the significant impact of tourism on the wider Dubai economy, accurate forecasting of tourism demand and its drivers is a fundamental input for key decisions on infrastructure investments, as well as for gauging the conjunctural situation and planning for the demand flow.

Song, Witt and Li (2009) explain that "Tourism demand is the foundation on which all tourism-related business decisions ultimately rest. Companies such as airlines, tour operators, hotels, cruise ship lines, and many recreation facility providers and shop owners are interested in the demand for their products by tourists. The success of many businesses depends largely or totally on the state of tourism demand, and ultimate management failure is quite often due to the failure to meet market demand. Because of the key role of demand as a determinant of business profitability, estimates of expected future demand constitute a very important element in all planning activities. It is clear that accurate forecasts of tourism demand are essential for efficient planning by tourism-related businesses, particularly given the perishability of the tourism product. For example, it is impossible for an airline to recoup the potential revenue lost by a flight taking off with empty seats."

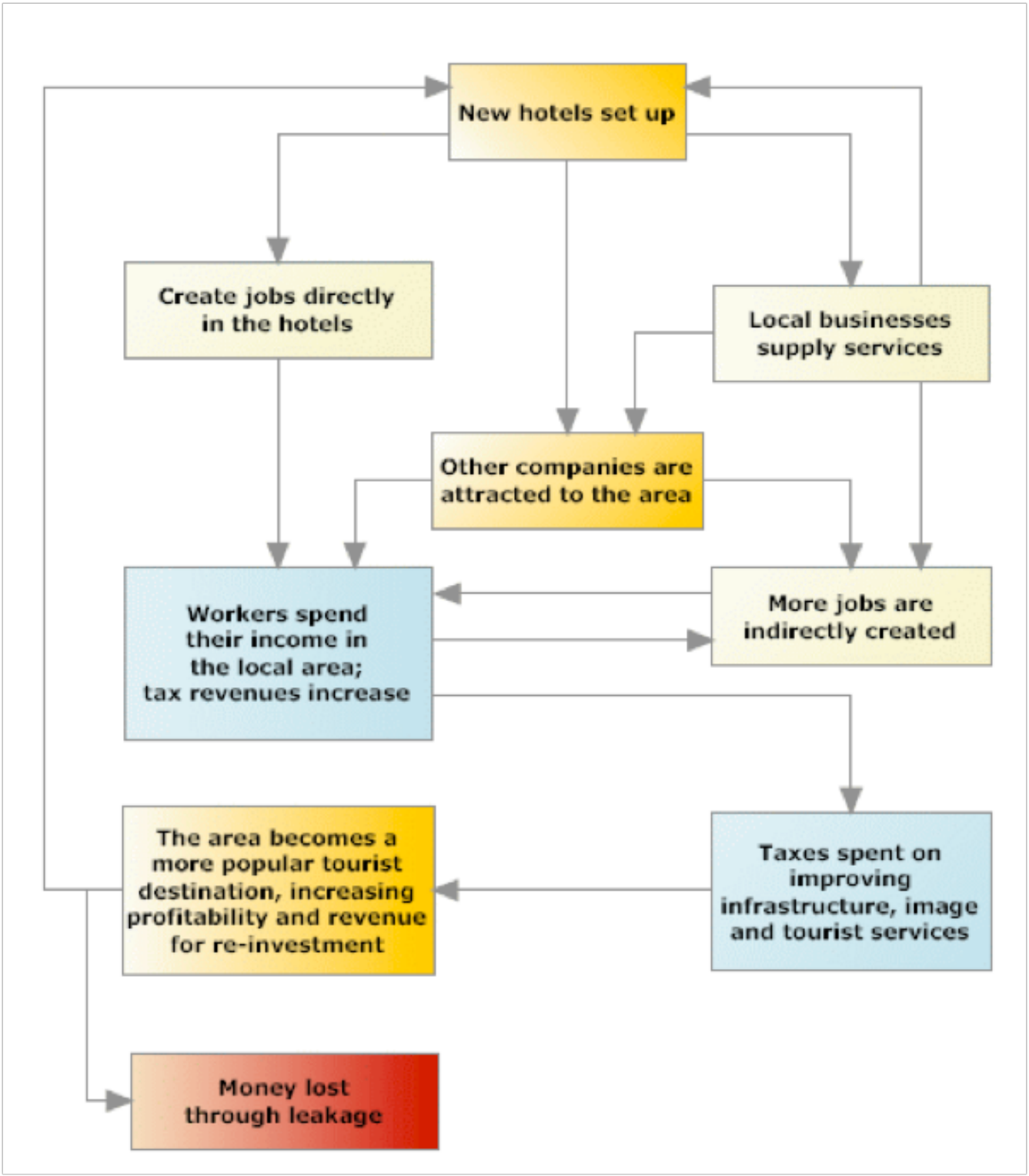
In this paper we conduct a forecasting exercise for the tourist sector in Dubai by focusing on a key indicator of tourism activity as measured by the number of guest nights. We use monthly data from the Dubai Department of Tourism and Commerce Marketing on guest night stays. Forecasting these numbers presents a series of challenges that have been tackled through different methodologies.[2] For the case of Dubai we face some unconventional challenges in terms of data availability, erratic factors and a dynamic economy that has experienced unprecedented growth and structural change, especially in the availability of hotels and recreational facilities.

We also test the predictive power of Google Trends data, which track internet searches based on selected keywords. The rationale is that with the increase in people who use web-based information sources and booking systems to plan their vacations, it is likely that the Google search data might prove useful to predict the number of visitors to Dubai.

The remainder of the paper is divided into three sections. Section 2 contains a description of the data and basic statistics on the Dubai tourism sector. Section 3 illustrates the statistical method and displays the results. Section 4 draws the conclusion and summarizes the findings. Two appendices provide additional details on the data and on Google Trends.

2 For a complete review, which is beyond the scope of this note, we direct the reader to Song, Witt and Li (2009).

Fig. 1 – Multiplier Effects from Tourism



2. Stylized Facts and Data Description

Dubai has become a leading tourist destination in the Middle East thanks to its developed transport system and logistics and accommodation facilities, as well as to its leading role as a conference and sporting events location. Furthermore, it hosts a large expatriate workforce coming from virtually every country in the world. The workforce is internationally mobile and encourages visits from relatives and friends. Finally, the booming service sector results in a derived demand for qualified professional consultants who often combine work assignments with a period of vacation. In short, the demand for tourism services in Dubai is well diversified by geographical source and income level, and often not entirely separated from business travel.

The Dubai Department of Tourism and Commerce Marketing maintains a comprehensive database of monthly series relating to hotel guests and hotel night stays. The Dubai Airport Free Zone Authority (DAFZA) also collects data on arrivals and departures.[3] Equally important are the data on hotel revenues, which provide an approximation of the sector value added.

