Timing matters: Crisis severity and occupancy rate forecasts in social unrest periods

Dr. Richard T. R. Qiu

Department of Integrated Resort and Tourism Management,

Faculty of Business Administration, University of Macau

Taipa, Macao, China

Email: richardqiu@um.edu.mo

Dr. Anyu Liu

(Corresponding author)

School of Hospitality and Tourism Management, University of Surrey,

Guildford, Surrey, GU2 7XH, United Kingdom

Email: anyu.liu@surrey.ac.uk

Dr. Jason L. Stienmetz

Department of Tourism and Service Management, Modul University Vienna

Am Kahlenberg 1, 1190 Wien, Austria

Email: jason.stienmetz@modul.ac.at

Yang Yu

Department of Information Systems,

College of Business, City University of Hong Kong

Kowloon Tong, Hong Kong, China

Email: yyu282-c@my.cityu.edu.hk

Purpose

The impact of demand fluctuation during crisis events is crucial to the dynamic pricing and revenue management tactics of the hospitality industry. The aim of this paper is to improve the accuracy of hotel demand forecast during periods of crisis or volatility, taking the 2019 social unrest in Hong Kong as an example.

Methodology

Crisis severity, approximated by social media data, is combined with traditional time-series models, including SARIMA, ETS and STL models. Models with and without the crisis severity intervention are evaluated to determine under which conditions a crisis severity measurement improves hotel demand forecasting accuracy.

Findings

Crisis severity is found to be an effective tool to improve the forecasting accuracy of hotel demand during crisis. When the market is volatile, the model with the severity measurement is more effective to reduce the forecasting error. When the time of the crisis lasts long enough for the time series model to capture the change, the performance of traditional time series model is much improved. The findings of this research is the incorporating social media data does not universally improve the forecast accuracy. Hotels should select forecasting models accordingly during crises.

Originality

The originalities of the study are as follows. First, this is the first study to forecast hotel demand during a crisis which has valuable implications for the hospitality industry. Second, this is also the first attempt to introduce a crisis severity measurement, approximated by social media coverage, into the hotel demand forecasting practice thereby extending the application of big data in the hospitality literature.

Keywords: Hotel Demand Forecast, Crisis Severity, Social Media Data, Hong Kong Social Unrest, Time Series Model, Forecast Combination

1. Introduction

Policy makers and managers of the hospitality and tourism sectors are increasingly concerned over the frequency and intensity with which disruptive social unrest events are occurring. While safety, peace and stability are understood to influence tourists' destination choice (Qiu *et al.*, 2018) and to stimulate tourism demand (Pratt & Liu, 2016), "shock" events and crisis, whether environmental, economic, or socio-political in nature, cause volatility and instability to tourism systems, which in turn increase perceived tourism risk and negatively impact destination image and demand. As such, in order to develop tourism systems and management strategies which are more resilient to periods of unrest and volatility, reliable demand forecasting becomes an essential tool upon which scenario planning and resource allocation can be based. To date, several key studies on crisis and tourism demand forecasting have been undertaken at the destination level, such as Blake and Sinclair (2003) for 9/11 terrorism attack in the US, Smeral (2010) for the global financial crisis and key outbound source markets in the world, and Page *et al.* (2012) for the global financial crisis and swine flu in the UK. However, more work is still needed which focuses on the hotel sector specifically. In particular, fluctuation in demand during crisis has significant implications on the dynamic pricing and revenue management tactics for the hospitality industry. Therefore, this research sets out to improve the forecasting of hotel room demand during periods of social unrest or volatility.

One of the challenges to predicting demand during a crisis is limited data availability. In traditional forecasting practice with low frequency data (e.g., annual or quarterly data), it is often difficult to quantify crisis events that are relatively short in duration. Prolonged crisis events, such as the current COVID-19 pandemic, which span quarters or even years, are the exception. Increasingly, the tourism and hospitality industries are utilizing big data generated through consumers' interactions with internet and mobile technology as an additional source of high frequency information for demand modelling. These data, such as search engine queries, website traffic, and social media posts are now frequently used to extract explanatory variables which may improve the accuracy of forecasting systems (Li *et al.*, 2020). Social media data's relevance is amplified during times of crisis and volatility, since information is of the utmost importance for all stakeholders (including current and potential travelers), and social media coverage (i.e., user-generated content posted on social media) is increasingly viewed as a fast and reliable source of

news. Based on this principle, previous studies have demonstrated the use of social media data as an approximation of crisis severity for the detection of crisis events (Atefeh & Khreich, 2015; Kankanamge *et al.*, 2020). Consequently, it is proposed that crisis severity, as measured by the magnitude of social media coverage during a crisis event, may also be useful in improving tourism and hotel demand forecasts. Based on this proposition, the present study investigates under which conditions such social media data may assist high frequency forecasts of hotel room demand.

2. Literature

Tourism demand forecasting is essential for the government and industry to optimize resource allocation and plan the future development of the industry and business (Song *et al.*, 2019; Wu *et al.*, 2017). Especially in the hotel sector, the accurate forecast of demand can improve the revenue management of hotels in their pricing strategies. The following section briefly reviews the methods used in hotel demand forecasting, the application of big data in hotel demand modeling and forecasting, and the works focusing on destination risk perception, social media coverage and tourism demand.

2.1 Hotel demand forecast

Different from tourism demand modeling and forecasting literature which can collect the tourism demand data by origin-destination pairs such as Li *et al.* (2019) and Lin *et al.* (2015), most hotel demand studies use aggregated data to measure demand as the data by source markets are usually not available. Therefore, the explanatory variables in the demand theory such as income of the source market and relative price between the destination and source market suggested by Song and Li (2008) are not applicable in hotel forecasting models.

In the hotel demand literature, an early attempt is Andrew *et al.* (1990) who applied Box-Jenkins and exponential smoothing (ETS) methods to predict monthly occupancy rates of a full-service hotel. They found that the former method performed better than the latter. Schwartz and Hiemstra (1997) compared the forecasting accuracy of the curve similarity approach, auto-regressive (AR) model, higher order polynomial model (HOPM) and the combination of AR and HOPM in predicting the advanced booking data. The daily forecasting results indicated the curve similarity approach is more accurate than the other methods. Weatherford and Kimes (2003) did a comprehensive forecasting analysis with six methods to predict the hotel demand of Choice and

Marriot Hotels, respectively. They concluded that ETS, moving average (MA) and pick-up model are the most robust methods. Lim *et al.* (2009) used data from New Zealand and concluded that Holt–Winters and Box–Jenkins auto-regressive moving average (ARMA) are the most accurate methods. In contrast, Ellero and Pellegrin (2014) suggested the ARMA family of models cannot beat pick-up models in the case of Italy. Pereira (2016) introduced the trigonometric trend and multiple seasonal patterns (TBATS) method into hotel demand forecasting and showed its superiority in addressing the complicated seasonality in high-frequency demand data. Zhang *et al.* (2017) integrated ensemble empirical mode decomposition (EEMD) with auto-regressive integrated moving average (ARIMA) to forecast the daily occupancy rate. The EEMD-ARIMA model improves the forecasting accuracy of the classical ARIMA model.

Law (1998) is the pioneer study which introduced the artificial neural network (ANN) method into the hotel demand forecasting field. He found that the ANN model outperforms the regression and Naïve models in the case of Hong Kong. Assaf and Tsionas (2019) integrated a recursive vector auto-regressive (VAR) model into the ANN framework. Their investigation suggested that the new method is more accurate than classic VAR models. In addition to the time series and artificial intelligent methods which are two popular approaches in tourism demand modeling and forecasting literature (Wu *et al.*, 2017), some scholars are keen to apply the simulation method to examine the forecasting accuracy in the hotel demand practice. For example, Zakhary *et al.* (2011) used a Monte Carlo algorithm to simulate the guest arrivals and occupancy rate forward in time. The simulation result is more accurate than the pick-up and ETS methods. Schwartz *et al.* (2016) included the booking information of competitors in a simulation model and confirmed the superiority of the new model. Schwartz *et al.* (2019) proposed a comprehensive forecasting system which integrates the machine learning (Law, 1998), human learning (i.e., judgmental forecast) (Yüksel, 2007) and forecast combination (Li *et al.*, 2019). Empirical results suggested the combination of the learning mechanism can significantly improve the forecasting accuracy in the long-run.

