Anagnostopoulou, S., Buhalis, D., Kountouri, I., Manousakis, E. and Tsekrekos, A. (2020), " The impact of online reputation on hotel profitability", International Journal of Contemporary Hospitality Management,Vol. 32(1), pp. 20-39. https://doi.org/10.1108/IJCHM-03-2019-0247

The impact of online reputation on hotel profitability

Seraina C. Anagnostopoulou[*] Dimitrios Buhalis[†] Ioanna L. Kountouri[‡]
Eleftherios G. Manousakis[‡] Andrianos E. Tsekrekos[§]

Purpose: The purpose of this study is to quantify the impact of online customer reputation on financial profitability.
Design/methodology/approach: Online reputation is captured by extracting the most recurring textual themes associated with customer satisfaction and dissatisfaction, expressed within positive vs. negative online guest reviews on Booking.com. Latent Semantic Analysis is used. Proxies of overall financial performance are manually constructed for the sample hotels, using financial data from the FAME (Financial Analysis Made Easy) database. Ordinary Least Squares is employed to gauge the effect of online customer reputation on financial profitability.
Findings: Empirical findings indicate that recurring textual themes from positive online reviews (in contrast to negative reviews) exhibit a higher degree of homogeneity and consensus. The themes repeated in *positive*, but not in *negative* reviews, are found to significantly associate with hotel financial performance. Results contribute to the discussion about the measurable effect of online reputation on financial performance.
Originality/value: Contemporary quantitative methods are used to extract online reputation for a sample of UK hotels, and associate this reputation with bottom-line financial profitability. The relationship between online reputation, as manifested within hotel guest reviews, and the financial performance of hotels is examined. Financial profitability is the result of revenue, minus cost incurred in order to offer a given level of service. Previous studies have mainly focused on basic measures of performance i.e. revenue generation, rather than bottom-line profitability. By combining online guest reviews from travel websites (Booking.com) with financial measures of enterprise performance (FAME), this study makes a meaningful contribution to the strategic management of hotel businesses.
Keywords: Quantitative content analysis; online hotel reviews; eWOM; hotel reputation; Latent Semantic Analysis; hotel performance
Acknowledgements: The last author gratefully acknowledges financial support, through grant EP-2337-01 of RC/AUEB

[*] University of Piraeus, Department of Banking and Financial Management, 80, M. Karaoli & A. Dimitriou St, 18534 Piraeus, Greece, Tel: +30 210 414 2189, Email: seraina@unipi.gr
[†] Bournemouth University, Department of Tourism and Hospitality, Email: dbuhalis@bournemouth.ac.uk
[‡] Athens University of Economics and Business, Department of Management Science and Technology, Email: ioannakoudouri@yahoo.gr, lgmanousakis@gmail.com
[§] Athens University of Economics and Business, Department of Accounting and Finance, Email: tsekrekos@aueb.gr

Introduction

Online customer reviews represent a form of electronic Word of Mouth (eWOM). Reviews are considered to be more successful in influencing consumer behavior, compared to traditional marketing tools (Phillips et al., 2017). They are typically independent of marketers' selling efforts, so they appear more trustworthy and credible in the eyes of consumers (Nieto et al., 2014). Online reviews are important sources of information that can shape the reputation of tourism enterprises (Au et al., 2014; Phillips et al., 2015).

Past research has established a significant effect of customer online review characteristics (i.e. review valence, volume, or dispersion) on hotel performance. This is expressed in the form of room bookings, sales, prices, revenue per available room (RevPAR) or occupancy (e.g. ; Anderson, 2012; Kim et al., 2015; Sainaghi et al., 2017; Viglia et al., 2016; Xie et al., 2014; Ye et al., 2009). However, the impact of user generated online reviews on business activities has focussed on sales and bookings, and less on the impact on bottom-line *financial* performance (Phillips et al., 2015). Hotel managers have to ensure the achievement of profitability high enough to cover investment, assets and the related fixed costs (Ben Aissa and Goaied, 2016). Although a close link between occupancy and profitability has been supported (Viglia et al., 2016), profitability, rather than room occupancy is the ultimate goal of every hotel manager (Jeffrey and Barden, 2000).

The purpose of this study is to quantify the effect of online reputation created through customer reviews on tourism business performance, following relevant calls by Phillips et al. (2015) and Anderson (2012). Increasingly, reputation depends on public opinion online, largely depending on the variety of images, opinion and discussions online (Buhalis and Inversini, 2014). User Generated Comments (UGC) create a digitalized WOM, and the aggregation of all this online feedback creates web reputation. Reputation is crucial for the desirability of tourism and hospitality products and services, and it is, therefore, a major asset that requires investment and being managed (Berezina et al., 2016; Bolton et al., 2004), Reputation influences the ability of organizations to improve their profitability and sustainable growth (Viglia et al., 2016; Xiang et al., 2017).

Hotel reputation among customers is manifested through the description and arguments used in online reviews. In our study, reputation is extracted by subjecting hotel review comments, characterized as 'positive' or 'negative' by customers themselves, from the travel website of Booking.com to textual analysis using Latent Semantic Analysis (LSA). LSA allows the extraction of underlying textual concepts from online guest reviews using a statistical way and is free of any reader/coder bias and judgement. Reviews on Booking.com were used as the site is widely thought of as more reliable than Tripadvisor.com. This is because Booking.com only invites reviews from people who have made a reservation and paid through the site, and, therefore, it can avoid fake reviews (Yahoo Finance, 2019). The statistical association of these themes is then examined, with respect to hotel profitability expressed in financial terms. This research focuses on independent hotels, rather than chain hotels. Independent hotels are self-proprietary properties not affiliated to any brand, whereas chain hotels belong to the same hotel group or consortium. Given these differences, rating patterns attracted by independent and chain hotels are not always comparable (Banerjee and Chua, 2016).

To the best of our knowledge, this is the first study to associate online review content with bottom line financial performance defined in a way that combines revenues, expenses and investment, as previous studies stop at defining financial performance in the form of RevPAR, bookings or booking income. To produce a satisfied customer and create positive online reputation, the hotel in question needs to make a relevant effort, and this effort implies sacrificing economic resources, or incurring costs to produce a given level of service. This cost is simply ignored when research stops at defining financial performance in terms of revenue or booking income.

The approach taken in this paper is to measure hotel performance in terms of achieved *financial results* that combine revenues against expenses made for service offered to customers and capital investment. This is accomplished by manually compiling a sample that consists of all UK unlisted, and independent (non-chain belonging) hotels located in 13 major cities, namely: Birmingham, Bradford, Bristol, Cardiff, Edinburgh, Glasgow, London, Leeds, Liverpool, Manchester, Newcastle, Sheffield, and Wolverhampton. Financial/accounting data from the FAME (Financial Analysis Made Easy) database were obtained for hotels with at least one online guest review on Booking.com. By combining online guest reviews from Booking.com with financial measures of enterprise performance (FAME) for the first time, this study makes a meaningful contribution to the strategic management of hotel businesses. It identifies which attributes are actually associated with 'bottom line' financial earnings per unit of capital investment.

Literature review and research hypotheses

Online hotel customer reviews and hotel performance

The internet has revolutionised hospitality marketing and distribution (Buhalis, 2000; Law et al., 2014; Paraskevas et al., 2011).

One stream of research has established a significant effect of guest review characteristics (review valence (rating), volume (number of reviews), or dispersion) on hotel performance. This performance is often expressed in the form of room bookings, sales, booking income, prices, RevPAR or occupancy (Viglia et al., 2016). Review scores and volume positively relate to bookings and/or sales, Average

Daily Rates and RevPAR, and also booking transaction value (Anderson, 2012; Kim et al., 2015; Torres et al., 2015; Xie et al., 2014; Ye et al., 2009).