Chart 1 depicts the monthly series on hotel statistics over the last 10 years. The line refers to the sum of guest nights figures from hotels and hotel apartments, the latter, which usually caters to families and longer-term visitors. Chart 2 shows the data on revenues aggregated by hotel and hotel apartments since 2004, when the data started to be released. Chart 3 shows the graph of total passenger traffic, including those in transit. Obviously not all arrivals relate to tourism, but it is a good proxy to explain the seasonal changes in figures.

The success of the tourism sector in Dubai and the impressive growth of air traffic at its airport over 10 years are summarized in Table 1 and Table 2.

Table 1: Expansion of the Tourism Sector in Dubai

Tourism Indicator	1998	2008	Average of Annual Growth Rate
Number of Operating Establishment	258	350	0.2
Number of Rooms	17,046	40,732	7
Number of Guests	2,184,292	6,273,291	9.9
Number of Guests nights	5,433,215	16,653,704	10.6
Revenues (in AED thousands)	2,063	13,245	20.3
Number of Aircrafts landed in Airport	123,352	270,349	8.3
Total Passengers Movement in Airport	9,732,202	37,441,440	14.5

Source: Yearly Statistical Book published by the Dubai Statistical Center (www.dsc.gov.ae)

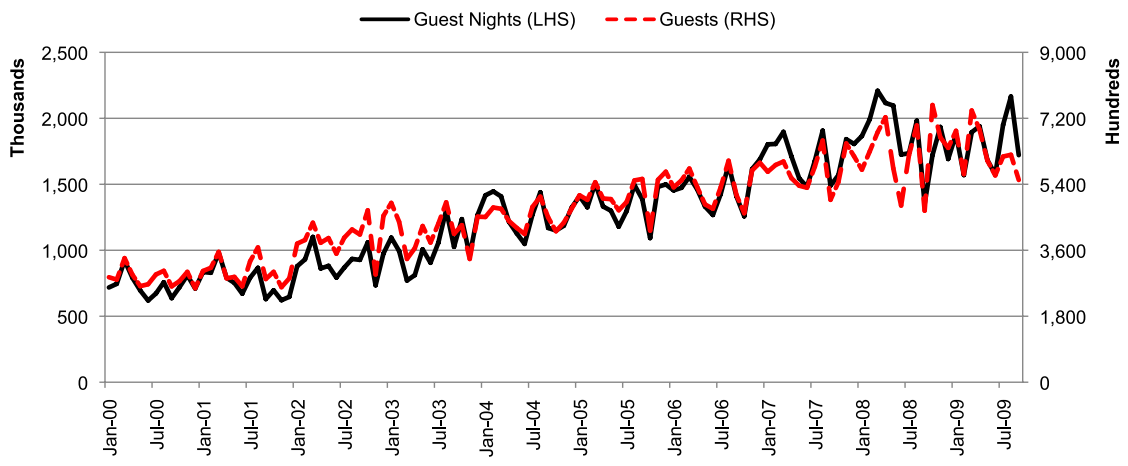
Table 2: Descriptive Statistics of Monthly Guest Nights Series

Statistic	Guest Nights	Guests	Revenues (in AED)	Rooms	Passengers	Buildings
Mean	1,279,697	469,483	895,000,000	42,780	2,152,654	436
Median	1,297,150	473,285	826,000,000	40,666	2,058,635	421
Std. Dev.	426,807	126,978	343,000,000	8,182	773,694	54
Maximum	2,208,599	756,208	1,730,000,000	59,372	3,768,965	534
Minimum	619,186	256,466	379,000,000	32,757	941,920	363
Sample Period	Y00M1-Y09M9	Y00M1-2009M9	Y04M1-Y09M9	Y04M1 - Y09M9	Y01M1 - Y09M10	Y04M1 - Y09M9
Observations	117	117	69	69	106	69

3 Arrivals from the sea are negligible in comparison, and data on land arrivals are not disclosed, but are likely to be minimal.

The graphs in Charts 1-4 clearly illustrate the importance of seasonal factors, which typically characterize the tourist sector, along with a clear upward trend. Apart from the strong summer effect, a number of other factors contribute to the monthly fluctuations, notably Ramadan -- whose timing varies with the Islamic Hijiri lunar calendar – sporting events, and the school calendar. Therefore a major task is to identify these factors and take them into account in building a forecasting model.

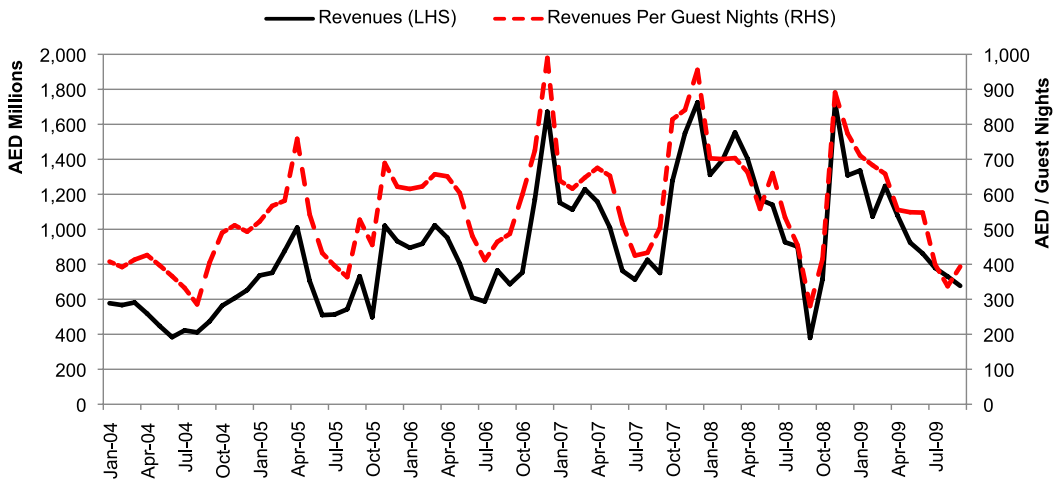
Chart 1: No. of Dubai Guests Nights in Hotels and Hotel Apartments



Source: Dubai Statistical Centre and Dubai Department of Tourism and Commerce Marketing

Hotel revenues on the other hand are distinctly cyclical with a weaker upward trend and a seasonal drop during the summer when the torrid heat discourages vacationers and the competition from destinations in more temperate zones is fiercer. The graph in Chart 2 shows a change around the summer of 2006. Before, the series had a lower average and a lower variance, afterwards the average and variance increased markedly, a structural change likely to be attributed to the increased appeal of Dubai as a tourist destination and the upward shift in the market segment to which the Dubai hotel industry caters.

