Annals of Tourism Research Empirical Insights 6 (2025) 100199

Contents lists available at ScienceDirect

Annals of Tourism Research Empirical Insights

journal homepage: www.sciencedirect.com/journal/annals-of-tourism-research-empirical-insights

Exploring serial patterns in negative hotel reviews

Rodion Skovoroda [a,1], Wei Yang [b,2], Bowei Chen [b,3,*], Trevor Buck [b,4]

[a] The Open University Business School, Walton Hall, Milton Keynes, MK7 6BT, UK
[b] Adam Smith Business School, University of Glasgow, 2 Discovery Place, Glasgow, G11 6EY, UK

ARTICLE INFO

Handling Editor: Dr. Lorenzo Masiero

Keywords:
Customer experience
Negative online reviews
Structural topic model
Serial correlation
Hospitality industry

ABSTRACT

This study examines the dynamics of negative customer reviews in the hospitality industry. Drawing on a dataset of 80,521 negative hotel reviews from Booking.com, we identify eleven serially correlated aspects of customer experience where current customer complaints predict future ones. These patterns suggest the presence of emerging issues that impact multiple customers over time and show that the size of the time series effects in hospitality can be substantial and important in practice. We attribute these findings to capacity constraints and the relatively slow pace of managerial corrective actions. The empirical findings have implications for academic researchers, service organizations and online platforms, offering insights that can inform strategic decisions, enhance managerial performance evaluation and improve overall customer experience.

1. Introduction

In this research, we address the following question: How can personal service providers use serial patterns in online reviews to evaluate managerial performance and service quality, especially when service quality cannot be directly observed? The evaluation of managerial performance has always been a central challenge for business management and economic theorists. For example, Drucker's (1954) guiding principles of management involved monitoring, evaluating and rewarding and penalizing managerial employees. In economics, agency theorists (e.g. Jensen & Meckling, 1976) considered the problem of hidden action in cases where managers' efforts and behaviours cannot be fully observed, and the related costs of aligning the interests of managerial employees and shareholders. In transaction cost analysis, Williamson (1975) identified the challenge in terms of the difference between the consummate and perfunctory performance of opportunistic managers.

Personal service industries, in particular, are characterized by customer participation (Susskind, Kacmar, & Borchgrevinkm, 2003), the simultaneous production and consumption (Bowen & Schneider, 1988), low trialability (You et al., 2015) and intangibility (Bowen & Ford, 2002). Collectively, these characteristics imply that managerial behaviour, service quality and performance may not be directly observed, obscuring monitoring of managers. In this context, online customer reviews have emerged as a potential solution, providing free, publicly available service quality data (Bauman & Tuzhilin, 2022; Papathanassis & Knolle, 2011; Pelsmacker, Tilburg, & Holthof, 2018; Tsai, Chen, Hu, & Chen, 2020; Ye, Xia, Zhang, Zhan, & Li, 2022; Zhang, Wang, Law, & Han, 2024). Unlike offline word-of-moth, which is difficult for firms to access (Dellarocas, 2003), online reviews are instantly visible and searchable to both potential customers and service providers, at negligible cost.

When customers complain, *intended* managerial reactions may be revealed by their narrative responses, and an extensive body of research has analyzed managers' narrative responses (e.g. Chang, Ku, & Chen, 2020; Proserpio & Zervas, 2017; Tran, Nguyen, Van Huynh, & Stangl, 2025). However, *realized* responses, or corrective actions, may be different from narrative intentions and more important. Prior research shows that firms are concerned that online negative reviews may damage future sales (Feng, Li, & Zhang, 2019; Kim, Lim, & Brymer, 2015), and hence may have an incentive to fix the service quality issues underlying negative reviews to avoid future sales losses (Gu, Park, &

* Corresponding author.
E-mail addresses: rodion.skovoroda@open.ac.uk (R. Skovoroda), wei.yang@glasgow.ac.uk (W. Yang), bowei.chen@glasgow.ac.uk (B. Chen), trevor.buck@glasgow.ac.uk (T. Buck).

[1] Rodion Skovoroda is a Senior Lecturer in Finance at Open University, researching managerial incentives and corporate governance.
[2] Wei Yang is a Senior Lecturer in Strategy at University of Glasgow, focusing on competitive strategy and corporate governance.
[3] Bowei Chen is a Professor of Business Analytics and Artificial Intelligence at University of Glasgow, specializing in application of AI within business contexts.
[4] Trevor Buck is a Research Professor at University of Glasgow, studying platform models and firm strategies.

https://doi.org/10.1016/j.annale.2025.100199
Received 4 June 2025; Received in revised form 1 October 2025; Accepted 20 October 2025
Available online 1 November 2025

Konana, 2012). Yet, as noted earlier, the nature, pace and effectiveness of managerial corrective actions are often not directly observable.