Another stream of research has used quantitative content analysis to extract consumer preferences from the text of customer online reviews. This enables users to discover relationships, patterns or trends in textual data (Singh et al., 2007; Stepchenkova et al., 2009). Li et al. (2015) use the Emerging Pattern Mining technique to identify emergent hotel features of interest and changes in customer preferences, extracted through online reviews. This is in line with similar methods applied by Xiang et al. (2009), and Chaves et al. (2012). Stringam and Gerdes (2010), and Stringam et al. (2010) also subject comments on online hotel distribution sites to textual analysis. They identify frequently used words, patterns of word usage, and their effect on hotel features rating. Li et al. (2013), Liu et al. (2013) and Berezina et al. (2016) analyze customer expectations,preferences, and common themes mentioned by satisfied and dissatisfied customers using text mining. Mazanec (2017) and Guo et al. (2017) apply latent Dirichlet allocation to provide results on the heterogeneity of perceptions across different demographic profiles. Xu and Li (2016) apply LSA to extract the determinants that create either customer satisfaction or dissatisfaction. . Liu et al. (2017) use big data to investigate the role of tourist cultural backgrounds for the formation of preferences for various hotel attributes. The emphasis of this research in on the extraction of consumer preferences via text mining of online reviews, and the identification of which hotel attributes are mostly associated with customer satisfaction.

Research hypotheses

WOM taking place via oral communication normally involves face-to-face or some form of interpersonal contact. This allows body language and voice intonations to reinforce the meaning of transmitted messages (Jeong and Jang, 2011; Nieto et al., 2014). In contrast, eWOM involves written feedback and lacks non-verbal cues (Nieto et al., 2014). Review contents are more influential than ratings (Fong et al., 2017), because consumers rely more on sentiment when making purchase decisions (Ong, 2012). Ratings also luck detailed evaluations (Racherla et al., 2013). For this reason, focusing on the textual descriptions of online customer reviews results in additional insights about which factors are important for customers to enhance (or reduce) the value of their stay. These factors (e.g. the quality of rooms, Internet provision, building, Phillips et al., 2017) are expected to shape hotel online reputation.

Customers evaluate the degree of satisfaction based on their perceptions of hotel attributes that they consider most important. These hotel attributes represent dimensions of satisfaction (Guo et al., 2017). The overwhelming majority of visitors regularly examine online reviews before booking hotels and also read online reviews before making purchase decisions (Tan et al., 2018). Decision making and consumption processes for tourism products tend to be different, because of their experiential and intangible nature (Tan et al., 2018).

The importance of customer ratings lies on their expected influence on hotel bookings, and ultimately, performance (Phillips et al., 2017). Research has provided extensive insights about which hotel attributes and quality of service factors are important to customers. Such factors include room cleanliness, convenience of location, value-for-money and friendliness of staff (Phillips et al., 2017). More specific factors, such as breakfast, food service, wellness or Wi-Fi are also important (Albayrak and Caber, 2015). Technology has been transforming the nature of service provision and the ability to understand context and dynamically cocreate tourism hospitality is increasingly critical in service excellence (Buhalis and Leung, 2018; Buhalis and Sinarta, 2019; Buhalis et al., 2019).

The literature on hotel industry uniformly indicates that there are many attributes that influence the customers' choice, ranging from location and room rate, to the quality of service, reputation, security, and cleanliness (Chu and Choi, 2000; Becerra et al., 2013). Certain hotel characteristics, e.g. service quality, are arguably important for all customers. However, some customers value certain features more than others (Becerra et al., 2013), so they should be expected to be more strongly associated with the revenue side of financial performance. Customer reviews are idiosyncratic by nature, as travellers have individualised preferences. They may weight differently the several attributes of a hotel, such as the quality of breakfast or hotel location depending on their internal and external context (Buhalis and Foerste, 2015; Neirotti et al., 2016). Sometimes, there is disagreement about the desirability of different hospitality services attributes among guests (Becerra et al., 2015). Different cultural backgrounds, needs and types of travel affect guest preferences for various hotel attributes e.g. value for money, cleanliness, and location (Liu et al., 2017).

Previous research on customer perceptions and decision making and online customer reviews has focused on marketing and impact (Phillips et al., 2017). The focus on the influence of online reviews on sales and bookings may well be justified by a difficulty to match review data with performance data. Research has shown that positive (negative) valence of online customer reviews is positively (negatively, or neutrally) associated with hotel performance (Phillips et al., 2017; Ye et al., 2009). WOM generates positive attitudes and increases the possibility of purchase, whilst negative WOM has the opposite effect (Melián-González et al., 2013). This is consistent with the established relationship between profit, growth and customer loyalty (Kim et al., 2015).

Game-theory research indeed addresses prices as signals and suggests that, in the long run, firms are able to charge a premium price only for quality service (Israely, 2002). Prices may work as signals of quality, and firms that do not provide quality service according to industry standards will have to settle for less than a premium price. Hence, quality, related to the level of service offered, is important for price determination (Israely, 2002). Torres et al. (2015) demonstrates that hotel ratings, ranking and number of reviews have a significant impact on a hotel's ability to generate revenues through its booking transactions.

This study examines the association between online hotel reputation, as captured though user-mined textual themes extracted from positive and negative online hotel reviews with different degrees of importance, and hotel performance, defining performance in a way that blends profit with investment made for the achievement of this profit. Evaluating the effect of online reputation, shaped by customer reviews to the hotel bottom-line financial performance offers a strategic management perspective. The service level provided requires the sacrifice of variable economic resources, in order to achieve quality and satisfactory hotel product attributes for customers. This is often reflected on financial statements through the incurrence of accounting costs. In order for service of a given level and quality to be offered to customers, so that customers appreciate the service in question in a positive (negative) way, the incurred revenue/cost relation must result in a positive (negative) net financial result. Profit also has to be consistent with the level of investment made in the production process, especially as the cost of debt is reflected into profitability via the creation of interest expenses. Therefore the following research hypotheses are developed:

H1: Positive online reputation, as manifested through textual themes explaining the variation in positive customer online review comments, is positively associated with hotel financial performance.

H2: Negative online reputation, as manifested through textual themes explaining the variation in negative customer online review comments, is negatively associated with hotel financial performance.

Latent Semantic Analysis (LSA) and online hotel guest reviews

LSA was introduced by Dumais et al. (1988) to extract the main underlying or latent textual meanings and concepts from a sample of text. Compared to other natural language processing techniques used in analyzing online hotel guest reviews (Stepchenkova et al., 2009; Stringam and Gerdes, 2010; Li et al., 2015), LSA has innate advantages. These include: statistical and objective nature; noise-reduction properties and ability to quantify text similarities; common 'latent' or hidden semantic structure, whilst handling synonymy and polysemy at the same time. It permits the identification of specific, latent textual themes within positive and negative guest reviews, which explain (according to different percentages of efficiency) the overall variability of reviews by hotel guests.

Before performing LSA, all positive online guest reviews ($d^{+} = 189{,}967$) in the sample written in English were grouped together into one textual collection called the "(positive) corpus", and similarly all negative online guest reviews ($d^{-} = 155{,}594$) into one "(negative) corpus". Each review on Booking.com has clear and separate areas for guest-written pros and cons, and as in Bjørkelund et al. (2012), these are quantitatively analysed separately. The total sample of $d = d^{+} + d^{-}$= 345,561 online hotel guest reviews and their characteristics are further discussed in section 4.1. The two corpora were pre-processed, as it is customary in LSA applications (see Mazis and Tsekrekos, 2017).