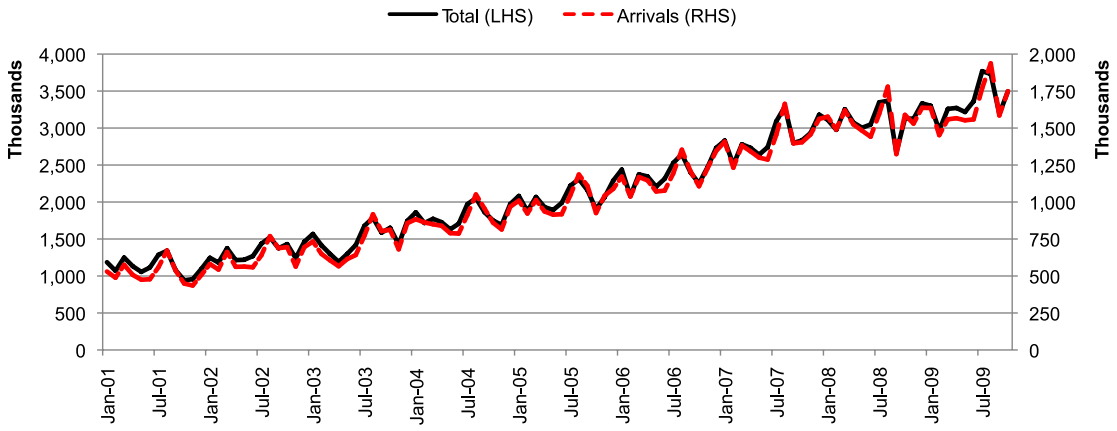
Chart 2: Revenues of Dubai Hotels and Hotel Apartments, Jan 04 – Sep 09



Source: Dubai Statistical Centre and Dubai Department of Tourism and Commerce Marketing

Another measure of the growing role of Dubai within the tourism industry is the airport arrivals (excluding transit passengers, which account for about 2% of the total). In a decade, the number has risen more than threefold (Chart 3).

Chart 3: Passenger Movements at Dubai International Airport, Jan 01 - Dec 09

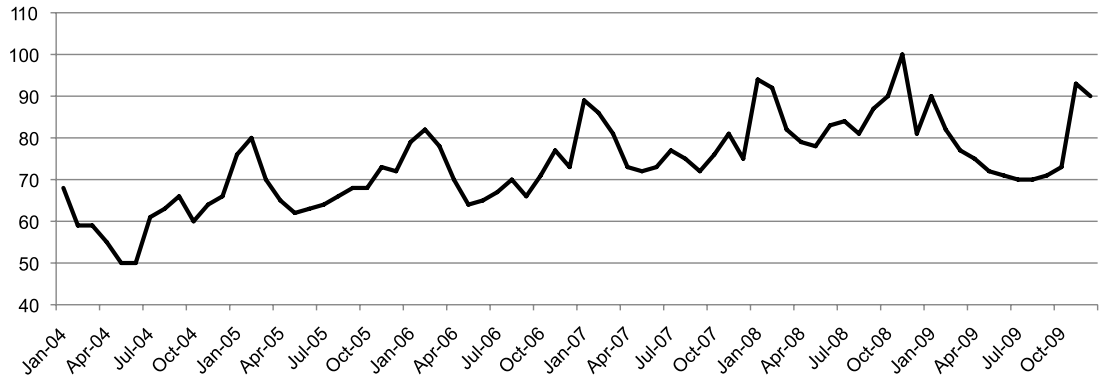


Source: Dubai Statistical Centre

In addition to the data from Dubai authorities, a useful gauge of interest in Dubai travel and tourism is constituted by the number of searches conducted on Google using certain keywords. Google has recently made available a wealth of data on searches done on its engine. By inserting a keyword one can determine how frequently it has been searched worldwide. Additional information can be gathered on the geographical distribution of the searches and the language of the search. Furthermore, some keywords are classified by categories and subcategories to form a coherent framework that can then be compared by country or other criteria. Appendix II contains a description and explanation of the Google Trends data.

Chart 4 displays the data relative to searches conducted under the category Travel with the keyword Dubai. This category is then subdivided into four subcategories for which the search range is also available: Hotels & Accommodations (75-100%), Air Travel (0-10%), Vacation Destinations (0-10%), Attractions & Activities (0-10%)

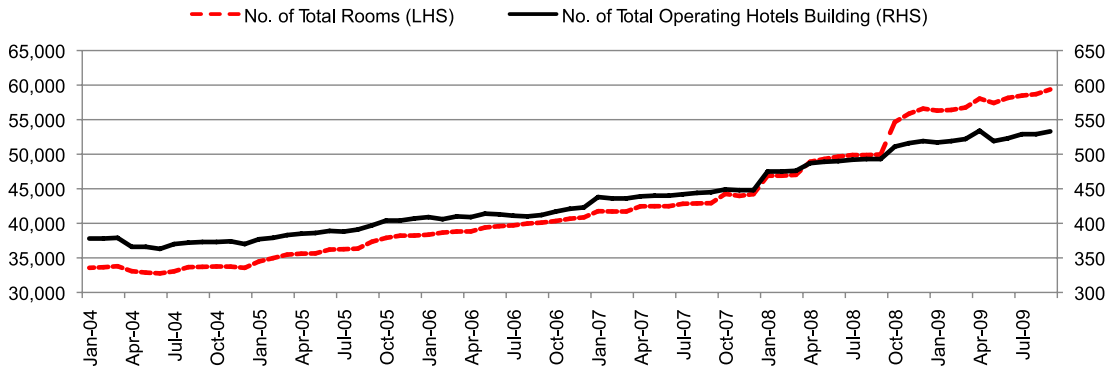
Chart 4: Internet Search for Dubai by Travel Category



Source: Google Insights for Search (www.google.com/insights/search/#)

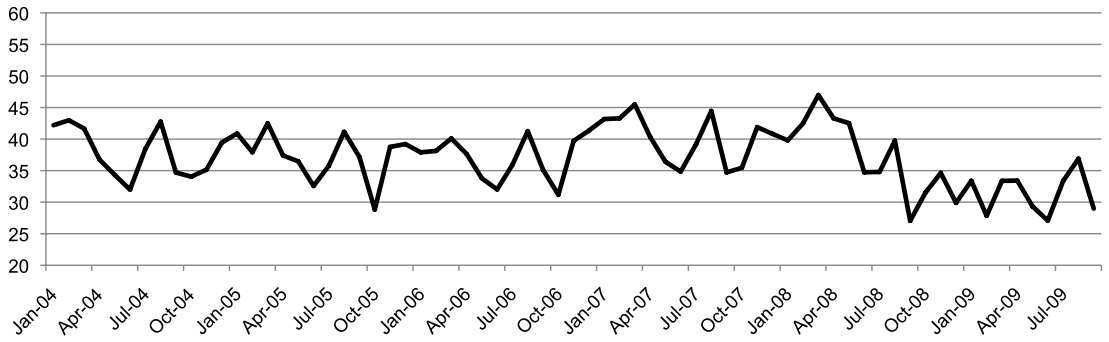
The numbers in the graph reflect how many searches have been conducted for a particular term, relative to the total number of searches done on Google in a given period over the category Travel with keyword Dubai. Hence the data do not represent absolute search volume numbers, because the data are normalized and presented on a scale from 0 to 100; in other words, each monthly figure is divided by the highest point in the sample period, which is then set equal to 100. This normalization is necessary to take into account scale effects deriving from the increase in internet users and to allow a comparison between geographical areas with different population sizes and internet penetration rates.