Compared to the tourism demand forecasting methods, less advanced methods are introduced into the hotel demand forecasting literature. Although classical forecasting models such as Naïve models, ETS and ARMA family models are difficult to beat (Song *et al.*, 2019), more emerging methods including interval forecast combination (Li *et al.*, 2019), scenario forecasting (Wu *et al.*,

2020), spatial forecasting (Wen et al., 2021) and hierarchical forecasts (Hu et al., 2021) should be introduced from the tourism demand forecasting field to the hotel demand forecasting literature.

2.2 Big data and hotel demand forecast

With the development of internet technology and mobile applications, search engines are frequently used in travel planning and tourism product selections (Li & Law, 2020; Volchek *et al.*, 2019). Thus, the big data which captures tourist and guest behaviors has become available in the tourism and hospitality industries (Li *et al.*, 2018; Song & Liu, 2017). Scholars have also started to employ big data into hotel demand forecasting practice. Pan *et al.* (2012) is the first attempt which applied search engine queries into a traditional ARIMAX model to predict the hotel demand of Charleston in the US. They found that the newly introduced variable not only improved the forecasting accuracy of corresponding ARIMA models, but also beat other econometric models including auto-regressive lag (ADL), time-varying parameter (TVP) and VAR models. Yang *et al.* (2014) included the web traffic data of the destination marketing organization (DMO) into the ARIMA model. They further confirmed Pan *et al.* (2012)'s conclusion that the introduction of big data can reduce the error of ARIMA family models in the forecast of hotel demand. Rivera (2016) extended Pan *et al.* (2012)'s research by using the data from Puerto Rico as the example. Zhang *et al.* (2019) further generalized the conclusions of previous studies by swapping Google for Baidu (a search engine in China) and using Hainan Province, China to conduct the forecasting practice. Both Rivera (2016) and Zhang *et al.* (2019) confirmed the superiority of the search engine queries data for forecasting hotel demand. The above literature suggested that search engine queries and web traffic are powerful indicators to predict hotel demand, thus Pan and Yang (2017) included the two variables in one model to examine the forecasting accuracy. Previous findings were confirmed again when the two variables were introduced into the models separately. However, when both variables were included into one model simultaneously, the improvement to the forecasting accuracy was only marginal.

2.3 Rationale of the Current Study

As indicated by Wu *et al.* (2017), compared with the large number of studies focusing on tourism demand forecasting, there is limited research shedding light on hotel demand and studies which forecast hotel demand during crisis events are even rarer. According to the best knowledge of the authors, Song *et al.* (2011) is the only study which focused on hotel demand forecasting and crisis.

They predicted the recovery of hotel demand after the Global Financial Crisis in 2008. The study demonstrated the importance of the tourism and hospitality industries in the economic recovery after the crisis. However, the forecast of demand fluctuations during a crisis will have more valuable implications to the hospitality industry, especially for the dynamic pricing strategy in revenue management.

Tourism risk perception is a well-established predictor of tourism demand for international tourism destinations, where demand decreases when perceived risks of the travelers exceed a threshold (Cui *et al.*, 2016; Garg, 2015). Major subjective factors influencing tourism risk perceptions include demographic variables (e.g., age, gender, income) and cognitive abilities, while objective factors are often categorized across multiple dimensions including physical/health/safety risk, financial/economic risk, psychological risk, performance risk, time risk, and social risk (Cui *et al.*, 2016). Safety and security are frequently identified as principal factors that influence destination choice and travel motivation, and when travelers perceive a loss of control over their personal safety, the trip is normally cancelled or postponed (Lai *et al.*, 2018; Reisinger & Mavondo, 2005). This phenomenon is well documented in the crisis management literature, where short-run tourism demand is often dramatically reduced after major events such as terrorism, natural disaster, or political upheaval (Araña & León, 2008; Schroeder *et al.*, 2018).

Importantly, media attention of such events plays a significant role in the image formation and perceived magnitude of risk for the tourism destination (Amara *et al.*, 2012; Garg, 2015). In addition to the traditional television and print media channels, social media is now a primary source of information for society. As such, social media coverage of a destination crisis (such as the 2019 social unrest in Hong Kong), as well as the destination's own use of social media, now effectively communicate the severity of a crisis event and significantly contribute to perceived risks (Assaker and O'Connor, 2020; Sigala, 2011). In fact, social media coverage, which is frequently considered a more trustworthy source of information by many travelers, is fast becoming a fundamental source of information used when evaluating the potential risks of visiting any destination (Assaker & O'Connor, 2020; Lai *et al.*, 2018; Schroeder & Pennington-Gray, 2014).

Thus, the originality of this study is the adaptation of social media coverage data to approximate the severity of a destination crisis event and to input this approximation into the classical time series models to forecast the daily occupancy rate of hotels within the destination. The aim of the

current study is not only to reduce the forecasting errors in hotel demand forecasting practice by employing crisis severity in the model, but also to investigate the timing of when the intervention variable should be integrated into the predictive model. Such investigation could provide informative suggestions to the industry, based on which the hotels can develop more reasonable pricing strategies and make themselves more resilient to crisis in the future.

3. Method

3.1 Background and Data

Large-scale social unrest occurred in Hong Kong throughout the second half of 2019. The series of protests, also known as the Anti-Extradition Law Amendment Bill Movement, were triggered by the proposition of a fugitive extradition amendment in early 2019. The protests escalated into violence after June 2019 and caught the attention of the world. The crisis had devastating impact on the economy of Hong Kong, especially the tourism industry. The volume of visitor arrivals in 2019 has dropped by more than 14% in comparison with 2018, and the decline is almost 40% when considering the second half of the year only. Many hotels have experienced abnormally low occupancy rates and have encountered serious problems in operating. Therefore, accurate forecasts of room demand during a crisis event are essential for hotels to control costs and survive .

Following previous studies such as Atefeh & Khreich (2015) and Kankanamge *et al.* (2020), social media data, in particular, the frequency of Twitter posts with key hashtag occurrences, are used as to approximate the crisis severity of the 2019 Hong Kong protests. Twitter is chosen not only because of data legitimacy (the Search API service is officially operated and managed by Twitter), but also due to the nature of the platform in comparison with other popular social media sites (e.g., Facebook, Instagram). As mentioned in Boukes (2019), Twitter derives “the most direct traffic to news websites” and generate more discussion on political events, whereas other social media sites prioritize “socializing” and feature “two-way personal communication”. One concern regarding the use of social media data in approximating crisis severity is the decline in social media coverage with time. This concern is not applicable to the current investigation because the 2019 Hong Kong social unrest consists of multiple events spreading over the second semester of 2019. Thus, fall-off in social media coverage was minimum given the short interval between each event. A supervised detection approach by hashtag statistics was used to detect the events of social unrest in Twitter streams (Atefeh & Khreich, 2015; Kankanamge *et al.*, 2020). The choice of hashtags

follows an earlier investigation by Leow (2019) and includes 22 out of 25 of the most popular Twitter hashtags associated with the social unrest in Hong Kong. Several hashtags identified by Leow (2019), such as #china, #hongkong and #policebrutality, were dropped due to their lack of specificity. Twitter's premium Search API service, which facilitates exhaustive queries of the entire Twitter archive, was used to determine the count of all tweets containing at least one of the following hashtags: #standwithhongkong, #hongkongprotests, #hkpolice, #standwithhk, #soshk, #hongkongpolice, #hkpoliceterrorism, #cuhk, #fightforfreedom, #hongkongers, #freehongkong, #hkpolicestate, #hkpolicebrutality, #polyU, #fightforrights, #fightfortheright, #resisttoexist, #hkpolicestate, #hkprotesters, #protecthkstudents, #sospolyu, or #soshongkong. No additional requirements (such as language or location of tweets) were enforced in the search protocol. In April 2020, a simple Python script was used with the Twitter premium Search API service to observe the daily aggregate count of tweets containing any of the above hashtags for each day between 1 June and 31 December 2019. In total, 214 daily observations were collected.