For brevity, our exposition here uses only the 'positive' corpus of online reviews. Each pre-processed review in the 'positive' corpus $i^{+} = 1, \dots, 189{,}967 = d^{+}$, was used to construct a $[t_i^{+} \times 1]$ vector that will be referred to as the *positive term-frequency vector*. The $d^{+} = 189{,}967$ different term-frequency vectors in the sample were subsequently merged, so as to keep any common words/terms only once. This resulted in a $t^{+} \times d^{+}$ matrix, with the number of rows, $t^{+} = max\{t_i^{+}\}, i^{+} = 1, \dots, d^{+}$, to reflect the unique words/terms across all positive online reviews, and the matrix columns, d^{+}, equal to the number of positive online guest reviews in the sample. The elements of each matrix row were demeaned to produce the so-called *term-document matrix*, denoted by $\boldsymbol{X}$, which expresses a unique term's frequency in each online review in relative terms, vis-à-vis its average frequency across all reviews. Finally, the latent textual "themes" from the sample online reviews (positive and negative) were extracted using Singular Value Decomposition (SVD) of the term-document matrix $\boldsymbol{X}$. SVD expresses the term-document matrix $\boldsymbol{X}$ as

$$\boldsymbol{X}_{t^{+} \times d^{+}} = \boldsymbol{U}_{t^{+} \times d^{+}} \boldsymbol{S}_{d^{+} \times d^{+}} \boldsymbol{V}^{T}_{d^{+} \times d^{+}} \tag{1}$$

Since the term-document matrix $\boldsymbol{X}$ is mean-centered, the columns of matrix $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{X}\boldsymbol{X}^{T}$ while the rows of $\boldsymbol{V}^{T}$ are the principal components of $\boldsymbol{X}$. Matrix $\boldsymbol{S}$ is diagonal, with the square roots of the eigenvalues of $\boldsymbol{X}^{T}\boldsymbol{X}$, appearing in its main diagonal, in decreasing order. The most important, latent textual theme in the (positive or negative) corpus correspond to the highest eigenvalue (first diagonal element of $\boldsymbol{S}$), with the second most-important appearing in the second, and so on.

Sample selection/methodology

Sample selection

Guest online reviews were extracted via auto-parsed web-crawling (as in Xie et al., 2014, among others) for all hotels with at least one review on Booking.com for a two year period. Hotels in the 13 largest UK cities according to population were selected and that included Birmingham, Bradford, Bristol, Cardiff, Edinburgh, Glasgow, London, Leeds, Liverpool, Manchester, Newcastle, Sheffield, and Wolverhampton (U.K. Census 2011, Mellinas et al., 2016; p. 81). According to Table 1, this process resulted in a total of 4,668 hotels, out of which London and Edinburgh were more strongly represented (2,797 and 631 hotels, respectively). Hotels pertaining to a branded chain were deleted. This was because research has consistently shown that chain hotels have improved efficiency (Ben Aissa and Goaied, 2016), economies of scale, and better performance in terms of revenues, occupancy and other performance metrics (Ivanova and Ivanov, 2015), as well as travellers' rating patterns (Banerjee and Chua, 2016). Hotel chains also have the ability to charge higher prices (Silva, 2015). Hotel chain affiliation would influence the observed level of financial performance of sample hotels and therefore the removal of chain-brand hotels resulted in a total of 3,670 hotels.

[Insert Table 1 about here.]

This sample was then matched to hotels included in the Financial Analysis Made Easy (FAME) database, compiled by Bureau Van Dijk. The FAME database has been consistently used in accounting and finance research, due to its coverage of financial data for private companies, enabling research that makes use of such companies (Ball and Shivakumar, 2005; Brav, 2009). The manual matching process resulted in a total of 325 different matches for private or unlisted hotels (198 in the London area). This represents a significant reduction in terms of observations, as the average number of hotels with online guest reviews from Booking.com also included in FAME is marginally less than 10% on a per city basis. The number of hotels with data on financial performance proxies is even lower, with a total number of hotels with data availability on profitability and total assets to be 159, for a total of 304 hotel-year observations. This significant drop in the number of usable observations with financial data available is consistent with the identified difficulty by past research to match online reviews with actual performance data (Phillips et al., 2017). However, this innovative approach, allows for the first time the impact examination of online reputation through the use of actual financial information and controls. This also represents an attempt to *'study what is important'* instead of limiting our sight to what can be measured readily (Phillips and Moutinho, 2014).

Table 1 shows that sample hotels receive on average 381.8 positive reviews in Glasgow (median 278.0), while in Bristol the average (median) is just 36.3 (36.0). Across all sample cities, one can observe a statistically significant higher number of average (and median) positive reviews, compared to the number of negative reviews. This is consistent with the finding of Bjørkelund et al. (2012) that hotel guest numerical scores on Booking.com shift towards the higher end of the scale.

The fact that online reviews primarily skew toward the positive end of the spectrum, as hotel ratings tend to be favorable, has been established by past research (Melián-González et al., 2013; Fong et al., 2017). However, relative to positive reviews, negative reviews have a stronger relationship with perceived helpfulness (Fong et al., 2017) and are more difficult to analyze semantically (Alaei et al., 2019). The number of positive vs. negative reviews reported in Table 1 does not exhibit a particularly high degree of discrepancy such as the one observed in other research (Fong et al., 2017).

Financial performance measurement

Hotel financial performance was measured in a way that combines revenues and costs incurred for providing service to customers, as well as investment. The performance indicators used are pre-tax profit margin, expressing profitability (unaffected by externally imposed taxation but taking into account all operating and financial expenses), as a percentage of hotel sales, and also pre-tax profit scaled by the level of investment in terms of total assets. Financial performance is measured using a proxy combining costs (operating and financial ones) and revenues. That ultimately combines the amount of investment made, with total assets used as a proxy for investment.

It should be noted at this point that at the single user level, guests who wrote a negative online comment should be the same guests that contributed to the revenue generated by the hotel in question. However, such guests were most certainly affected by the (inefficient) incurrence of costs made by the hotel in order to provide them with a given level of service, which turned out to be unsatisfactory for them given their negative review. In every case, the scope of the study is to examine the effect of positive and negative online reputation at the aggregate level, under the expectation that aggregate or cumulative effect of positive and negative online reputation (via relevant reviews) should affect profitability within a given financial year, reflecting the way of measuring financial performance.

Main regression model – Control variables

The following specification was used for examining the significance of themes identified in positive and negative online reviews for hotel financial performance:

$$PBTTA_{i,t} \text{ or } PBTS_{i,t} = a_0 + \sum_{l=1}^{3 \text{ or } 5} \beta_l Positive\ Themes_{l,i,t} + \sum_{m=1}^{3 \text{ or } 5} \gamma_m Negative\ Themes_{m,i,t} + a_1 Ln\left(TA_{i,t}\right)$$

$$+a_2 Lev_{i,t} + \frac{a_3 CA_{i,t}}{CL_{i,t}} + a_4 City_i + a_5 Ln\left(No\ of\ Positive\ Reviews_{i,t}\right)$$

$$+a_6 Ln\left(No\ of\ Negative\ Reviews_{i,t}\right) + a_7 Rating_i + e_{i,t} \qquad (2)$$

where $PBTTA_{i,t}$ and $PBTS_{i,t}$ are Profit before Tax divided by Total Assets (ROA) and divided by Sales (Profit margin), respectively. Equation (2) is estimated using OLS. $Positive\ Themes_{l,i,t}$ ($Negative\ Themes_{m,i,t}$) regressors are defined as the prevalence on hotel i in year t of the main three (or five, among robustness checks) textual themes identified in the positive (negative) hotel guest reviews, estimated via LSA as described in section 3. The selection of control variables follows past research (Ben Aissa and Goaied, 2016). The natural logarithm of Total Assets $Ln\left(TA_{i,t}\right)$ is used as a hotel size proxy, given the association of hotel size with performance (Pine and Phillips, 2005; Kim et al., 2015). $Lev_{i,t}$ is a proxy for financial leverage, defined as the sum of short-term loans and long-term debt as provided by FAME, divided by Total Assets, consistent with arguments on an eventual association of financial indebtness with profitability (Ben Aissa and Goaied, 2016). Hotel companies tend to have high proportion of fixed assets, and it is generally easier for firms carrying a significant proportion of fixed assets to issue debt (Phillips and Sipahioglou, 2004). $CA_{i,t}/CL_{i,t}$ represents the current ratio, defined as current assets divided by current liabilities, and is a proxy for hotel liquidity, or the ability of the firm to cover short-term obligations with its more liquid or current assets. $City_i$ is a variable taking values from 1 to 13 depending on the city where the hotel is based, while $Rating_i$ stands for the overall rating of hotel i, as reported by Booking.com. Controls for the number of positive or negative reviews used for positive and negative theme extraction $Ln\left(No\ of\ Positive\ Reviews_{i,t}\right)$ and $Ln\left(No\ of\ Negative\ Reviews_{i,t}\right)$ were included, following the association of review volume with hotel performance (Blal and Sturman, 2014).