Chart 5 - Hotels and Apartments Capacity



Source: Dubai Department of Tourism and Commerce Marketing

Chart 6 - Hotels and Apartment Occupancy Rate (%)



Source: Dubai Department of Tourism and Commerce Marketing

3. Forecasting Model

We proceed to forecast the total numbers of guest nights starting from a univariate ARIMA model. We then compare the results to an OLS where we introduce seasonal and erratic factors and then probe the usefulness of Google search data as a leading indicator of tourist arrivals.

3.1 ARIMA model

In view of the strong upward trend in the number of guest nights recorded over the years, we focus on the growth rate of this variable. We start our exercise with the standard univariate ARIMA forecasting model following the classic methodology proposed by Box and Jenkins. Based on conventional diagnostic statistics for goodness of fit, such as R-square, Akaike Information Criterion and Schwartz Information Criterion, the most appropriate model to describe the data is an MA (2), with a seasonal component (Equation 1). The estimates are reported in Table 3.

Equation 1: ARIMA Model for Guest Nights

$$d(\log(GuestNight_t)) = C + \beta\varepsilon_{t-1} + \gamma\varepsilon_{t-2} + \delta\varepsilon_{t-12}$$

Table 3: Univariate ARIMA Parameter Estimates for Guest Nights

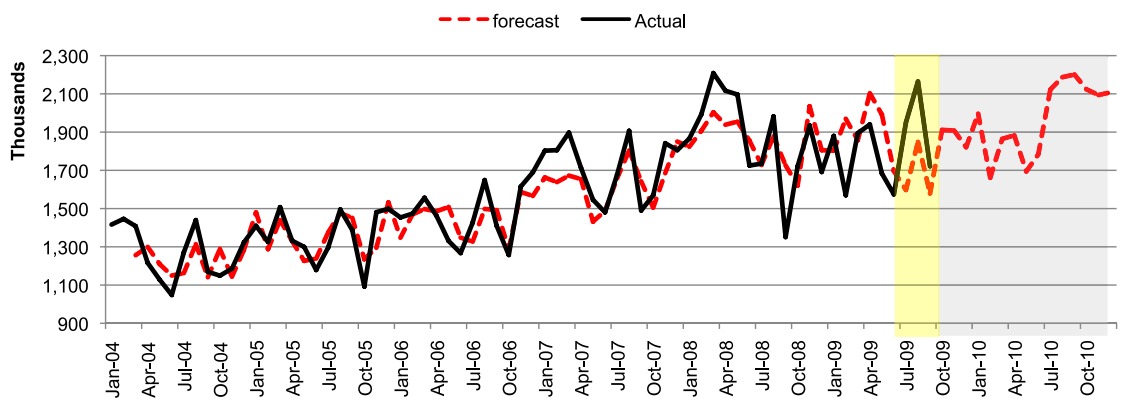
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.006705	0.001492	4.493435	0.0000
MA(1)	-0.553237	0.112439	-4.920349	0.0000
MA(2)	-0.421351	0.111335	-3.784528	0.0003
SMA(12)	0.854543	0.031980	26.72111	0.0000

R-squared	0.637772	Mean dependent var	0.002862
Adjusted R-squared	0.620792	S.D. dependent var	0.134350
S.E. of regression	0.082732	Akaike info criterion	-2.089390
Sum squared resid	0.438057	Schwarz criterion	-1.958831
Log likelihood	75.03927	Hannan-Quinn criter.	-2.037659
F-statistic	37.56139	Durbin-Watson stat	1.966231
Prob(F-statistic)	0.000000	Mean Absolute Percent Error	6.649684

The in-sample and out-of-sample forecasted values are depicted in Chart 7. The yellow area shows the out-of-sample forecasts over the period Jul-Sep 2009 which can be compared with actual values. The grey area shows the out-of-sample forecasts over the period Oct 09-Dec 10 i.e. beyond the sample horizon. Overall the fit is acceptable, but not extremely precise, with a Mean Absolute Percent Error of the in-sample forecast of 6.65 (Table 3). The profile of the forecast matches the pattern of the actual data, but with a major drawback: the size of the rebound over the summer of 2009 is underestimated.

This ARIMA exercise was conducted mostly to provide a benchmark against which to evaluate the alternative approach in the next subsections.

Chart 7: Actual & Forecasted Guest Nights with ARIMA



Note: The yellow area depicts out-of-sample forecasts over Jul-Sep 09 and actual data; the grey area depicts forecasts beyond the horizon of sample data.

3.2 Seasonal and erratic factors

Seasonality is a major factor influencing tourism worldwide, and our region is no exception. However for Dubai, other factors such as Ramadan and sporting events have a distinct impact.

In this section we identify the major seasonal and erratic factors with the aim to construct a number of dummy variables to capture their effects. As an example, Table 4 provides a summary on the number and classification of major events as recorded by the Dubai Department of Tourism and Commerce Marketing in its Dubai Calendar in each month of March since January 2004. A detailed description of each factor and its relevance to the forecast exercise is provided below.

3.2.1 Sports Factor

Sporting events turned out to be one of the biggest influencers on tourism, in particular the Dubai Tennis APT Tournament (which also gives rise to the highest peak in Google searches related to Dubai). Traditionally the tournament starts in the last days of February and lasts until the first Saturday of March (the exception was in 2009 when it took place in mid March). Another major event, the Dubai World Cup – the horse race with the highest prize in the world – takes place in the last week of March. For this reason we have inserted a Sports Dummy which is 1 in March and zero for all other months. The main sports events, ranked by the number of Google searches, are listed in Table 5.

Table 4: Dubai Calendar of Events in March

Time/event	Conference & exhibition	Cultural & festival	Sport	No. of events
March - 04	13	0	8	21
March - 05	29	0	12	41
March - 06	23	3	13	39
March - 07	24	2	1	27
March - 08	31	6	4	41

Source: Dubai Department of Tourism and Commerce Marketing

Table 5: Top Search Terms

Rank	Search Terms	Index
1	Dubai tennis	100
2	Dubai golf	90
3	Dubai classic (snooker)	55
4	Dubai open	40
5	Dubai rugby	35
6	Dubai desert classic (golf)	35
7	Dubai Sevens (rugby)	30
8	Dubai cup (horse race)	39
9	Atp Dubai (Tennis)	30

Source: Google Insights for Search (www.google.com/insights/search/#)

3.2.2 Islamic Calendar Factor

The share of hotel guests from predominantly Muslim countries in Dubai was roughly 60% for the period 2004-2009; hence it is to be expected that events related to the Hijri calendar affect travel. The Holy Month of Ramadan (which in relation to the Gregorian calendar shifts every year by two weeks) marks the low season for traveling. The sharp drop in guest nights in October between 2004 and 2006 and September afterwards, justifies the insertion of a Ramadan dummy.