The data on occupancy rate of Hong Kong hotels is collected from STR, which contains daily aggregate occupancy rates of five hotel classes (luxury, upper upscale, upscale, upper midscale, and midscale) from 1 January to 31 December 2019. The original classification in STR data also includes economy class, but this class is omitted in the present study due to data unavailability during the investigation period. To fit with the tweet counts data, for each of the remaining five classes, 214 observations from 1 June to 31 December 2019 are included in the investigation.

The summary of sample statistics is provided in Figure 1. The blue solid line illustrates the trend of average occupancy rate of each hotel class (tweet counts), and the box plot in each month describes the dispersion of occupancy rate (tweet counts) within the month. In the present study, the box represents the data within the first and the third quartiles, and the high and low ends of the whiskers are the maximum and minimum values of the monthly occupancy rate (tweet counts), respectively. It can be observed that the occupancy rates of all hotel classes have started significant declines in August as the crisis increased in severity (indicated by the surge in tweet counts). In the same month, the standard deviations of occupancy rates revealed an increasing trend, indicating a higher volatility in the market. Among the five hotel classes, the occupancy rate of luxury hotels suffers the most impact from the social unrest, dropping from 86.86% in March to 43.88% in December, whereas the upper midscale and midscale hotels exhibit the most resilience, with a 30%

drop from the peak to the bottom. In terms of market volatility, the average standard deviation of occupancy rate is approximately 1.5 higher than that prior to the social unrest period for all classes except for the upscale class. The market volatility reaches a peak in October for upper midscale and midscale hotels, after which the occupancy rates stayed low but became relatively stable. The market volatility of luxury and upper upscale hotels peaked one month later than that of upper midscale and midscale hotels. In November 2019, the occupancy rates of luxury and upper upscale hotels have reached 73.0% and 81.7% on 12 November, respectively, but rapidly dropped to 38.8% and 44.1% on 17 November, respectively, in responding to a severe strike on 11 November. The number of tweets with relevant hashtags has increased exponentially since the beginning of the social unrest, from double digit tweets at the beginning of June to more than 1 million in mid-November. The surge in the number of tweets demonstrated the dynamics of the crisis severity during the social unrest.

[Figure 1 about here]

3.2 Model

Three frequently adopted time-series models, namely seasonal autoregressive integrated moving average (SARIMA) model, exponential smoothing (ETS) model, and seasonal and trend decomposition using Loess (STL) model, are used as the benchmark models to test the effectiveness of crisis severity in improving forecast accuracy. Influences of the crisis severity are further integrated into the benchmark models to reflect the impact of crisis severity on the occupancy rate of hotels. The models with severity intervention are denoted as SARIMAX, ETSX, and STLX, respectively, with their properties to be discussed below. Forecast combinations are frequently proved to be effective in improving forecast accuracy and stability (Song *et al.*, 2019; Wong *et al.*, 2007). In the current investigation, a simple average combination of three benchmark models is included as one candidate in the comparison. Correspondingly, the simple average combination of three intervention models is also involved.

3.2.1 Seasonal Autoregressive Integrated Moving Average (SARIMA)

SARIMA model is a member of ARMA family, which was originally proposed by Box and Jenkins (1970). It is very popular in the forecasting literature and is also proven to perform well in terms of forecast accuracy (Song *et al.*, 2019). A SARIMA model integrates an AR component, a moving MA component, and seasonality into one model, and is usually specified as $ARIMA(p,d,q)(P,D,Q)_m$, where p is non-seasonal AR order, d is non-seasonal degree of differencing, q is non-seasonal MA order, P is seasonal AR order, D is seasonal degree of differencing, Q is seasonal MA order, and m is the time cycle of the seasonal pattern. In the current context of daily occupancy rate forecasts, the time cycle is set to be $m = 7$, reflecting the weekly seasonality in daily data. The estimation of SARIMA model is done with the `auto.arima()` function of the “forecast” package (Hyndman *et al.*, 2019) in the R software, in which the parameters of the model are automatically determined by the Box-Jenkins approach.

3.2.2 Exponential Smoothing (ETS)

ETS model is another popular tool in the time-series analysis of tourism demand (Song *et al.*, 2019). The initial appearance of the model in statistical literature can be traced back to Brown (1959) and Holt (1957) and it reached the current form with decades of innovative development (Hyndman and Athanasopoulos, 2018). The model includes the analysis of the level, the trend, and the seasonality of a time-series, by utilizing multiple applications of smoothing (averaging) on the data with exponentially decreasing weights. The general form of the model is written as $ETS(e, t, s)$, where e denotes the error type, which can be additive (A) or multiplicative (M); t denotes the trend type which can be “no trend” model (N), additive trend (A), additive damped trend (A_d), or multiplicative trend (M); and s denotes the seasonality type, which can be “no seasonality” model (N), additive seasonality (A), and multiplicative seasonality (M). In the present study, ETS model is estimated with `es()` function of the “smooth” package (Svetunkov, 2020) in the R software, in which the error type, trend type, and seasonality type are determined automatically by Akaike Information Criteria (AIC) comparison.

3.2.3 Seasonal and Trend Decomposition using Loess (STL)

As suggested by its name, STL model performs forecasts by decomposing the seasonal and trend components of a time-series. The decomposition method was initially developed by Cleveland *et al.* (1990) in the light for overcoming the disadvantages of classical decomposition methods (e.g., SEATS and X11). Compared to SEATS and X11, STL allows more flexible seasonality frequency

such as daily data, exhibits better robustness against outliers, and provides more control to the user (such as the changing rate of seasonality pattern and the smoothness of the trend-cycle component) (Hyndman & Athanasopoulos, 2018). In the general setting, STL decomposes a time-series into three parts: the seasonal component (S), the trend component (T), and the irregular component (R). Forecasting using STL involves two steps: the seasonal component (S) and the seasonal adjusted component (T + R) are decomposed first and performed with time-series forecasts separately. The forecasts of the seasonal component and the seasonally adjusted component are then combined to form the forecast of the origin time-series. In the present study, following the default choice in Hyndman *et al.* (2019), the seasonal component is forecast using seasonal naïve model and the seasonally adjusted component is forecasted using ETS model. The estimations and forecasts are done by `stlm()` function in “forecast” package (Hyndman *et al.*, 2019) in the R software, in which the parameters of the model are automatically optimized and chosen by AIC comparison.

3.2.4 Time-series Models with Exogenous Variables

The intervention models (SARIMAX, ETSX, and STLX) extend the corresponding time-series model by including exogenous variables in the time-series analysis. In the current context, the natural logarithm of tweet counts of different lags are included as exogenous variables to measure the crisis severity. Time lags of tweet counts data from one-day-backwards to 28-days-backwards (four weeks) are considered and the combinations of these lagged variables are introduced into the intervention models. In each round of forecasts, the model with the set of lags achieving the lowest AIC is chosen as the best for each corresponding intervention model. It is worth noting that the model fit of different forecast horizons are conducted separately, so that only the appropriate subset of the lagged variables is considered. For example, in the 14-day-ahead forecast models, tweet counts of 14 or more-days-backwards are the only considerations. This treatment guarantees that there is no need to “forecast” crisis severity in order to generate the forecast of occupancy rates. In the present study, the exogenous variables are introduced into the model by the “xreg” argument in the `auto.arima()` function (SARIMAX) and `es()` function (ETSX and STLX).