Star rating (and corporate affiliation) represent signals of quality, with a corresponding influence on pricing (Israeli, 2002). Distinct pricing patterns have been identified in the presence of hotel vertical differentiation, for hotels with more stars, which have the ability to offer smaller discounts and charge higher prices (Becerra et al., 2013). Given this association between the number of hotel stars and financial performance, standard errors were clustered according to the number of hotel stars.

Results

Theme extraction/interpretation

Table 2 summarizes results from applying SVD to term-document matrices (positive and negative). Panel A (B) of Table 2 summarizes the contribution of each underlying textual theme to the "variability" of positive (negative) online hotel guest reviews.

[Insert Table 2 about here.]

For *positive* online reviews, the first five, most important (highest eigenvalue) textual themes can collectively account for 54.20% of the total variance in the reviews. In contrast, five latent textual themes can only account for 29.95% of the total variation in the *negative* sample hotel reviews. One would need nearly twenty textual themes in order to account for more than half (54.04%) of the total variation in *negative* customer hotel reviews in the sample. This finding, that positive online reviews, as a corpus, appear more homogenous and concise (as opposed to negative online reviews that are lengthier, more detailed, multi-faceted and more difficult to analyze semantically), is in line with the conclusions of Alaei et al. (2019) and the findings of Bjørkelund et al. (2012) in their opinion-mining study of Booking.com. There exists more guest consensus about the hotel attributes considered satisfactory, compared to hotel attributes found dissatisfying.

Interpretation of the extracted textual themes is a difficult and subjective task. Using the loadings of ***U*** from the SVD allows one to measure the representation of each unique term to a particular textual theme. Figure 1 plots the loadings of the ten most important terms from textual Themes 1-5.

[Insert Figure 1about here.]

The key terms for Theme 1 in positive online reviews are *location, station*, *shop*, *underground, convenient*, *restaurant*, *distance*, *position*, *centre* and *attraction.* This theme is interpreted as ***"Convenience of location"***. The key terms for Theme 2 are *staff, friend, helpful, reception*, *attend*, *pleasant*, *polite*, *professional, efficient* and *concierge,* which can be interpreted as ***"Staff attentiveness and professionalism"***. Theme 3 is mainly associated with terms such as *clean*, *comfort*, *price*, *money*, *easy*, *shower*, *bed*, *quiet, wifi,* and *convenient.* These terms seem to suggest an interpretation such as ***"Room quality and value-for-money"***. Keywords in theme 4 include *breakfast, facility, kitchen, decor, restaurant, tradition, cook, food, include* and *continental*. This Theme is interpreted as ***"Breakfast and food quality"***. Theme 5 is predominantly associated with the terms *decor*, *atmosphere*, *luxury, fabulous, building, facility, relax, ambience, pool* and *spa*, which can be interpreted as ***"Amenities, ambience and luxury"***.

The key terms in Theme 1 from negative online reviews are *bathtub, loud, bedroom, bill, improve, wardrobe, garden, clean, key and lock.* This Theme is interpreted as ***"Room condition"***. The key terms in Theme 2 are *breakfast, price, include, food*, *toast*, *coffee*, *continental*, *expensive, egg* and *tea*. These terms seem to suggest an interpretation such as ***"Breakfast and food quality"***. The main terms associated with Theme 3 are *shower*, *bathroom*, *toilet*, *water*, *share*, *expensive*, *price*, *air condition, hot* and *cold,* which can be interpreted as ***"Bathroom and air condition value-for-money".*** Theme 4 can be best-described via terms such as *noise, floor, loud, road, stairs, lift, sleep, traffic, hear* and *wall*. This Theme is interpreted as ***"Noise and lack of quiet"***. Finally**,** Theme 5 is associated with the terms *staff, unfriendly*, *location, pressure, helpful, impolite, reception, car, complain* and *arrive*, which seem to suggest an interpretation like ***"Staff politeness and reception attentiveness upon arrival"***.

Positive and negative review themes and hotel financial performance

Table 3 reports estimation results for Equation (2). For brevity, the first three (and five in robustness tests) positive and negative textual themes are examined. Positive Theme 1 (convenience of location), and positive Theme 3 (overall quality of the room), are the ones most significantly associated with hotel profitability in terms of ROA and profit margins. Their coefficients are strongly positive and significant for most model specifications, with and without the inclusion of any hotel-specific control variables.

Positive Theme 2 (quality and attentiveness of staff) is also positively significant, but only when hotel-specific controls are excluded from the equation. However, for negative themes, results are less strong: negative Theme 3 (quality of the bathroom and air-conditioning), is only significant for explaining ROA when no hotel-specific controls are included, while negative Theme 2 (quality of breakfast) is again only significant in just two model specifications.

[Insert Table 3 about here.]

Regarding hotel-specific controls, a strong and significant negative effect of financial leverage on hotel financial profitability, is observed, confirming the need to examine the effect of this factor for hotel financial performance (Ben Aissa and Goiaed, 2016). The liquidity variable, $CA_{i,t}/CL_{i,t}$, is found to be positively and statistically significant in affecting hotel financial profitability in all regression specifications. Findings do not change even after including review rating among independent variables.

Table 4 reports results when the analysis is repeated by estimating Equation (2) with positive vs. negative themes separately. Results confirm previous findings on a positive and significant effect of positive Themes 1 and 3 on financial profitability. Negative themes are uniformly not significant, as in Table 3, consistent with the notion that semantic constructs from negative online reviews are multi-faceted and weaker. Results from Table 4 also confirm a strongly negative association between leverage and profitability, and a strongly positive association between hotel liquidity and profitability. Untabulated results are virtually indistinguishable if the estimation is repeated using the first five (instead of three) textual themes from online reviews.

[Insert Table 4 about here.]

Discussion and conclusions

Conclusions

This study examines the effect of online hotel reputation, extracted through guest reviews, on hotel financial performance, measured in a way that combines revenues, costs and investment. Latent Semantic Analysis is applied in order to mine customer impressions from hotel reviews, identifying the textual themes within positive and negative online reviews that are most important in explaining their overall variation.

Findings indicate that the themes that recur in guest descriptions within positive reviews have a greater power to explain the overall variation of positive reviews, compared to themes from negative reviews. It takes five themes to explain more than half of the variability of positive reviews, when the same number of themes does not explain even 30% of the variability of negative reviews. This is consistent with a greater degree of consensus among guests about the hotel attributes they consider satisfactory, compared to the (more idiosyncratic and multi-faceted) features that they found dissatisfying. Results finally indicate that positive, but not negative, review themes are significantly associated with hotel financial performance. This provides support on positive reviews being carried through to explain profitability in accounting terms. It indicates that costs incurred in order to achieve a given level of service positively materialize into financial profitability.

Most importantly, findings show that identified textual themes explaining *positive* guest reviews, *positively* and significantly associate with accounting profitability. Textual themes explaining *negative* reviews do *not* consistently have a significant effect on profitability.

Theoretical implications

Positive online reputation is carried through financial performance, while this does not occur for negative reputation. This is in contrast to marketing beliefs that negative online reviews can be catastrophic for reputation and profitability. This is because negative reviews, consistent with previous findings, tend to be seen as more idiosyncratic and multi-faceted, in comparison to positive reviews, so their effect does not appear to be carried through profitability. The findings appear to contradict past research from the marketing field, indicating that negative reviews influence consumers' decision making more than positive reviews (Phillips et al., 2017). This evidence, the more dispersed or guest-dependent negative reviews failing to translate negatively and significantly into bottom-line profitability, is the first of its kind indicating whether negative online reputation turns out to be harmful in bottom-line financial terms. However, findings are in accordance with evidence on decomposed valence scores by Phillips et al. (2017), indicating that only positive reviews have an impact on RevPAR, with the effect of negative reviews being insignificant, as the negativity effect is not strong enough to be carried through to performance.