3.2.3 School Exam Factor

Final school exams take place usually during June, so families with school age children tend to avoid traveling in that month. Thus we inserted an Exam dummy for the month of June.

3.2.4 Summer Factor

Over the northern hemisphere, summer, especially August, is peak season for holidays, not least because schools are closed. Furthermore, Dubai holds the Summer Surprises festival, which essentially entails large discounts by retailers, as well as extensive family-focused entertainment across the city's malls, thereby attracting shoppers from a host of countries (the retail sector accounts for over 30% of Dubai's GDP). To capture this effect, we add a Summer dummy.

To highlight the effect of all these factors, in Chart 8 we plot their occurrence and the graph of guest nights over Jan 2004 – Sep 2009. Even from a cursory visual inspection, it is clear that each of the factors has a notable impact that an ARIMA would not entirely capture.

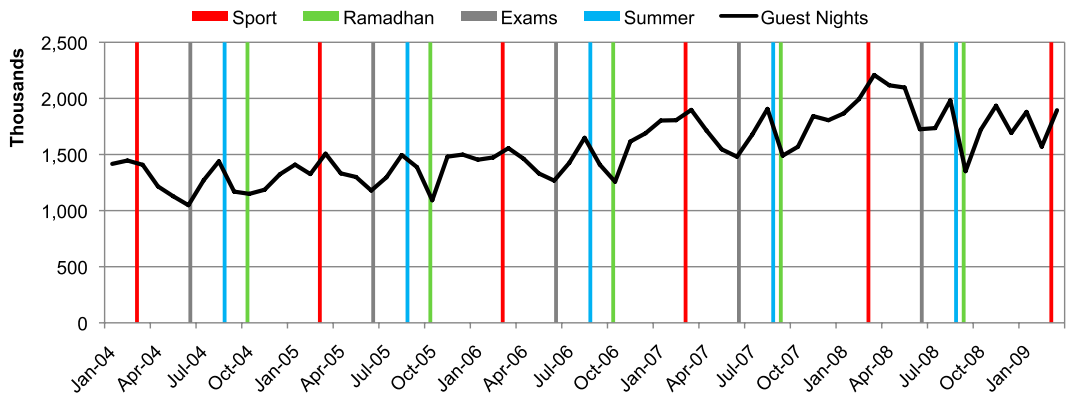
We run an OLS regression with data in logarithms, including the various seasonal factors, the lagged dependent variable and a linear trend to capture the growth factor:

Equation 2: OLS Model with Seasonal and Erratic Factors

LOG(GUESTNIGHT) = C(1) + C(2)*SUMMER + C(3)*SPORT + C(4)*EXAM + C(5)*RAMADAN + C(6)*TREND + C(7)*LOG(GUESTNIGHT(-1))

The results of the estimation are presented in Table 6 and the plot of the forecast, both in sample and out of sample, in Chart 9, where again the shaded area contains the out-of-sample forecasts.

Chart 8: Seasonal & Erratic Factors' Influence on Guest Nights



Source: Dubai Statistical Centre and Dubai Department of Tourism and Commerce Marketing

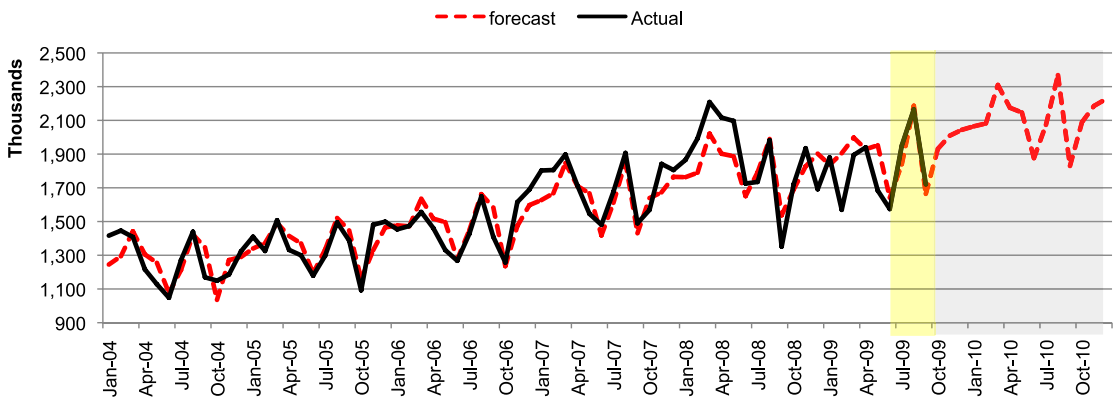
Table 6: Estimated Parameters & Tests for OLS Model in Eq. 2

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	9.557290	1.105923	8.641910	0.0000
C(2)	0.097523	0.032682	2.983982	0.0041
C(3)	0.096697	0.032797	2.948356	0.0045
C(4)	-0.135925	0.032797	-4.144436	0.0001
C(5)	-0.205265	0.032748	-6.267938	0.0000
C(6)	0.005023	0.000731	6.874745	0.0000
C(7)	0.301472	0.080869	3.727900	0.0004

R-squared	0.843112	Mean dependent var	14.25147
Adjusted R-squared	0.827930	S.D. dependent var	0.180788
S.E. of regression	0.074993	Akaike info criterion	-2.246911
Sum squared resid	0.348687	Schwarz criterion	-2.020263
Log likelihood	84.51844	Hannan-Quinn criter.	-2.156992
F-statistic	55.53123	Durbin-Watson stat	1.583856
Prob(F-statistic)	0.000000	Mean Absolute Percent Error	6.482477

The main improvement over the ARIMA model is the lower mean absolute percent error and the closer capture of turning points and peaks. All variables are highly significant and the turnaround in the summer of 2009 is predicted with remarkable precision. Exams and Ramadan have the expected dampening effect on tourism, whereas sport and summer have the corresponding positive effect.

Chart 9: Actual & Forecasted Guest Nights with OLS Model



Note: The yellow area depicts out of sample forecasts over Jul-Sep 09 and actual data; the grey area depicts forecasts beyond the sample horizon

3.3 Combining dummy variables and Google data

Data on tourism are released with a lag and therefore are of limited use in gauging the current state of the sector and to make forecasts for the months ahead. In other words, one has to wait months for the data to be released and in the meantime it is hard for hotel managers and other hospitality outfits to assess whether demand is buoyant or stagnant. By contrast, Google data on search volumes are available at weekly frequency and with minimal delay. Therefore, it would be highly useful if the Google data had some predictive properties over guest nights[4].

We collected monthly Google trends data and ran an augmented OLS regression using the same seasonal and erratic factors dummies as in Equation 2 but adding Google search data (Equation 3). Table 7 reports the results, while Chart 10 shows the in-sample and out-of-sample forecasts. To obtain the out-of-sample forecasts beyond Sep 2009, we have inserted the forecasted data provided by Google trends on keywords searches up to end 2010.