3.2.5 Forecast combination

It is well acknowledged in the forecasting literature that “no model can universally outperform all others in terms of forecasting accuracy” (Song *et al.*, 2019, p. 355). Model combination is therefore a popular choice to enhance the stability of forecasts. Throughout the literature, the

forecasts are usually combined by either simple average (e.g., Coshall & Charlesworth, 2011) or weighted average according to forecast performance (e.g., Coshall, 2009). In the present study, since the social unrest is causing significant variations in the data, which leads to unstable forecast performance, simple average is adopted to combine the forecast models. Specifically, the average of forecasts of SARIMA, ETS, and STL forms the combined forecast of time-series model (CB), and the average of forecasts of SARIMAX, ETSX, and STLX constitutes the combined forecast of intervention model (CBX).

3.2.6 Forecast rolling and accuracy measures

In the tourism and hospitality literature, forecasts are usually conducted and averaged over a time window to enhance the robustness of accuracy comparison (Li *et al.*, 2019). The test set in most of the studies is usually a fixed period of time at the end of a target time-series (e.g., Li *et al.*, 2019; Volcheck *et al.*, 2019). In the present study, however, one fixed test set would not be sufficient in examining the influence of the crisis severity. As revealed by the data (see Figure 1), volatility of the occupancy rates was changing across different classes of hotels as well as different time horizons. Therefore, the effectiveness of the crisis severity would not be universal throughout the investigation period. In the present study, two test sets of size one month are considered. More specifically, two sets of forecasts are generated, one with the data from June 2019 to October 2019, using October 2019 (31 days) to test the forecast accuracy, and the other with data from July 2019 to November 2019, using November 2019 (30 days). The choice of the two test sets is based on the crisis severity. December 2019 is excluded from the study not only because of the subsiding trend of the crisis, but also due to the potential noise caused by the anecdotes of a new disease (later named as COVID-19) in the latter half of the month.

To evaluate and compare the forecast performance of the intervention models and benchmark time-series models, the mean absolute percentage error (MAPE) and root mean squared error (RMSE) which are the most popular forecasting accuracy measurements (Wu *et al.*, 2017), are calculated as:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|a_i - \hat{a}_i|}{a_i},$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (a_i - \hat{a}_i)^2},$$

where $\hat{a}_i$ and a_i represent the forecast value and actual value of occupancy rate, respectively; and n is the width of the test set. Smaller values of accuracy measures demonstrate better forecasting ability of the corresponding model. Other forecast accuracy measures, such as mean absolute error (MAE), mean squared error (MSE), root mean squared percentage error (RMSPE), and mean absolute scaled error (MASE) are also available upon request.

4. Findings

Table 1 presents the point forecast accuracy measures of each forecast model of selected forecast horizon¹, with the best model for each class and horizon marked in dark bold. The performance rankings of each forecast model from 1-period-ahead ($H = 1$) to 14-period-ahead ($H = 14$) forecasts of all classes with October as the test set are presented in Figure 2. Lower positions (towards 1 on the vertical axis) in the Figures represent lower accuracy measures and thus better forecasts. The dashed lines are the rankings of time-series benchmark models, whereas the solid lines are the rankings of intervention models. The corresponding models (i.e., STL vs STLX, SARIMA vs SARIMAX, ETS vs ETSX, and CB vs CBX) are marked with the same color. The left and right panels of Figure 2 correspond to the accuracy measures of MAPE and RMSE respectively.

In general, the forecasts are very good in comparison with some previous studies on forecasting daily occupancy rate (e.g., MAPEs between 10% and 20% in Schwartz *et al.*, 2016; MAPEs above 10% in Zhang *et al.*, 2017). The models seem even more promising given that the forecast periods in the presents study are amid volatile social unrest. Taking October as the test set, SARIMAX model generally provides the best forecasts of occupancy rates of all classes and all forecast horizons, followed by CBX model, as demonstrated in both Table 1 and Figure 2. In general, the intervention models are more accurate than their counterparts on most forecast horizons. This general observation indicates the strong ability of crisis severity data in improving the forecast

¹ Only selected forecast horizons are presented due to the limited space. The full set of forecast accuracy measurements are available upon request.

accuracy of time-series models during the social unrest period. Although previous research has attempted to use the crisis severity to predict the tourism demand recovery from COVID-19 (Liu et al., 2021), this is the first study to empirically disclose the effectiveness of crisis severity on the improvement of forecasting accuracy during the crisis.

[Table 1 about here]

[Figure 2 about here]

The story is even more interesting in the forecasting results taking November as the test set. While intervention models still show superiority in some classes and forecast horizons, the ranking results are more stochastic. Table 2 and Figure 3 present the forecast measures of selected horizons and the rankings of all horizons, respectively. In terms of the magnitude of the forecast accuracy measures, the best model of each class and horizon still perform well with lower than 6% MAPEs in short horizons and around 20% MAPEs in long horizons. The ranking of different models, however, are not as straightforward as those forecasts taking October as the test set. For luxury and upper upscale hotels, intervention models still show significant superiority by outperforming time-series models in most forecast horizons. Nevertheless, the winning model is no longer unique. SARIMAX, ETSX, STLX, and CBX, have been identified as the best model for 10, 17, 13, and 14 times, respectively, in different measures and horizons. This observation coincides with the argument that no single model can be regarded as superior for all markets and all times (Song and Li, 2008; Song *et al.*, 2019; Qiu et al., 2021).

For upscale, upper midscale, and midscale hotels, the superiority of intervention models is less obvious. Among 84 forecast comparisons across hotel classes, forecast measures, and forecast horizons, the intervention models outperform time-series models 55 times. For the remaining 29 occasions, time-series models provide the best forecasts and the crisis severity data has lost its power in improving the forecast accuracy. This observation can relate to the change in volatility in the markets of upscale, upper midscale, and midscale hotels. As shown in Figure 1, after the

significant increases in market volatilities in October, the markets of upper midscale and midscale hotels become more stable in November, whereas the volatility burst in the market of upscale hotels has stopped. In November, the social unrest still had its severe impact as revealed by the average occupancy rates. However, the impact becomes relatively consistent across the month.

This finding is different from the previous research using lower frequency models, such as Li and Law (2020) and Li *et al.* (2020). The information included in low frequency demand series (e.g., annual and quarterly data) is significantly less than that in high frequency series (e.g., daily data). As a result, in low frequency models, the introduction of social media data can help to improve the forecasting accuracy in normal periods. In contrast, in high frequency models the information included in the historical data of the demand variable is rich enough to generate accurate forecasts and, thus, the contribution of social media data is limited. Nonetheless, the information included in the historical data of the demand variable is not enough to explain the strong fluctuation during the crisis, and the rich and timely information of social media data can effectively measure the crisis severity and compensate for the information gap. Consequently, during crisis events, the performance of high frequency forecasting models can be improved by including crisis severity, as measured by social media, as an intervention.

[Table 2 about here]

[Figure 3 about here]

Another observation regards the forecast accuracy measurement of RMSE. For upscale, upper midscale, and midscale hotels, while MAPE reveals indecisive results regarding the comparison between time-series models and intervention models, RMSE generally indicates a better forecast accuracy of the intervention models. Among the 42 comparisons across different hotel classes and horizons using RMSE, 38 occasions favor intervention models. Different forecast accuracy measurements may have varying results (Liu et al. 2020). The different performance of MAPE and RMSE in this study is related to the structure of RMSE, where each forecast error is weighted

by the magnitude of itself by squaring the error. In contrast to MAPE, where the errors are equally weighted, RMSE penalize big errors more than small ones. Therefore, RMSE favor models with smoother forecasts rather than methods providing good forecasts in most periods but bad forecasts during shocks. As an example, Figure 4 demonstrates the 1-day-ahead forecasts of upper midscale hotels of all models during 13 November 2019 and 19 November 2019, the week after the 11 November strike. The upper panel presents the actual occupancy rates and the lower panel displays the forecast errors. The occupancy rates have dropped significantly from 85.35% to 57.25% in one week and a considerable amount of errors are inevitably generated. It can be observed that on date 13, 14, 15, and 19, where the forecast errors are considerably big, the intervention models outperform time-series models, whereas on date 17 and 18, when the forecast errors are small, time-series models win. Intervention models would therefore be recommended by RMSE and time-series models by MAPE.