Evidence extends Berezina et al. (2016), who use a text-mining approach on online reviews. They identify common categories that are used in both positive and negative reviews, and our paper extracts these categories and further examines which of them actually translate into financial profitability. Overall, findings indicate a significantly positive association between the themes explaining positive reviews and ROA. This implies that revenues are worth the costs incurred, per unit of investment, for a given level of service to be offered to customers. It further provides tangible evidence to Neirotti et al. (2016) discussion on the extent to which online review ratings are actually carried to financial profitability well and above their effect on sales. Neirotti et al. (2016) are critical about the fact that the increase in perceived importance of user-generated reviews in online communities for travellers might shift hotel competition from unit profit margin to volumes, and to higher room occupancy rates. Evidence indicates that positive reviews actually have a significant association with bottom line financial profitability, well and above volumes. In other words, evidence shows that in contrast to fears expressed by past literature that online reviews may only relate to bookings, we put this expectation into empirical testing and show that their impact goes all the way down and is translated into profitability. This constitutes an attempt to shift the performance measurement impact of online hotel reviews in terms of quantification, which is the ultimate business goal. It also opens relevant possibilities for future research to shift from a focus on revenues or booking income into bottom-line results. The quantification of online reputation and its business effect into bottom-line terms represents a natural extension of previous research, given that construction of positive reputation among customers comes through the sacrifice of resources to offer value to customers.

Practical implications

Results show that the two most crucial (for financial profitability) hotel attributes discussed in positive reviews are hotel location and room quality. The geographic location of a hotel, closely related to the level of investment made in the business, is its most unchangeable attribute, that cannot be altered or improved by the staff once a hotel property has been built (Xie et al., 2014). The findings are consistent with this investment being worth its costs in terms of translation into financial profitability. Hotel managers should also understand how hotel location affects their bottom-line profitability, and adjust their strategy accordingly.

Findings also indicate that hotel-specific factors financial in nature, such as financial leverage (and liquidity), are significant for explaining accounting profitability. This is consistent with the fact that hotel profitability analysis should be performed by combining business-related, with financial factors. The fact that the UK market is not subject to the same seasonal concentration as other countries (Fernández-Morales et al., 2016) helps towards results generalization for different markets with similar characteristics.

Limitations and future directions

This study unavoidably has a few limitations. First, the study is ultimately based on a limited sample from UK hotels due to the need to combine data from online customer reviews with information on financial performance and appropriate control variables. Secondly, only one online platform, Booking.com, is employed for online guest review extraction. However, this platform is widely accredited for the most authentic reviews as reviewers need to have made a reservation and paid on the site before being invited to submit a review. Nevertheless, this research would benefit from the analysis of customer reviews from other platforms and longer time periods, in support of result generalization, in light of recent evidence on significant variation across platforms in terms of linguistic characteristics, semantic features, sentiment, and rating (Xiang et al., 2017). Third, there is no control for qualitative guest characteristics, such as impressions made on repeat vs. first time visitors (Jarvis et al., 2016), business vs. leisure visitors, or local vs. foreign guests. Finally, as is the case with all textual analysis applied to any customer comments characterized as overall positive, or negative, the results may be affected by the existence of so-called dual valence reviews (Fong et al., 2017).

References

Albayrak, T., and Caber, M. (2015), "Prioritization of the hotel attributes according to their influence on satisfaction: A comparison of two techniques", *Tourism Management,* Vol. 46, pp. 43–50.

Alaei, A. R., Becken, S. and and Stantic, B. (2019), "Sentiment analysis in tourism: Capitalizing on big data", *Journal of Travel Research*, Vol. 58, No. 2, pp. 175–191.

Anderson, C. (2012), "The impact of social media on lodging performance", *Cornell Hospitality Report*, Vol. 12, No. 5, pp. 6-11.

Au, N., Buhalis, D., and Law, R., (2014), "Online complaining behavior for mainland China hotels: The perception of Chinese and non-Chinese customers", *International Journal of Hospitality and Tourism Administration*, Vol. 15, No. 3, pp. 248-274

Ball, R., and Shivakumar, L. (2005), "Earnings quality in UK private firms: Comparative loss recognition timeliness", *Journal of Accounting and Economics,* Vol. 39, No. 1, pp. 83-128.

Banerjee, S., and Chua, A.Y.K. (2016), "In search of patterns among travellers' hotel ratings in TripAdvisor", *Tourism Management*, Vol. 53, pp. 125-132.

Becerra, M., Santaló, J., and Silva, R. (2013), "Being better vs. being different: differentiation, competition, and pricing strategies in the Spanish hotel industry", *Tourism Management,* Vol. 34, pp. 71–79.

Ben Aissa, S., and Goaied, M. (2016), "Determinants of Tunisian hotel profitability: The role of managerial efficiency", *Tourism Management,* Vol. 52, pp. 278-287.

Berezina. K., Bilgihan, A., Cobanobglu, C., and Okumus, G. (2016), "Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews", *Journal of Hospitality Marketing & Management*, Vol. 25, No. 1, pp. 1-24.

Bjørkelund, E., Burnett, T.H., and Nørvåg, K. (2012), "A study of opinion mining and visualization of hotel reviews", in *Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services* (pp. 229-238). ACM.

Blal, I., and Sturman, M. (2014), “The differential effects of the quality and quantity of online reviews on hotel room sales”, *Cornell Hospitality Quarterly,* Vol. 55, No. 4, pp. 365-375.

Bolton, G.E., Katok, E., and Ockenfels, A. (2004), “How effective are electronic reputation mechanisms? An experimental investigation”, *Management Science*, Vol. 50, No. 11, pp. 1587-1602.

Brav, O. (2009), “Access to capital, capital structure, and the funding of the firm”, *Journal of Finance,* Vol. 64, No. 1, pp. 263-308.

Buhalis, D., (2000), “Information technology in tourism: Past, present and future”, *Tourism Recreation Research*, Vol. 25, No. 1, pp. 41-58.

Buhalis, D., and Foerste, M., (2015), “SoCoMo Marketing for Travel and Tourism: Empowering co-creation of value”, *Journal of Destination Marketing & Management*, Vol. 4, No. 3, pp. 151–161.

Buhalis, D., Harwood, T., Bogicevic, V., Viglia, G., Beldona, S., and Hofacker, C., (2019), “Technological disruptions in services: Lessons from tourism and hospitality”, *Journal of Service Management*, https://doi.org/10.1108/JOSM-12-2018-0398.

Buhalis, D. and Inversini, A., (2014), “Tourism branding, identity, reputation co-creation, and word-of-mouth in the age of social media”, in *Tourism management, marketing, and development* (pp. 15-40). Palgrave Macmillan, New York.

Buhalis, D., and Leung, R., (2018), “Smart Hospitality – Interconnectivity and Interoperability towards an Ecosystem”, *International Journal of Hospitality Management*, Vol. 71, pp. 41-50.

Buhalis, D., and Sinarta, Y., (2019), “Real-time co-creation and nowness service: Lessons from tourism and hospitality”, *Journal of Travel and Tourism Marketing*, Vol. 36, No. 5, pp. 563-582.

Chaves, M. S., Gomes, R., and Pedron, C. (2012), “Analysing reviews in the web 2.0: Small and medium hotels in Portugal”, *Tourism Management,* Vol. 33, No. 5, pp. 1286-1287.

Chu, R.K.S., and Choi, T. (2000), “An importance-performance analysis of hotel selection factors in the Hong Kong hotel industry: A comparison of business and leisure travellers", *Tourism Management*, Vol. 21, No. 4, pp. 363-377.

Dumais, S.T., Furnas, G.W., Landauer, T.K. and Deerwester, S. (1988), “Using latent semantic analysis to improve information retrieval”, in *Proceedings of CHI'88: Conference on Human Factors in Computing*, New York: ACM, pp. 281-285.