Equation 3: OLS Regression with Google Trend Data

LOG(GUESTNIGHT)=C(1)+C(2)*LOG(GUESTNIGHT(-1)) + C(3)*SPORT + C(4)*RAMADAN + C(5)*EXAM + C(6)*SUMMER + C(7)*LOG(GOOGLE) + C(8)*TREND

The overall fit improves noticeably (as demonstrated by the lower Mean Absolute Percent Error) and in 2010 they indicate healthy growth but high volatility, even more pronounced than in 2009.

Table 7: OLS Regression with Google Data & Dummies

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	9.737523	1.037892	9.382018	0.0000
C(2)	0.275425	0.076240	3.612625	0.0006
C(3)	0.095611	0.030733	3.111017	0.0028
C(4)	-0.187210	0.031233	-5.993970	0.0000
C(5)	-0.109094	0.031926	-3.417079	0.0011
C(6)	0.114975	0.031136	3.692597	0.0005
C(7)	0.003768	0.001215	3.100924	0.0029
C(8)	0.003923	0.000771	5.088771	0.0000

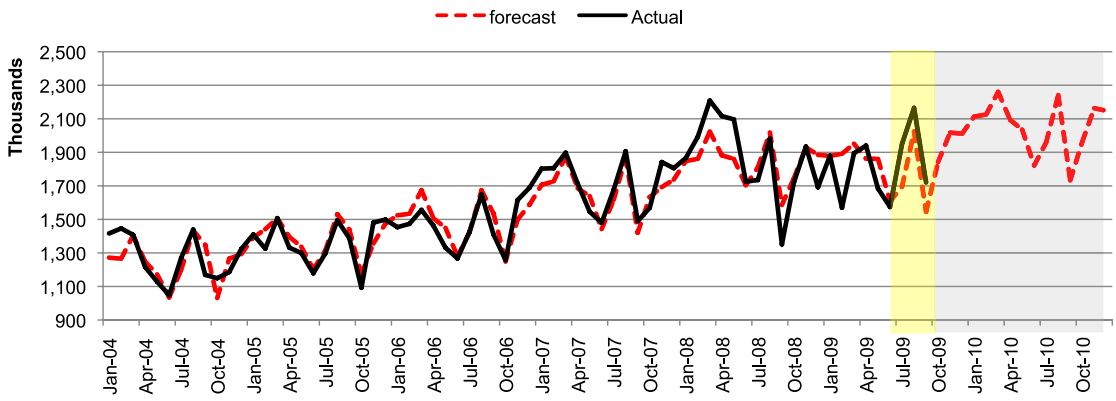
R-squared	0.864476	Mean dependent var	14.25147
Adjusted R-squared	0.848924	S.D. dependent var	0.180788
S.E. of regression	0.070270	Akaike info criterion	-2.364305
Sum squared resid	0.301206	Schwarz criterion	-2.105278
Log likelihood	89.56852	Hannan-Quinn criter.	-2.261540
F-statistic	55.58628	Durbin-Watson stat	1.559221
Prob(F-statistic)	0.000000	Mean Absolute Percent Error	5.975611

4 The most common statistical test to verify whether a variable might help to predict another is the Granger Causality Test. It is well known that the name is not a correct description of its properties, because causality cannot be ascertained: Granger causality detects precedence and information content of one variable with respect to the other. The Table below reports the results of the test which indeed confirms that Google searches included in the category Dubai Travel, Granger causes Tourism Guest Nights in Dubai.

Pair-wise Granger Causality Tests

Null Hypothesis:	F-Statistic	Prob.
Google does not Granger Cause NIT_GST		0.0029
NIT_GST does not Granger Cause TRAVEL_M	0.31953	0.7279

Chart 10: Actual & Forecasted of Guest Nights with OLS with G-Trends Model



Note: The yellow area depicts out-of-sample forecasts over Jul-Sep 09 and actual data; the grey area depicts out-of-sample forecasts beyond the sample horizon.

At this point we can compare out-of-sample forecasts at different horizons obtained with each of the three models to provide a complete picture. Table 8 displays the MSE of these forecasts.

All three methods are rather robust. Considering that the forecasting period has been one of the most turbulent in recent economic history a MSE of about 60,000 units for a monthly figure in the order of 2 million guest nights is reasonably accurate for most practical purposes. It is likely that once the situation stabilizes the forecasting power of these models will improve further.

All in all the addition of Google Trend data on Dubai Travel is not particularly useful especially at shorter horizons (using lagged Google Trend data leads to even worse results[5]). This is somewhat puzzling but we interpret this outcome as an indication that people conduct searches after they have decided to travel to Dubai to look for information on their destination, e.g. venues, transportation, restaurants, and in any case they do not search Google much in advance. In practice Google data are useful because they are disclosed at weekly intervals and in a timely fashion.

Table 8: Comparison among Mean Squared Errors for out-of-Sample Forecast

Period	ARIMA	OLS	OLS with G-Tend
3 Months	222,821	42,674	134,732
6 Months	96,263	56,551	69,033
9 Months	89,332	75,456	69,033
12 Months	59,663	69,091	69,033

5 The results of these regressions are not reported here, but are available on request from the authors.

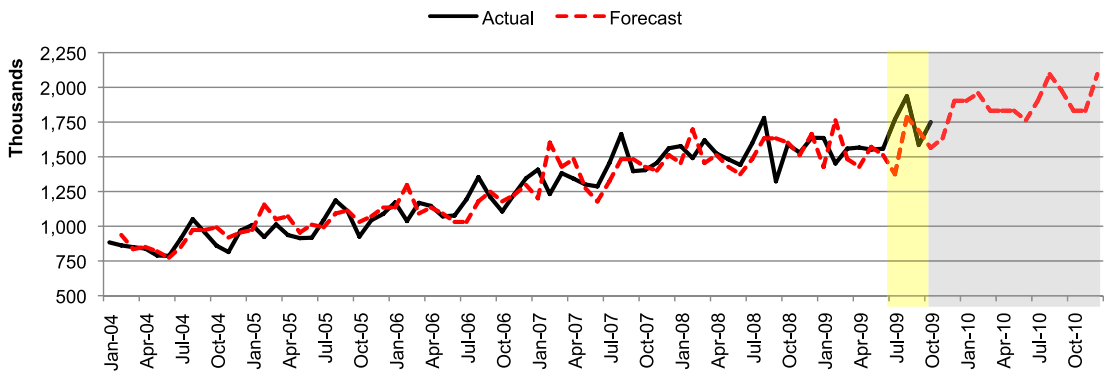
We wondered if there was some subsection of Dubai Travel data that could display a better predictive power. Given the scale of online air ticket sales worldwide we postulated that Google searches on air travel could bear a relation with the arrivals at the Dubai Airport the next month. Table 9 and Chart 11 show that the hypothesis was correct and actually the fit is extremely good for such a simple regression. Essentially passenger traffic can be broadly gauged a month in advance through Google data. In turn, arrivals are also very much correlated with guest nights (about 80%), so in general Google Air Travel series does help to produce a reliable forecast of guest nights.