[Figure 4 about here]

5. Implications and Conclusions

The following implications can be inferred from the findings and observations of the present study. In general, the inclusion of crisis severity measured by social media data is found to be effective to improve accuracy in the forecast of hotel room occupancy rate during times of social unrest. Amid the volatile period of social unrest, time-series models cannot adapt to the speed of change in data while modeling and forecasting the hotel room demand. The social media coverage indicated by tweet count data can provide researchers and practitioners with useful information regarding the changes in market demand when the market becomes more volatile. The intervention of crisis severity in the model would remedy the inadaptability of time-series models and improve the forecast accuracy.

Nevertheless, the effectiveness of data is not universal. The market response to social unrest is not static in a sense that market volatility is dynamic through time. In the current context of the 2019 Hong Kong protests, the monthly standard deviation of hotel occupancy rates grew larger in the first few months of the social unrest and started to decline in the later months. The decline in

the standard deviation of hotel occupancy rates did not indicate an end of the crisis event, but rather it is a sign that the negative impact of the social unrest became relatively consistent through time. Correspondingly, the performance of time-series models is dynamic throughout the entire period of a crisis event. The model predictability is relatively poor in the first few months of social unrest, where the market is more volatile. This is the period in which social media data become very effective in improving forecast accuracy. However, with the decrease of market volatility and when the time-series model has adapted to the changes in data, the forecast performance is improved. The crisis severity data will lose their power and become insignificant.

We therefore argue that, in order to achieve the most accurate forecast of hotel demand, it is very important to select the most appropriate forecasting model. During normal periods, hotels should maintain the regular forecasting models for routine operation. However, when a crisis hits the market, hotels should look for more advanced models to do the forecast and control the timing of integrating a crisis severity variable (such as Twitter hashtag frequency) into predictive models. The relative market volatility should be considered as one potential indicator to determine the effectiveness of such social media data. When the volatility is above a certain range, models with social media information should be used and the cut-off value to trigger the social media model should be specifically determined by the historical data of various hotels.

Some theoretical implications also emerge from the above observations. It is obvious from the findings that the choice of test set is vital to the comparisons of model accuracy. Due to the stochastic feature of market demand during social unrest, it is inappropriate to use a unique and fixed test set for the comparison throughout the entire period. The dynamic of market volatility leads to changes of forecasting performance of different models. Therefore, to achieve accurate and robust forecasting results, multiple test sets should be considered and examined for the analysis during social unrest periods. Furthermore, while each forecast accuracy measure provides error information from a different perspective and the choice of measure should align with decision makers' utility, we would recommend a higher priority on the forecast accuracy measures of the second moments (e.g., RMSE, RMSPE) in the analysis of hotel demand during social unrest, e.g., in comparison to first moment counterparts (e.g., MAE, MAPE, MASE). By construction, the accuracy measures of second moments take the square of each error and focus on the dispersion of forecast errors. Such accuracy measures penalize big errors more than small ones. In the forecast

of market demand in regular periods, it is usually argued that the accuracy measures of second moments can be easily influenced by a few large outliers and provide unstable forecasts (e.g., Andersen *et al.*, 1999). The case is reversed during a social unrest period, in which most of the data points are “outliers”. It is therefore beneficial for researchers and practitioners to utilize forecast accuracy measures of second moments and pay closer attention to the forecasts of these “outliers”.

The findings of the present study are not achieved without limitations. First, the results are obtained from the observations of five hotel classes in Hong Kong, and may not be generalized to other geographical settings, especially non-tourism environments. Second, due to data availability and legitimacy concerns, the crisis severity in the present study is approximated by tweet counts. Ideally, given the large volume of Mainland Chinese visitors in Hong Kong’s inbound market, Weibo or Baidu data would be more suitable for the purpose of the present study. Unfortunately, these sources of social media data are not made publicly available due to legitimacy and privacy concerns. Therefore, we call for future research efforts to consider other time and locations of social unrest as well as a more comprehensive index of crisis severity. In addition, it would be interesting to extend the crisis severity measurement to a broader context, compositing the statistics of different hashtags and integrating the information from multiple sources, such as search query, website traffic, and social media posts. Such investigation could benefit the most from today’s big data environment and facilitate more advanced analysis of hotel (or tourism) demand during social unrest.

Reference

- Amara, D.; El Samadicy, A.; and Ragheb, M. (Eds.) (2012), "Tourists' risk aversion and willingness to take risks: the case of tourists visiting Egypt after 25th January revolution." *6th World Conference for Graduate Research in Tourism, Hospitality and Leisure*, Vol. 22.
- Andersen, T.G., Bollerslev, T., and Lange, S. (1999), "Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon", *Journal of Empirical Finance*, Vol. 6 No. 5, pp.457-477.
- Andrew, W.P., Cranage, D.A., and Lee, C.K. (1990), "Forecasting hotel occupancy rates with time series models: An empirical analysis", *Hospitality Research Journal*, Vol. 14 No. 2, pp.173-182.
- Araña, J.E. and León, C.J. (2008), "The impact of terrorism on tourism demand", *Annals of Tourism Research*, Vol. 35 No. 2, pp.299-315. DOI: 10.1016/j.annals.2007.08.003.
- Assaf, A.G. and Tsionas, M.G. (2019), "Forecasting occupancy rate with Bayesian compression methods", *Annals of Tourism Research*, Vol. 75, pp.439-449.
- Assaker, G. and O'Connor, P. (2020), "eWOM platforms in moderating the relationships between political and terrorism risk, destination image, and travel intent: The case of Lebanon", *Journal of Travel Research*, DOI: 10.1177/0047287520922317.
- Atefeh, F. and Khreich, W. (2015), "A survey of techniques for event detection in Twitter", *Computational Intelligence*, Vol. 31 No. 1, pp.132-164.
- Blake, A. and Sinclair, M.T. (2003), "Tourism crisis management: US response to September 11", *Annals of Tourism Research*, Vol. 30 No.4, pp.813-832.
- Boukes, M. (2019). "Social network sites and acquiring current affairs knowledge: The impact of Twitter and Facebook usage on learning about the news", *Journal of Information Technology & Politics*, Vol. 16 No. 1, pp.36-51. DOI: 10.1080/19331681.2019.1572568
- Box, G. and Jenkins, G. (1970). *Time Series Analysis, Forecasting and Control*, Holden Day, San Francisco, CA.
- Brown, R.G. (1959). *Statistical Forecasting for Inventory Control*, McGraw/Hill, New York, NY.