Fernández-Morales, A., Cisneros-Martínez, J.D., and McCabe, S. (2016), “Seasonal concentration of tourism demand: Decomposition analysis and marketing implications”, *Tourism Management,* Vol. 56, pp. 172-190.

Fong, L.H.N., Lei, S.S.l., and Law, R. (2017), “Asymmetry of hotel ratings on TripAdvisor: Evidence from single- versus dual-valence reviews”, *Journal of Hospitality Marketing & Management,* Vol. 26, No. 1, pp. 67-82.

Guo. Y., Barnes, S.J., and Jia, Q. (2017), “Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation”, *Tourism Management,* Vol. 59, pp. 467-483.

Israeli, A.A. (2002), "Star rating and corporate affiliation: Their influence on room price and performance of hotels in Israel", *International Journal of Hospitality Management,* Vol. 21, No. 4, pp. 405-424.

Ivanova, M., and Ivanov, S. (2015), "Affiliation to hotel chains: Hotels' perspective", *Tourism Management Perspectives*, Vol. 16, pp. 148-162.

Jarvis, D., Stoeckl, N., and Liu, H.-B. (2016), "The impact of economic, social and environmental factors on trip satisfaction and the likelihood of visitors returning", *Tourism Management,* Vol. 52, pp. 1-18.

Jeffrey, D., and Barden, R.R.D. (2000), "An analysis of daily occupancy performance: A basis for effective hotel marketing?", *International Journal of Contemporary Hospitality Management,* Vol. 12, No. 3, pp. 179-189.

Jeong, E., and Jang, S. (2011), "Restaurant experiences triggering positive electronic word-of-mouth (eWOM) motivations", *International Journal of Hospitality Management,* Vol. 30, No. 2, pp. 356-366.

Kim, W., Lim, H., and Brymer, R. (2015), "The effectiveness of managing social media on hotel performance", *International Journal of Hospitality Management,* Vol. 44, pp. 165-171.

Law, R., Buhalis, D., and Cobanoglu, C., (2014), "Progress on information and communication technologies in hospitality and tourism", *International Journal of Contemporary Hospitality Management*, Vol. 26, No. 5, pp. 727-750.

Li, G., Law, R., Vu, H. Q., and Rong, J. (2013), "Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet integral", *Tourism Management,* Vol. 36, pp. 321-330.

Li, G., Law, R., Vu, H. Q., Rong, J., and Zhao, X.R. (2015), "Identifying emerging hotel preferences using Emerging Pattern Mining technique", *Tourism Management,* Vol. 46, pp. 311-321.

Liu, S., Law, R., Rong, J., Li, G., and Hall, J. (2013), "Analyzing changes in hotel customers' expectations by trip mode", *International Journal of Hospitality Management,* Vol. 34, pp. 359-371.

Liu, Y., Teichert, T., Rossi, M., Li, H., and Hu, F. (2017), "Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews", *Tourism Management*, Vol. 59, pp. 554-563.

Mazis, P.G. and Tsekrekos A. E. (2017), "Latent semantic analysis of the FOMC statements", *Review of Accounting and Finance*, Vol. 16, No. 2, pp. 179-217.

Silva, R. (2015), "Multimarket contact, differentiation, and prices of chain hotels", *Tourism Management,* Vol. 48, pp. 305–315.

Mazanec, J. A. (2017), "Determining long-term change in tourism research language with text-mining methods", *Tourism Analysis*, Vol. 22, No. 1, pp. 75-83.

Melián-González, S., Bulchand-Gidumal, J., and González López-Valcárcel, B. (2013), "Online customer reviews of hotels: As participation increases, better evaluation is obtained", *Cornell Hospitality Quarterly,* Vol. 54, No. 3, pp. 274-283.

Mellinas, J.P., Martínez María-Dolores, S.M., and Bernal García, P. (2016), “Effects of the Booking.com scoring system”, *Tourism Management*, Vol. 57, pp. 80-83.

Neirotti, P., Raguseo, E., and Paolucci, E. (2016), “Are customers’ reviews creating value in the hospitality industry? Exploring the moderating effects of market positioning”, *International Journal of Information Management,* Vol. 36, No. 6, Part A, pp. 1133-1143.

Nieto, J. Hernández-Maestro, R. M., and Muñoz-Gallego, P.A. (2014), “Marketing decisions, customer reviews, and business performance: The use of the Toprural website by Spanish rural lodging establishments”, *Tourism Management.* Vol. 45, pp. 115-123.

Ong, B.S. (2012), “The perceived influence of user reviews in the hospitality industry”, *Journal of Hospitality Marketing and Management,* Vol. 21, No. 5, pp. 463-485.

Paraskevas., A., Katsogridakis, I., Law, R., and Buhalis, D. (2011), “Search engine marketing: Transforming search engines to hotel distribution channels”, *The Cornell Hotel Restaurant Administration Quarterly*, Vol. 52, No. 2, pp. 200-208.

Pine, R., and Phillips, P. (2005), “Performance comparisons of hotels in China”, *International Journal of Hospitality Management*, Vol. 24, No. 1, pp. 57-73.

Phillips, P., Barnes, S., Zigan, K., and Schegg, R. (2017), “Understanding the impact of online reviews on hotel performance: An empirical analysis”, *Journal of Travel Research,* Vol. 56, No. 2, pp. 235-249.

Phillips, P., and Moutinho, L. (2014), “Critical review of strategic planning research in hospitality and tourism”, *Annals of Tourism Research,* Vol. 48, pp. 96-120.

Phillips. P., and Sipahioglou, M.A. (2004), “Performance implications of capital structure: Evidence from quoted UK organizations with hotel interests”, *The Service Industries Journal,* Vol. 24, No. 5, pp. 31-51.

Phillips, P., Zigan, K., Santos Silva, M.M, and Schegg, R. (2015), “The interactive effects of online reviews on the determinants of Swiss hotel performance: A neutral network analysis”, *Tourism Management,* Vol. 50, pp. 130-141.

Racherla, P., Connolly, D.J, and Christodoulidou, N. (2013), “What determines consumers' ratings of service providers? An exploratory study of online traveler reviews”, *Journal of Hospitality Marketing and Management,* Vol. 22, No. 2, pp. 135-161.

Sainaghi, R., Phillips, P., and Zavarrone, E. (2017), “Performance measurement in tourism firms: A content analytical meta-approach”, *Tourism Management,* Vol. 59, pp. 36-56.

Singh, N., Hu, C. and Roehl, W. S. (2007), “Text mining a decade of progress in hospitality human resource management research: Identifying emerging thematic development”, *International Journal of Hospitality Management*, Vol. 26, No. 1, pp. 131-147.

Stepchenkova, S., Kirilenko, A. P. and Morrison, A. M. (2009), “Facilitating content analysis in tourism research”, *Journal of Travel Research*, Vol. 47, No. 4, pp. 454-469.

Stringam, B.B., and Gerdes, J.J. (2010), “An analysis of word-of-mouse ratings and guest comments of online hotel distribution sites”, *Journal of Hospitality Marketing and Management*, Vol. 19, No. 7, pp. 773-796.

Stringam, B. B., Gerdes, J.J., and Vanleeuwen, D.M. (2010), “Assessing the importance and relationships of ratings on user-generated traveler reviews”, *Journal of Quality Assurance in Hospitality and Tourism*, Vol. 11, No. 2, pp. 73-92.

Tan, H., Lv, X., Liu, X., and G., D. (2018), “Evaluation nudge: Effect of evaluation mode of online customer reviews on consumers’ preferences”, *Tourism Management*, Vol. 65, pp. 29-40.

Torres, E.N., Singh, D., and Robertson-Ring, Al. (2015), “Consumer reviews and the creation of booking transaction value: Lessons from the hotel industry”, *International Journal of Hospitality Management*, Vol. 50, pp. 77-83.

Viglia, G., Minazzi, R., Buhalis, D., (2016), “The influence of e-word-of-mouth on hotel occupancy rate”, *International Journal of Contemporary Hospitality Management*, Vol. 28, No. 9, pp. 2035– 2051.