Table 9 – OLS Regression with Dubai Air Travel Searches

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	12.99081	0.059556	218.1283	0.0000
C(2)	0.019002	0.001073	17.70513	0.0000

R-squared	0.826074	Mean dependent var	14.02362
Adjusted R-squared	0.823439	S.D. dependent var	0.235532
S.E. of regression	0.098969	Akaike info criterion	-1.759054
Sum squared resid	0.646458	Schwarz criterion	-1.693775
Log likelihood	61.80785	Hannan-Quinn criter.	-1.733188
F-statistic	313.4717	Durbin-Watson stat	1.995712
Prob(F-statistic)	0.000000	Mean Absolute Percent Error	7.554

Chart 11: Actual & Forecasted Arrivals at Dubai Airport



Note: The yellow area depicts out-of-sample forecasts over Jul-Sep 09 and actual data. Beyond September the forecasts are based on the predicted values for Dubai Air Travel calculated by Google.

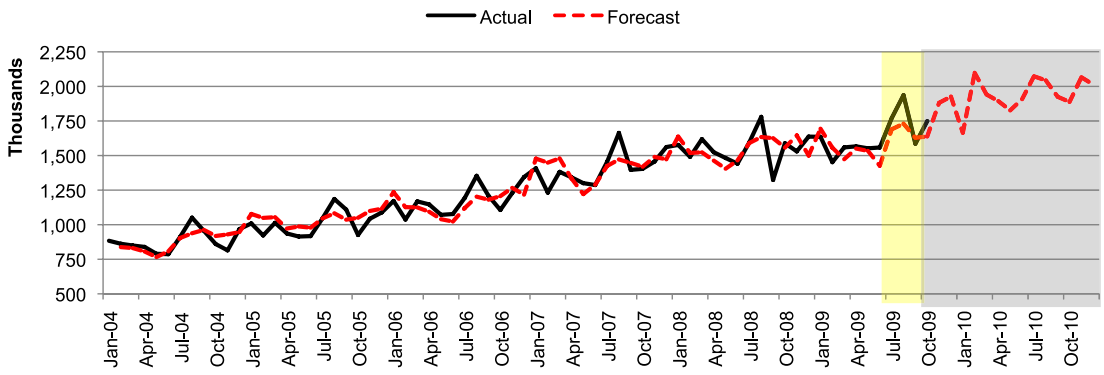
Adding on the right hand side the lagged arrivals yields an even better fit, as presented in Table 10 and in Chart 12, with Air Travel searches retaining their informational content.

Table 10 – OLS with Dubai Air Travel Searches and Lagged Arrivals

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	8.722488	1.091009	7.994883	0.0000
C(2)	0.013769	0.001775	7.759017	0.0000
C(3)	0.324515	0.084066	3.860259	0.0003

R-squared	0.913995	Mean dependent var	14.02390
Adjusted R-squared	0.911427	S.D. dependent var	0.239268
S.E. of regression	0.071209	Akaike info criterion	-2.404489
Sum squared resid	0.339736	Schwarz criterion	-2.308125
Log likelihood	87.15712	Hannan-Quinn criter.	-2.366212
F-statistic	356.0109	Durbin-Watson stat	2.347276
Prob(F-statistic)	0.000000		

Chart 12: Actual & Forecasted Arrivals at Dubai Airport



Note: The yellow area depicts out-of-sample forecasts over Jul-Sep 09 and actual data. Beyond September the forecasts are based on the predicted values for Dubai Air Travel calculated by Google.

4. Conclusions

Tourism is one of the main drivers of Dubai's economy in terms of GDP and even more in terms of employment share. This paper has presented a set of forecasting exercises on key variables related to the tourist sector in Dubai. We started from a simple univariate ARIMA model, which in itself provides a reasonable fit. However for practical purposes, e.g. for a hotel manager that needs to plan for the months ahead, an ARIMA model is not very useful because the data required for this forecast are released with a substantial delay.

In order to improve the fit and strengthen the practical relevance of the forecasting exercise we tried an alternative methodology. We proceeded along two lines: a) we conducted a more detailed analysis of seasonal factors, such as the school calendar or sporting events, and a non-seasonal but time-related factor, i.e. Ramadan; b) we included data from Google Trends that are related to internet keyword searches collected under the heading Dubai Travel and at a finer resolution under Dubai Air Travel.

Simple OLS regressions with seasonal and erratic factors have proven to be more accurate than ARIMA forecasts. Google Dubai Travel data (which are available at weekly frequency), are helpful for improving the accuracy of the long-term forecast for guest nights, while at shorter horizon their usefulness is debatable. However Google data on Air Travel searches (i.e. by people buying air tickets on line) are extremely useful to forecast the arrivals at the Dubai airport, which in turn are highly correlated with guest nights.

Appendix I: Dataset

Table 10 reports the metadata and the source for the series used in this paper. The series are reported below and they can be downloaded from

http://www.dubaitourism.ae/EServices/Statistics/HotelStatistics/tabid/167/language/en-US/Default.aspx.

Table A1: Metadata

Variable	Description	Period	Source
Guest Nights	Number of Dubai Hotels and Hotel Apartments Guest Nights	(Y00M01,Y09M09)	Dubai Tourism
Guests	Number of Dubai Hotels and Hotel Apartments Guests	(Y00M01,Y09M09)	Dubai Tourism
Google	Google Trends Data for Dubai search term categorized by travel	(Y04M01,Y09M12)	Google Trends
Ramadan	Dummy that takes the value 1 during Ramadan and 0 otherwise	(Y04M01,Y09M12)	Hijri Calendar
Summer	Dummy that takes the value 1 in August and 0 otherwise	(Y04M01,Y09M12)	Gregorian Calendar
Exam	Dummy that takes the value 1 in June and 0 otherwise	(Y04M01,Y09M12)	Gregorian Calendar
Sport	Dummy that takes the value 1 in March and 0 otherwise	(Y04M01,Y09M12)	Gregorian Calendar

Table A2: Guests Series

Month/Year	2004	2005	2006	2007	2008	2009
Jan	450,805	510,028	531,697	574,180	579,743	685,913
Feb	476,673	497,485	550,751	592,741	629,707	567,760
Mar	473,174	545,707	582,987	602,175	681,803	741,345
Apr	439,885	501,206	532,705	557,127	722,875	688,882
May	420,916	499,534	485,130	536,201	583,726	604,726
Jun	404,255	469,133	473,285	530,524	481,938	564,116
Jul	478,256	491,612	531,862	588,242	619,488	615,695
Aug	506,302	550,529	604,331	659,333	700,590	620,533
Sep	449,506	554,547	511,257	497,603	467,938	551,733
Oct	411,825	413,651	461,067	545,839	756,208	N/A
Nov	435,373	552,220	577,886	651,413	667,953	N/A
Dec	473,754	574,351	598,712	616,420	639,330	N/A