- Cleveland, R.B., Cleveland, W.S., McRae, J.E., and Terpenning, I.J. (1990), "STL: A seasonal-trend decomposition procedure based on Loess", *Journal of Official Statistics*, Vol. 6 No. 1, pp.3-33.
- Coshall, J.T. (2009), "Combining volatility and smoothing forecasts of UK demand for international tourism", *Tourism Management*, Vol. 30 No. 4, pp.495-511.
- Coshall, J.T. and Charlesworth, R. (2011), "A management orientated approach to combination forecasting of tourism demand", *Tourism Management*, Vol. 32 No. 4, pp.759-769.
- Cui, F., Liu, Y., Chang, Y, Duan, J. and Li, J. (2016), "An overview of tourism risk perception", *Nat Hazards*, Vol. 82 No. 1, pp.643-658. DOI: 10.1007/s11069-016-2208-1.
- Ellero, A. and Pellegrini, P. (2014), "Are traditional forecasting models suitable for hotels in Italian cities?", *International Journal of Contemporary Hospitality Management*, Vol. 26 No. 3, pp.383-400.
- Garg, A. (2015), "Travel risks vs tourist decision making: A tourist perspective", *International Journal of Hospitality & Tourism Systems*, Vol. 8 No. 1, pp.1-9. DOI: 10.21863/ijhts/2015.8.1.004.
- Holt, C.C. (1957), "Forecasting seasonals and trends by exponentially weighted moving averages", *ONR Memorandum*, Vol. 52, Pittsburgh, PA: Carnegie Institute of Technology.
- Hu, M., Qiu, R. T., Wu, D. C., & Song, H. (2021). "Hierarchical pattern recognition for tourism demand forecasting". *Tourism Management*, Vol. 84, DOI: 104263.
- Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. (2nd ed.), OTexts, available at: <https://otexts.org/fpp2/>
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M. and Razbash, S. (2019), "Package 'forecast'", available at: <https://cran.r-project.org/web/packages/forecast/forecast.pdf>
- Kankanamge, N., Yigitcanlar, T., Goonetilleke, A., & Kamruzzaman, M. (2020). "Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets", *International Journal of Disaster Risk Reduction*, 42, 101360.

- Lai, I.K.W., Hitchcock, M., Lu, D. and Liu, Y. (2018), "The influence of word of mouth on tourism destination choice: Tourist–resident relationship and safety perception among mainland Chinese tourists visiting Macau", *Sustainability*, Vol. 10 No. 7, p.2114. DOI: 10.3390/su10072114.
- Law, R. (1998), "Room occupancy rate forecasting: a neural network approach", *International Journal of Contemporary Hospitality Management*, Vol. 10 No. 6, pp.234-239.
- Leow, G. (2019), "Analysis of tweets on the Hong Kong protest movement 2019 with Python", *Towards Data Science*, available at: <https://towardsdatascience.com/analysis-of-tweets-on-the-hong-kong-protest-movement-2019-with-python-a331851f061> (accessed 7 April 2020)
- Li, G., Wu, D.C., Zhou, M. and Liu, A. (2019), "The combination of interval forecasts in tourism", *Annals of Tourism Research*, Vol. 75, pp.363-378.
- Li, J., Xu, L., Tang, L., Wang, S., & Li, L. (2018). "Big data in tourism research: A literature review". *Tourism Management*, Vol.68, pp. 301-323.
- Li, H., Hu, M. and Li, G. (2020), "Forecasting tourism demand with multisource big data", *Annals of Tourism Research*, Vol. 83, DOI: 10.1016/j.annals.2020.102912.
- Li, X., & Law, R. (2020). "Forecasting tourism demand with decomposed search cycles". *Journal of Travel Research*, Vol. 59, No. 1, pp. 52-68.
- Lim, C., Chang, C. and McAleer, M. (2009), "Forecasting h (m) otel guest nights in New Zealand", *International Journal of Hospitality Management*, Vol. 28 No. 2, pp.228-235.
- Lin, V.S., Liu, A. and Song, H. (2015), "Modeling and forecasting Chinese outbound tourism: An econometric approach", *Journal of Travel & Tourism Marketing*, Vol. 32 No. 1-2, pp.34-49.
- Liu, A., Lin, V. S., Li, G., & Song, H. (2020). "Ex Ante Tourism Forecasting Assessment". *Journal of Travel Research*, 0047287520974456.
- Liu, A., Vici, L., Ramos, V., Giannoni, S., & Blake, A. (2021). "Visitor arrivals forecasts amid COVID-19: A perspective from the EUROPE team". *Annals of Tourism Research*, 103182.

- Page, S., Song, H. and Wu, D.C. (2012), "Assessing the impacts of the global economic crisis and swine flu on inbound tourism demand in the United Kingdom", *Journal of Travel Research*, Vol. 51 No. 2, pp.142-153.
- Pan, B., Wu, D.C. and Song, H. (2012), "Forecasting hotel room demand using search engine data", *Journal of Hospitality and Tourism Technology*, Vol. 3 No. 3, pp.196-210.
- Pan, B. and Yang, Y. (2017), "Forecasting destination weekly hotel occupancy with big data", *Journal of Travel Research*, Vol. 56 No. 7, pp.957-970.
- Pereira, L.N. (2016), "An introduction to helpful forecasting methods for hotel revenue management", *International Journal of Hospitality Management*, Vol. 58, pp.13-23.
- Pratt, S. and Liu, A. (2016), "Does tourism really lead to peace? A global view", *International Journal of Tourism Research*, Vol. 18 No. 1, pp.82-90. DOI: 10.1002/jtr.2035.
- Qiu, R.T.R., Masiero, L. and Li, G. (2018). "The psychological process of travel destination choice", *Journal of Travel & Tourism Marketing*, Vol. 35 No. 6, pp.691–705.
- Qiu, R. T., Wu, D. C., Dropsy, V., Petit, S., Pratt, S., & Ohe, Y. (2021). "Tourist arrival forecast amid COVID-19: A perspective from the Asia and Pacific team". *Annals of Tourism Research*, 103155.
- Reisinger, Y. and Mavondo, F. (2005), "Travel anxiety and intentions to travel internationally: implications of travel risk perception", *Journal of Travel Research*, Vol. 43 No. 3, pp.212-225. DOI: 10.1177/0047287504272017.
- Rivera, R. (2016), "A dynamic linear model to forecast hotel registrations in Puerto Rico using Google Trends data", *Tourism Management*, Vol. 57, pp.12-20.
- 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.31-38.
- Schroeder, A. and Pennington-Gray, L. (2014), "Perceptions of crime at the Olympic Games: What role does media, travel advisories, and social media play?", *Journal of Vacation Marketing* Vol. 20 No. 3, pp.225-237.

- Schroeder, A., Pennington-Gray, L., Kim, H., & Liu-Lastres, B. (2018). "Using the media's tweets to broaden previous conceptualizations of political travel risks", *Tourism Management Perspectives*, Vol. 26, pp.107–117. DOI: 10.1016/j.tmp.2018.03.001
- Schwartz, Z. and Hiemstra, S. (1997), "Improving the accuracy of hotel reservations forecasting: curves similarity approach", *Journal of Travel Research*, Vol. 36 No. 1, pp.3-14.
- Schwartz, Z., Uysal, M., Webb, T. and Altin, M. (2016), "Hotel daily occupancy forecasting with competitive sets: A recursive algorithm", *International Journal of Contemporary Hospitality Management*, Vol. 28 No. 2, pp.267-285.
- Schwartz, Z., Webb, T., van der Rest, J.P.I. and Koupriouchina, L. (2019), "Enhancing the accuracy of revenue management system forecasts: The impact of machine and human learning on the effectiveness of hotel occupancy forecast combinations across multiple forecasting horizons", *Tourism Economics*, DOI: 10.1177/1354816619884800.
- Sigala, M. (2011), "Social media and crisis management in tourism: Applications and implications for research." *Information Technology & Tourism* Vol. 13 No. 4, pp.269-283.
- Song, H. and Li, G. (2008), "Tourism demand modelling and forecasting -- A review of recent research", *Tourism Management*, Vol. 29 No. 2, pp.203-220.
- Song, H., Lin, S., Witt, S. F. and Zhang, X. (2011), "Impact of financial/economic crisis on demand for hotel rooms in Hong Kong". *Tourism Management*, Vol. 32 No. 1, pp.172-186.
- Song, H., & Liu, H. (2017). "Predicting tourist demand using big data". In *Analytics in Smart Tourism Design* (pp. 13-29) edited by Xiang and Fesenmiar. Springer, Cham.
- Song, H., Qiu, R.T.R. and Park, J. (2019), "A review on tourism demand forecasting: Launching the *Annals of Tourism Research* Curated Collection on tourism demand forecasting", *Annals of Tourism Research*, Vol. 75, pp.338-36.
- Svetunkov, I. (2020), "Package 'smooth'", available at: <https://cran.r-project.org/web/packages/smooth/smooth.pdf>
- Volchek, K., Liu, A., Song, H. and Buhalis, D. (2019), "Forecasting tourist arrivals at attractions: Search engine empowered methodologies", *Tourism Economics*, Vol. 25 No. 3, pp.425-447.