Xiang, Z., Du, Q., Ma, Y., and Fan, W. (2017), “A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism”, *Tourism Management,* Vol. 58, pp. 51-65.

Xiang, Z., Gretzel, U. and Fesenmaier, D. R. (2009), “Semantic representation of tourism on the Internet”, *Journal of Travel Research*, Vol. 47, No. 4, pp. 440-453.

Xie, K., Zhang, Z, and Zhang, Z. (2014), “The business value of online consumer reviews and management response to hotel performance”, *International Journal of Hospitality Management*, Vol. 43, No. 1, pp. 1-12.

Xu, X., and Li, J. (2016), “The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach”, *International Journal of Hospitality Management*, Vol. 55, pp. 57-69.

Yahoo Finance (2019), *Fake hotel reviews? Never on Booking.com, says CEO* by Alanna Petroff, available at https://news.yahoo.com/fake-hotel-reviews-never-booking-com-says-ceo-173948451.html (accessed April 6, 2019),

Ye, Q., Law, R., and Gu, B. (2009), “The impact of online user reviews on hotel room sales”, *International Journal of Hospitality Management,* Vol. 28, No. 1, pp. 180-182.

Table 1: The Table reports, per sample city, the number of available independent hotel entities with at least one positive or negative online guest review on Booking.com, with financial data from the FAME database. ***, **, * indicate that a paired, two-sided *t*-test rejects the null hypothesis that the number of positive comments and the number of negative comments are from populations with equal means, while †††, ††, † indicate that a paired, two-sided Wilcoxon signed rank test rejects the null hypothesis that the difference between the median of positive comments and the median of negative comments is zero (at 1%, 5% and 10% significance levels, respectively).

	UK city												
	Birmingham	Bradford	Bristol	Cardiff	Edinburgh	Glasgow	Leeds	Liverpool	London	Manchester	Newcastle	Sheffield	Wolverhampton
Hotels with at least one review on booking.com	111	38	108	135	631	183	74	164	2,797	230	130	50	17
Excluding chained-brand hotels	77	29	87	116	552	158	57	145	2,082	202	113	39	13
Of which, Hotels that have data on FAME	8	3	4	6	38	10	5	21	198	15	11	3	3
	10.39%	10.35%	4.60%	5.17%	6.86%	6.33%	8.77%	14.48%	9.51%	7.43%	9.74%	7.69%	23.08%
Positive (+) guest reviews													
Mean	289.6***	166.6*	36.3***	219.9*	158.4***	381.8***	317.6***	135.5***	176.2***	198.4***	123.3***	111.0**	244.0*
Median	201.0†††	94.5††	36.0†††	83.0†††	71.0†††	278.0†††	189.0†††	62.0†††	116.0†††	177.0†††	89.5†††	127.0††	300.5
St. Dev.	250.2	191.2	19.8	305.8	247.9	281.8	341.0	171.9	255.7	142.4	130.9	93.7	132.5
Maximum	879	505	80	1,000	1,249	1,119	1,101	903	3,105	561	643	247	328
Minimum	16	8	12	62	2	15	53	3	1	12	4	3	47
Negative (-) guest reviews													
Mean	240.2***	108.9*	28.6***	160.1*	119.1***	318.5***	248.9***	101.8***	147.4***	160.0***	96.6***	86.0**	164.3*
Median	166.0†††	72.5††	27.0†††	57.0†††	54.0†††	253.0†††	146.0†††	48.0†††	100.0†††	145.0†††	62.0†††	72.0††	164.0
St. Dev.	214.8	117.4	18.7	216.3	188.4	243.2	274.4	138.6	219.0	110.9	107.0	84.2	87.5
Maximum	661	310	75	687	967	1,016	911	711	2,614	413	520	217	271
Minimum	7	8	10	36	1	11	35	2	0	9	3	1	58

Table 2: The Table summarizes the amount each theme contributes to the overall understanding of the text corpus based on the matrix of singular values (**S**). Panel A (B) reports the contribution of textual themes from Positive (Negative) online reviews/comments.

Panel A. Variance of Positive comments, explained by themes

Themes	Singular Value	EigenValue	Variance explained	Cumulative Variance explained
1	1.8357	3.3699	18.97%	18.97%
2	1.5321	2.3472	13.21%	32.18%
3	1.3233	1.7512	9.86%	42.03%
4	1.1580	1.3409	7.55%	49.58%
5	0.9056	0.8200	4.62%	54.20%
6	0.7420	0.5506	3.10%	57.30%
7	0.6775	0.4589	2.58%	59.88%
8	0.5575	0.3108	1.75%	61.63%
9	0.5393	0.2908	1.64%	63.27%
10	0.5215	0.2720	1.53%	64.80%
20	0.3776	0.1426	0.80%	75.12%
100	0.1191	0.0142	0.08%	91.97%
500	0.0235	0.0006	0.00%	99.53%
1000	0.0043	0.0000	0.00%	99.99%
1174	0.0000	0.0000	0.00%	100.00%

Panel B. Variance of Negative comments, explained by themes

Themes	Singular Value	EigenValue	Variance explained	Cumulative Variance explained
1	1.6524	2.7304	15.55%	15.55%
2	0.9350	0.8743	4.98%	20.53%
3	0.8167	0.6670	3.80%	24.33%
4	0.7156	0.5120	2.92%	27.24%
5	0.6896	0.4756	2.71%	29.95%
6	0.6318	0.3992	2.27%	32.23%
7	0.6154	0.3787	2.16%	34.38%
8	0.5962	0.3555	2.02%	36.41%
9	0.5849	0.3421	1.95%	38.36%
10	0.5708	0.3259	1.86%	40.21%
20	0.4204	0.1767	1.01%	54.04%
100	0.1637	0.0268	0.15%	82.92%
500	0.0364	0.0013	0.01%	98.71%
1000	0.0084	0.0001	0.00%	99.96%
1174	0.0000	0.0000	0.00%	100.00%

Table 3 The Table reports estimation results of Equation (2). The dependent variables are $PBTTA_{i,t}$ =*Profit bef. Tax*$_{i,t}$/$TA_{i,t}$ and $PBTS_{i,t}$ =*Profit bef. Tax*$_{i,t}$/$Sales_{i,t}$. *T*-statistics, estimated via standard errors clustered by the number of hotel 'stars', are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

Specification	(1)		(2)		(3)	
	Dependent variable		Dependent variable		Dependent variable	
	$PBTTA_{i,t}$	$PBTS_{i,t}$	$PBTTA_{i,t}$	$PBTS_{i,t}$	$PBTTA_{i,t}$	$PBTS_{i,t}$
Independent var.						
$PosTheme1_{i,t}$	10.661* (2.56)	23.398*** (18.00)	6.002* (2.75)	15.459*** (8.01)	5.675** (4.02)	11.791 (2.13)
$PosTheme2_{i,t}$	21.351* (2.63)	26.090*** (7.01)	2.845 (0.88)	8.955 (1.67)	6.605* (3.07)	20.892 (1.60)
$PosTheme3_{i,t}$	16.856*** (8.00)	21.948*** (6.67)	12.352 (2.33)	14.753** (3.71)	16.264*** (5.80)	25.992* (2.88)
$NegTheme1_{i,t}$	-15.192 (-1.58)	-51.066 (-1.24)	1.464 (0.10)	-25.287 (-0.55)	-2.549 (-0.20)	-20.235 (-0.49)
$NegTheme2_{i,t}$	-3.870 (-1.49)	-14.418 (-1.52)	-3.374 (-0.98)	-12.307* (-2.44)	-4.056 (-1.75)	-16.373 (-1.69)
$NegTheme3_{i,t}$	-6.206** (-3.43)	-5.124 (-0.94)	0.154 (0.05)	-0.156 (-0.04)	1.282 (0.54)	0.155 (0.04)
$Ln(TA_{i,t})$			0.101 (2.20)	-0.014 (-0.14)	0.087 (1.44)	-0.073 (-0.52)
$Lev_{i,t}$			-1.052*** (-39.67)	-0.275 (-2.34)	-1.066*** (-33.95)	-0.304 (-1.94)
$CA_{i,t}/CL_{i,t}$			0.047** (4.59)	0.140** (3.49)	0.049** (5.49)	0.1574* (2.49)
$City_i$	0.011 (0.34)	0.052* (2.36)	0.044*** (8.77)	0.024* (2.90)	0.043*** (7.04)	0.021 (1.75)
$Ln(No\ of\ Positive\ Reviews_{i,t})$	-0.687 (-1.66)	1.202 (1.54)	0.096 (0.45)	1.948 (2.03)	-0.492 (-0.87)	0.429 (0.69)
$Ln(No\ of\ Negative\ Reviews_{i,t})$	0.759 (1.99)	-0.894 (-1.40)	-0.117 (-0.57)	-1.751 (-2.27)	0.446 (0.78)	-0.247 (-0.34)
$Rating_{i,t}$					0.214 (1.48)	0.817 (0.85)
Intercept	-0.518** (-3.02)	-2.027 (-1.82)	-0.532 (-0.92)	-1.305*** (-9.28)	-1.905** (-3.71)	-7.244 (-1.02)
R^2	0.068	0.237	0.891	0.378	0.893	0.408
No. Obs	304	113	120	107	120	107