Table A3: Guest Nights Series

Month/Year	2004	2005	2006	2007	2008	2009
Jan	1,416,447	1,409,742	1,453,615	1,802,708	1,865,791	1,879,874
Feb	1,446,076	1,325,581	1,473,492	1,805,480	1,992,722	1,569,767
Mar	1,408,038	1,507,237	1,556,496	1,897,442	2,208,599	1,894,538
Apr	1,215,782	1,331,861	1,461,710	1,713,387	2,116,448	1,939,924
May	1,127,886	1,299,520	1,330,835	1,546,077	2,096,553	1,683,684
Jun	1,048,079	1,178,763	1,267,320	1,480,200	1,725,074	1,574,168
Jul	1,270,327	1,297,150	1,426,410	1,677,183	1,735,125	1,950,590
Aug	1,439,500	1,495,121	1,648,921	1,906,534	1,982,078	2,165,122
Sep	1,169,282	1,387,288	1,409,082	1,489,906	1,351,242	1,720,755
Oct	1,149,235	1,092,219	1,257,339	1,569,151	1,721,380	N/A
Nov	1,185,995	1,480,965	1,615,329	1,841,783	1,934,525	N/A
Dec	1,323,440	1,499,040	1,689,477	1,805,624	1,691,530	N/A

Table A4: Google Series

Month/Year	2004	2005	2006	2007	2008	2009
Jan	68	76	79	89	94	90
Feb	59	80	82	86	92	82
Mar	59	70	78	81	82	77
Apr	55	65	70	73	79	75
May	50	62	64	72	78	72
Jun	50	63	65	73	83	71
Jul	61	64	67	77	84	70
Aug	63	66	70	75	81	70
Sep	66	68	66	72	87	71
Oct	60	68	71	76	90	73
Nov	64	73	77	81	100	93
Dec	66	72	73	75	81	90

Appendix II: Google Trends

This appendix reports the illustration of Google search data contained in the Google Trend website

1.1 What is a category?

Category refers to a classification of industries or markets, which are commonly referred to as verticals. For example, the Entertainment/Music category may include music genres, recording artists, recordings, performances, instruments, and music-related merchandise, and the Health category may include health resources, education, products, and health-related services. Please note that at this time, not all languages support the category feature. When you select a particular category, the data for your search term will be restricted to that category. For example, if you select the Travel category for the search term cruise, the data you see will be restricted to that specific category. If you don't choose a category, then the data would be comprised across all categories.

1.2 What do the percentages next to the category suggestions refer to?

When you enter in a search term, Insights for Search automatically determines which categories that term may fall under; the range represents the percentage of searches containing that term, which are classified into a particular category. If you've selected a category beforehand, you may see a subcategory breakdown instead. Let's look at an example. Suppose you entered car as your search term, without selecting a category. The categories suggested would likely include Automotive (25-50%) and Travel (10-25%). Now, suppose you entered car again, but selected Automotive as the category. You would then see subcategories suggested instead, such as Vehicle Shopping (10-25%) and Auto Parts (5-10%).

1.3 Does the category breakdown always add up to 100%?

No, the sum of the ranges may not always total 100% because when there are multiple categories, the breakdown is shown in ranges (example: Entertainment: 25-50%, Shopping: 10-25%, Health: 25-50%). Moreover, not all categories are shown.

2.1 How is the data derived?

Google Insights for Search analyzes a portion of Google web searches to compute how many searches have been done for the terms you've entered, relative to the total number of searches done on Google over time. This analysis indicates the likelihood of a random user to search for a particular search term from a certain location at a certain time. Our system also eliminates repeated queries from a single user over a short period of time, so that the level of interest isn't artificially impacted by these types of queries.

Say you've entered the search term tea, setting your location parameter to Scotland, and your time parameter to March 2007. In order to calculate the popularity of this term among users in Scotland in March of 2007, Insights for Search examines a percentage of all searches for tea within the same time and location parameters. The results are then shown on a graph, plotted on a scale from 0 to 100. The same information is also displayed graphically by the geographic heat map.

2.2 Are the data normalized?

Yes. All the results in Google Insights for Search are normalized, which means that we've divided the sets of data by a common variable to cancel out the variable's effect on the data. Doing so allows the underlying characteristics of the data sets to be compared. If we didn't normalize the results and displayed the absolute rankings instead, data from regions generating the most search volume would always be ranked high.

Let's consider the following examples to highlight some of the key points of normalization:

*Canada and Fiji show the same percentages for the term 'hotel.' Does this mean that they have the same amount of search volume for that term?
Just because two regions show the same percentage for a particular term doesn't mean that their absolute search volumes are the same. Data from these two regions -- with significant differences in search volumes -- can be compared equally because the data has been normalized by the total traffic from each respective region. So, we can assume that users in both Fiji and Canada are equally likely to search for the term hotel.
*New York doesn't appear on the list for the term 'haircut.' Does this mean that people in New York don't search for this term at all?

Remember, Google Insights for Search shows the likelihood of users in a particular area to search for a term on Google on a relative basis. So, just because New York isn't on the top regions list for haircut doesn't necessarily mean that people there don't search for that term at all. Consider the following scenarios. It could be that people in New York:

- Don't use Google to find a barber or hair salon
- Use a different term for haircut-related searches
- Search for so many other topics unrelated to haircuts, that searches for haircut comprise a small portion of the search volume from New York as compared to other regions

2.3 How are data scaled?

The data is displayed on a scale of 0 to 100, after normalization; each point on the graph has been divided by the highest point, or 100.

For example, let's suppose that interest for the term skiing surged in the month of November in Sweden. Our system designates that peak as 100. Now let's suppose that interest decreased significantly in December, where the next highest peak was approximately half of what it was in November. That peak would then be designated as 50, and so on.

References

R. Carter Hill, William E. Griffiths and George G. Judge, "Undergraduate Econometrics", Second Edition

Haiyan Song, Stephen F. Witt, Gang Li "The Advanced Econometrics of Tourism Demand", Taylor & Francis, 2009

Hyunyoung Choi and Hal Varian, "Predicting the Present with Google Trends", Google Inc., Draft Date April 10, 2009

"Dubai: Calendar of Events", Department of Tourism and Commerce Marketing, Government of Dubai

"Hotel Statistics Summary 1998-2007", Department of Tourism and Commerce Marketing, Government of Dubai

"Hotel Statistics Reports", Department of Tourism and Commerce Marketing, Government of Dubai

"2008 Statistical Yearbook – Emirate of Dubai", Dubai Statistics Centre, Government of Dubai

Dubai International Financial Centre (DIFC)
The Gate, Level 14
P.O.Box 74777, Dubai, UAE
Tel: +971 4 363 2222
Fax: +971 4 362 2333
Email: info@difc.ae
Web: www.difc.ae