- Weatherford, L.R. and Kimes, S.E. (2003), "A comparison of forecasting methods for hotel revenue management", *International Journal of Forecasting*, Vol. 19 No. 3, pp.401-415.
- Wen, L., Liu, C., Song, H., & Liu, H. (2021). "Forecasting tourism demand with an improved mixed data sampling model". *Journal of Travel Research*, Vol.60 No.2, pp.336-353.
- Wong, K.K., 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.1068-1078.
- Wu, D. C., Cao, Z., Wen, L., & Song, H. (2020). "Scenario forecasting for global tourism". *Journal of Hospitality & Tourism Research*, 1096348020919990.
- Wu, D.C., Song, H. and Shen, S. (2017), "New developments in tourism and hotel demand modeling and forecasting", *International Journal of Contemporary Hospitality Management*, Vol. 29 No. 1, pp.507-529.
- Yang, Y., Pan, B. and Song, H. (2014), "Predicting hotel demand using destination marketing organization's web traffic data", *Journal of Travel Research*, Vol. 53No. 4, pp.433-447.
- Yüksel, S. (2007), "An integrated forecasting approach to hotel demand", *Mathematical and Computer Modelling*, Vol. 46 No. 7-8, pp.1063-1070.
- Zakhary, A., Atiya, A.F., El-Shishiny, H. and Gayar, N.E. (2011), "Forecasting hotel arrivals and occupancy using Monte Carlo simulation", *Journal of Revenue and Pricing Management*, Vol. 10 No. 4, pp.344-366.
- Zhang, B., Pu, Y., Wang, Y. and Li, J. (2019), "Forecasting hotel accommodation demand based on LSTM model incorporating internet search index", *Sustainability*, Vol. 11 No. 17, p.4708.
- Zhang, G., Wu, J., Pan, B., Li, J., Ma, M., Zhang, M. and Wang, J. (2017), "Improving daily occupancy forecasting accuracy for hotels based on EEMD-ARIMA model", *Tourism Economics*, Vol. 23 No. 7, pp.1496-1514.

Table 1(a) Forecast performance – Luxury class (October)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0903	0.0877	0.0813	0.0805	0.0826	0.0831	0.0808	0.0770
<i>2</i>	0.1495	0.1383	0.1406	0.1373	0.1420	0.1292	0.1385	0.1323
<i>3</i>	0.1837	0.1614	0.1688	0.1682	0.1684	0.1505	0.1637	0.1624
<i>7</i>	0.1097	0.1205	0.1097	0.1116	0.1097	0.1099	0.1078	0.1059
<i>14</i>	0.1396	0.1515	0.1396	0.1426	0.1392	0.1289	0.1385	0.1345
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	6.2284	5.8738	5.5999	5.6367	5.7148	5.7680	5.6060	5.5852
<i>2</i>	9.6750	8.9388	8.8413	8.8377	8.9862	8.5698	8.7623	8.7075
<i>3</i>	11.7833	11.0983	10.8009	10.8670	10.8621	10.5505	10.6162	10.6534
<i>7</i>	7.9340	8.7207	7.9342	8.0472	7.9248	7.6046	7.8673	7.6209
<i>14</i>	10.6398	11.4368	10.6400	10.8538	10.6575	9.4474	10.6454	10.2923

Table 1(b) Forecast performance – Upper Upscale class (October)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0804	0.0746	0.0742	0.0741	0.0732	0.0612	0.0729	0.0662
<i>2</i>	0.1336	0.1183	0.1318	0.1249	0.1282	0.0941	0.1293	0.1105
<i>3</i>	0.1700	0.1505	0.1709	0.1617	0.1677	0.1197	0.1672	0.1488
<i>7</i>	0.1112	0.1237	0.1112	0.1098	0.1067	0.0936	0.1046	0.0965
<i>14</i>	0.1244	0.1527	0.1244	0.1252	0.1245	0.1083	0.1239	0.1188
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	6.2388	5.8726	5.8543	5.8077	5.9039	5.1501	5.8507	5.4522
<i>2</i>	9.9023	9.1043	9.6783	9.3214	9.6009	7.6200	9.6060	8.6120
<i>3</i>	12.5467	11.8544	12.3756	11.9749	12.3168	9.6069	12.3766	11.1850
<i>7</i>	9.5884	10.5798	9.5886	9.7339	9.3282	8.0208	9.2513	8.7786
<i>14</i>	10.2544	12.9562	10.2547	10.7312	10.2490	9.0413	10.2399	9.8725

Table 1(c) Forecast performance – Upscale class (October)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0702	0.0646	0.0657	0.0641	0.0656	0.0628	0.0654	0.0614
<i>2</i>	0.1069	0.0843	0.1016	0.0952	0.1020	0.0780	0.1013	0.0919
<i>3</i>	0.1310	0.0971	0.1276	0.1150	0.1260	0.0928	0.1244	0.1120
<i>7</i>	0.0938	0.0801	0.0923	0.0855	0.0879	0.0691	0.0879	0.0789
<i>14</i>	0.0818	0.0695	0.0798	0.0766	0.0786	0.0712	0.0783	0.0745
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	6.0627	5.8685	5.7256	5.7171	5.7332	5.4962	5.7176	5.5075
<i>2</i>	8.2314	6.8605	7.9036	7.4716	7.8780	6.2981	7.8485	7.2359
<i>3</i>	10.2738	8.2878	9.9922	9.3014	9.9334	7.7356	9.9566	9.1246
<i>7</i>	7.7587	6.7677	7.5877	7.2705	7.2012	5.5107	7.1904	6.6566
<i>14</i>	6.8031	5.7761	6.6311	6.3765	6.5522	5.7736	6.5270	6.2161

Table 1(d) Forecast performance – Upper Midscale class (October)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0818	0.0807	0.0743	0.0774	0.0752	0.0754	0.0736	0.0724
<i>2</i>	0.1416	0.1360	0.1325	0.1340	0.1336	0.1012	0.1322	0.1211
<i>3</i>	0.1825	0.1734	0.1739	0.1750	0.1737	0.1205	0.1712	0.1552
<i>7</i>	0.1443	0.1414	0.1443	0.1360	0.1405	0.1093	0.1382	0.1244
<i>14</i>	0.1218	0.1790	0.1218	0.1372	0.1217	0.1058	0.1214	0.1156
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	7.4290	7.4033	6.7879	7.0107	6.8695	6.7045	6.7524	6.6418
<i>2</i>	10.9811	10.2095	10.3324	10.2976	10.4121	8.3618	10.2977	9.5389
<i>3</i>	13.8322	13.2680	13.3680	13.2263	13.3955	9.9706	13.3380	12.0974
<i>7</i>	11.9140	11.6370	11.9143	11.5480	11.6223	8.4092	11.3717	10.1887
<i>14</i>	10.5865	15.5638	10.5869	11.9793	10.5704	8.9298	10.4408	9.9707