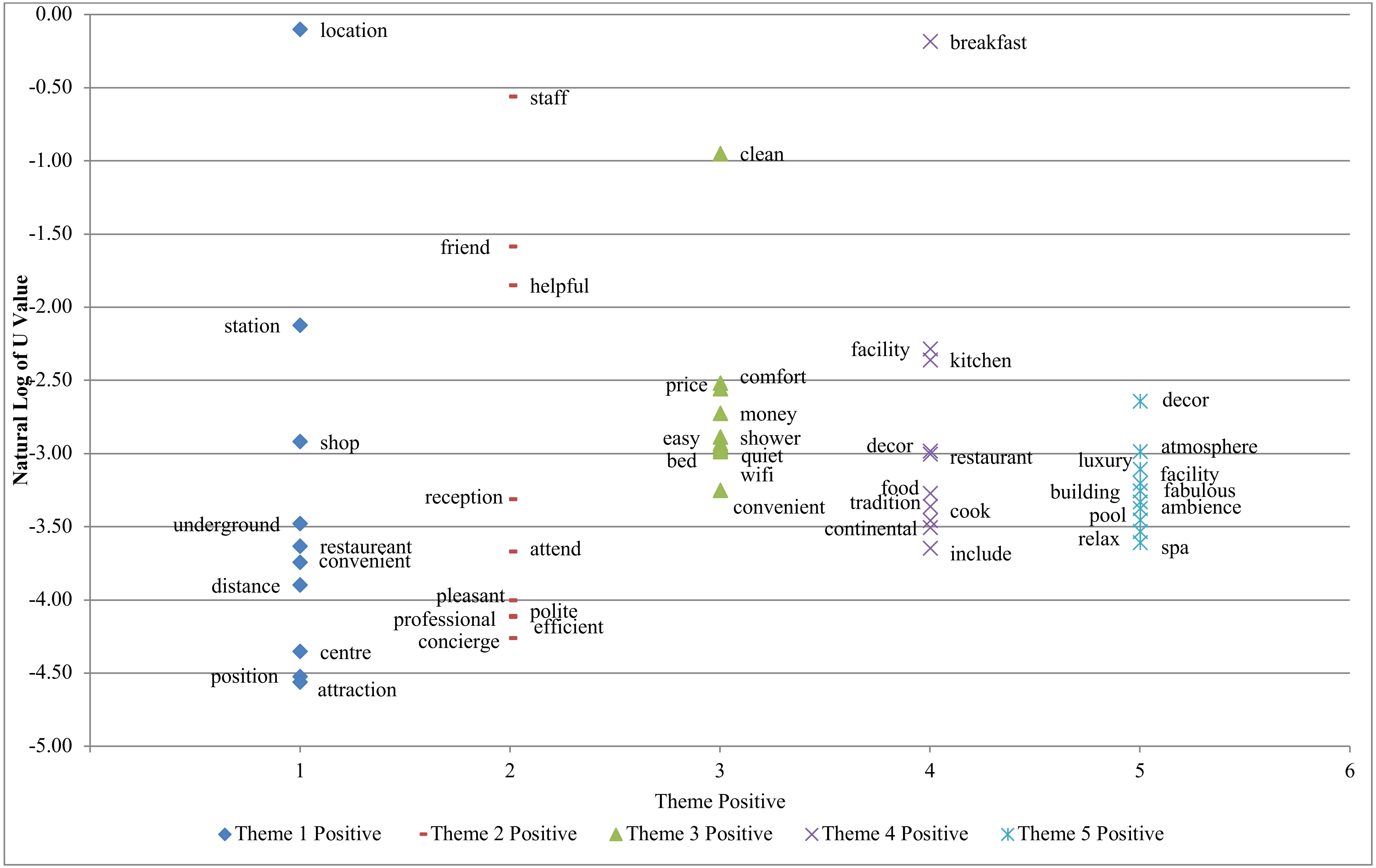
Table 4: The Table reports estimation results of Equation (2). The dependent variables are $PBTTA_{i,t}$ =*Profit bef. Tax*$_{i,t}$/$TA_{i,t}$ and $PBTS_{i,t}$ =*Profit bef. Tax*$_{i,t}$/$Sales_{i,t}$. *T*-statistics, estimated via standard errors clustered by the number of hotel 'stars', are reported in parentheses. ***, ** and * indicate significance at the 1%, 5% and 10% level, respectively.

	(4)		(5)	
	Dependent variable		Dependent variable	
	$PBTTA_{i,t}$	$PBTS_{i,t}$	$PBTTA_{i,t}$	$PBTS_{i,t}$
Independent var.				
$PosTheme1_{i,t}$	6.035**	10.956*		
	(3.69)	(2.41)		
$PosTheme2_{i,t}$	6.820	15.404		
	(2.02)	(1.24)		
$PosTheme3_{i,t}$	16.197**	22.448*		
	(3.66)	(2.51)		
$NegTheme1_{i,t}$			-1.532	-20.411
			(-0.14)	(-0.40)
$NegTheme2_{i,t}$			-2.996	-15.226
			(-1.07)	(-1.55)
$NegTheme3_{i,t}$			-3.799	-2.658
			(-1.58)	(-0.72)
$Ln(TA_{i,t})$	0.085	0.064	0.0302	0.141
	(1.75)	(0.54)	(0.45)	(1.46)
$Lev_{i,t}$	-1.052***	-0.296	-1.065***	-0.337*
	(-22.29)	(-2.04)	(-14.30)	(-3.02)
$CA_{i,t}/CL_{i,t}$	0.047***	0.174	0.059***	0.1782***
	(11.87)	(2.10)	(6.15)	(6.98)
$City_i$	0.047**	0.014	0.023	0.001
	(3.91)	(0.95)	(1.53)	(0.03)
$Ln(No\ of\ Positive\ Reviews_{i,t})$	0.430	0.587	0.600	0.398
	(0.74)	(1.33)	(1.22)	(1.45)
$Ln(No\ of\ Negative\ Reviews_{i,t})$	-0.381	-0.509	-0.671	-0.028
	(-0.68)	(-1.27)	(-1.46)	(-0.07)
$Rating_{i,t}$	0.165	0.624	0.127	0.790
	(0.93)	(0.71)	(0.65)	(1.58)
Intercept	-1.540	-5.316	-1.175	-7.258
	(-2.22)	(-0.88)	(-2.14)	(-1.67)
R^2	0.891	0.369	0.874	0.374
No. Obs	120	107	120	107

Figure 1: Characteristic terms for Themes 1 through 5

The Figure displays the ten most important terms, according to term contribution to themes (**U** matrix) for each of the first five themes, for positive (Panel A) and negative (Panel B) reviews.

Panel A, Positive Themes



Panel B, Negative Themes