Table 1(e) Forecast performance – Midscale class (October)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.1089	0.0915	0.1010	0.0977	0.0991	0.0870	0.0985	0.0932
<i>2</i>	0.1506	0.1289	0.1540	0.1425	0.1518	0.1105	0.1519	0.1373
<i>3</i>	0.1835	0.1463	0.1921	0.1702	0.1903	0.1253	0.1870	0.1613
<i>7</i>	0.1502	0.1217	0.1499	0.1353	0.1438	0.1044	0.1415	0.1172
<i>14</i>	0.1315	0.1146	0.1318	0.1256	0.1281	0.1004	0.1262	0.1174
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	9.2166	8.6271	8.7877	8.6011	8.6648	8.1933	8.6533	8.3385
<i>2</i>	12.1351	10.4300	12.1696	11.2222	12.0260	9.8984	11.9930	10.9243
<i>3</i>	14.6829	11.8229	15.1937	13.5603	15.0946	10.9249	15.0408	13.2134
<i>7</i>	12.5000	10.1240	12.4602	11.3993	11.8302	9.1044	11.5085	9.7175
<i>14</i>	10.7799	9.8496	10.7834	10.3725	10.5245	8.7344	10.5497	9.7883

Table 2(a) Forecast performance – Luxury class (November)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0687	0.0653	0.0667	0.0629	0.0662	0.0620	0.0643	0.0587
<i>2</i>	0.1155	0.1343	0.1192	0.1152	0.1155	0.1108	0.1087	0.1045
<i>3</i>	0.1456	0.1880	0.1601	0.1587	0.1522	0.1346	0.1448	0.1404
<i>7</i>	0.1571	0.2512	0.1571	0.1879	0.1538	0.1656	0.1521	0.1495
<i>14</i>	0.2293	0.3702	0.2293	0.2733	0.2093	0.2656	0.2052	0.2176
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	4.2388	4.1327	4.0915	3.9783	4.0595	3.8268	3.9718	3.7471
<i>2</i>	6.8471	7.7073	7.1229	6.9420	6.9296	6.6621	6.7142	6.4929
<i>3</i>	8.6648	10.4927	9.2837	9.1321	8.9506	8.3777	8.6382	8.3851
<i>7</i>	10.9477	15.0540	10.9477	11.9911	10.1209	9.7204	9.9688	9.8365
<i>14</i>	12.7597	20.0288	12.7598	14.5444	11.9310	13.7103	11.7388	12.3403

Table 2(b) Forecast performance – Upper Upscale class (November)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0730	0.0752	0.0607	0.0671	0.0616	0.0607	0.0596	0.0596
<i>2</i>	0.1226	0.1515	0.1051	0.1227	0.1037	0.1226	0.1026	0.1096
<i>3</i>	0.1527	0.1908	0.1347	0.1541	0.1355	0.1500	0.1320	0.1368
<i>7</i>	0.1362	0.1898	0.1362	0.1458	0.1334	0.1549	0.1332	0.1375
<i>14</i>	0.2413	0.3103	0.2413	0.2640	0.2207	0.2502	0.2209	0.2289
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	5.6442	5.2930	4.8347	5.0389	4.9124	4.4766	4.8327	4.4908
<i>2</i>	9.2589	10.1413	8.4456	8.8991	8.3874	8.3184	8.4392	7.8424
<i>3</i>	11.5753	13.3959	10.8015	11.4042	10.6964	10.5856	10.7347	10.0628
<i>7</i>	11.6180	13.8584	11.6180	12.0808	11.1645	11.1375	11.3681	11.1999
<i>14</i>	15.6710	20.0657	15.6710	16.9307	14.8469	15.7957	14.8694	15.2101

Table 2(c) Forecast performance – Upscale class (November)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0487	0.0532	0.0463	0.0454	0.0474	0.0532	0.0469	0.0455
<i>2</i>	0.0765	0.0818	0.0726	0.0727	0.0732	0.0893	0.0729	0.0749
<i>3</i>	0.0970	0.0918	0.0895	0.0890	0.0891	0.1070	0.0882	0.0905
<i>7</i>	0.0827	0.0931	0.0826	0.0839	0.0833	0.1003	0.0835	0.0833
<i>14</i>	0.1067	0.1165	0.1064	0.1091	0.1076	0.1289	0.1080	0.1090
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	4.2407	4.1091	3.9091	3.8193	3.9563	4.1059	3.9273	3.7540
<i>2</i>	6.6739	6.1226	6.4520	5.9762	6.4111	6.4987	6.4713	6.0009
<i>3</i>	8.1164	7.0147	7.9089	7.1666	7.7391	7.6514	7.8480	7.1739
<i>7</i>	7.5115	7.3157	7.4986	7.3085	7.3014	7.2061	7.3325	6.9700
<i>14</i>	8.8809	8.6636	8.8538	8.7285	8.7269	8.7132	8.6614	8.1466

Table 2(d) Forecast performance – Upper Midscale class (November)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0608	0.0748	0.0511	0.0616	0.0518	0.0621	0.0512	0.0544
<i>2</i>	0.1017	0.1278	0.0910	0.1053	0.0911	0.1134	0.0910	0.0949
<i>3</i>	0.1240	0.1577	0.1121	0.1290	0.1118	0.1314	0.1120	0.1156
<i>7</i>	0.0960	0.1117	0.0960	0.0955	0.0951	0.1162	0.0952	0.0960
<i>14</i>	0.1070	0.1647	0.1070	0.1193	0.1080	0.1330	0.1086	0.1107
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	5.2628	5.8825	4.5921	5.0837	4.6443	5.0364	4.6053	4.5851
<i>2</i>	8.4066	10.0578	7.7253	8.4660	7.7695	8.4974	7.7451	7.6450
<i>3</i>	10.0728	12.0904	9.3977	10.1831	9.3833	9.8870	9.4329	9.0795
<i>7</i>	9.1217	10.0722	9.1218	9.3041	8.7772	9.2834	8.7740	8.8309
<i>14</i>	9.3275	13.1531	9.3276	10.1912	9.3856	10.2476	9.3829	9.5372

Table 2(e) Forecast performance – Midscale class (November)

MAPE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	0.0719	0.0742	0.0576	0.0635	0.0566	0.0641	0.0566	0.0550
<i>2</i>	0.0940	0.1102	0.0737	0.0881	0.0762	0.0972	0.0755	0.0787
<i>3</i>	0.1114	0.1316	0.0925	0.1055	0.0921	0.1166	0.0917	0.0932
<i>7</i>	0.0977	0.1013	0.0980	0.0979	0.1011	0.0970	0.1019	0.1039
<i>14</i>	0.1171	0.1399	0.1175	0.1188	0.1249	0.1231	0.1271	0.1305
RMSE								
<i>H</i>	STL	SARIMA	ETS	CB	STLX	SARIMAX	ET SX	CBX
<i>1</i>	6.0727	5.9521	4.8091	5.3367	4.8431	5.1562	4.7622	4.6416
<i>2</i>	8.3591	8.6049	6.6575	7.3875	6.7970	7.4611	6.7134	6.5163
<i>3</i>	10.0003	10.3188	8.1545	8.9182	8.3292	8.8791	8.2301	7.8775
<i>7</i>	8.2008	8.6073	8.1950	8.2103	8.3129	7.3487	8.3597	7.8330
<i>14</i>	9.6884	11.7240	9.7162	10.1087	10.2137	8.8935	10.3395	10.1549

Figure 1 Sample trends of occupancy rates and Tweet counts

Figure 2(a) Forecast performance – Luxury class (October)

Figure 2(b) Forecast performance – Upper Upscale class (October)

Figure 2(c) Forecast performance – Upscale class (October)

Figure 2(d) Forecast performance – Upper Midscale class (October)

Figure 2(e) Forecast performance – Midscale class (October)

Figure 3(a) Forecast performance – Luxury class (November)

Figure 3(b) Forecast performance – Upper Upscale class (November)

Figure 3(c) Forecast performance – Upscale class (November)

Figure 3(d) Forecast performance – Upper Midscale class (November)

Figure 3(e) Forecast performance – Midscale class (November)

Figure 4 Example forecasts of upper midscale hotels