This paper extends the literature in a new direction by arguing that, beyond signalling contemporaneous service quality, online reviews across consecutive customer cohorts can also reveal realized corrective actions through serial patterns such as autocorrelation. If managerial corrective actions influence subsequent reviews, then analysing the serial correlation of reviews allows researchers and owners to infer these actions indirectly, even when they are not directly observable.

Serially correlated customer complaints generally suggest a systematic relationship between the experiences of consecutive customer cohorts, revealing patterns that may indicate underlying service issues or operational inefficiencies. In particular, a positive serial correlation in the proportion of monthly customer complaints that are related to a particular area or topic means that high values for one month tend to be followed by high values, and low values tend to be followed by low values. Note that both ‘high’ and ‘low’ are defined relative to some long-term average trend. In this way, a positive correlation indicates a pattern of persistence or clustering in the data over time.

Serially correlated customer complaints can reveal service areas affected by time-variant resource requirements and capacity constraints (e.g., Daw, Hampshire, & Pender, 2025). Specifically, within the hospitality industry, consider negative feedback regarding hotel facilities such as the gym and pool. Such complaints will tend to peak during periods of high demand and frequent usage, which push these facilities to their capacity limits. Given that periods of high customer demand tend to cluster, it is likely that criticisms of hotel facilities will exhibit (some degree of) serial correlation.

However, the implications of serially correlated complaints extend beyond capacity constraints. By analysing this serial correlation, we can indirectly assess management’s effectiveness in addressing and responding to emergent issues. Taking another example from the hospitality industry, imagine a sudden spike in online reviewer complaints about unhelpful or unfriendly staff in a particular hotel recorded in month t, but this specific complaint consistently diminishes and returns to its long-term average level in month $t + 1$. It may then be inferred that managers successfully resolved the emergent underlying issue via corrective action, e.g. by organizing training sessions and/or replacing rude employees. On the other hand, serially correlated negative reviews about unhelpful or unfriendly service that take a relatively longer time to be resolved would indicate a persistent service failure and a degree of ineffectiveness on the part of the management.

Prior empirical studies have explored the serial correlation in overall customer satisfaction and review scores (Fang, Luo, & Jiang, 2013; Ravichandran & Deng, 2023). While the average review score provides useful information about overall customer experience, however, it may mask important heterogeneity across individual customer experience areas. For instance, changes in average scores over time might mask a number of simultaneous, distinct, and possibly divergent trends in individual customer experience categories. Moreover, the estimates of serial correlation in average review scores might reflect a composite mix of both stronger and weaker serially correlated customer experience categories. In contrast, recent studies have focused on sentiment analysis and topic classification of customer reviews, but these studies have not examined time-series dynamics in depth (Hu, Zhang, Gao, & Bose, 2019; Kolomoyets & Dickinger, 2023; Le, Phan-Thi, Nguyen, & Nguyen, 2025; Mensah, Odame, Ankrah, Obuobisa-Darko, & Hinson, 2025; Tsai et al., 2020; Yucel, Dag, Oztekin, & Carpenter, 2022).

Against this backdrop, our paper presents the first systematic analysis of serial correlation across a wide range of individual customer experience areas within a single industry – hospitality. Our empirical approach is summarised in Fig. 1. Using the Structural Topic Model (Roberts et al., 2014), we decompose negative reviews into distinct customer experience dimensions or categories (topics). We then analyze the prevalence of serial correlation across these topics using dynamic panel data techniques. This combination forms a unique longitudinal approach that is quick, cost-effective, and scalable.

While our method tracks service topics rather than individual customers over time, the strength of serial correlation tests based on panel data lies in their robustness to self-selection bias, noise (e.g., idiosyncratic customer preferences), and unobserved correlated heterogeneity in negative reviews.

Applying this methodology to 80,521 negative customer reviews from *booking.com*, we uncover several novel insights. First, we identify topics with statistically significant positive serial correlation in monthly complaints. These topics include feedback about room sizes, hotel location, pool and internet facilities, parking, and spatial aspects (e.g. view) which could arguably be attributed to capacity constraints in these areas. Secondly, and perhaps surprisingly, complaints about unhelpful and unfriendly staff also display strong positive serial correlation over a period extending up to five months, pointing to potential managerial inefficiencies in addressing underlying customer experience issues. Conversely, topics such as food and dining, cleanliness, bed quality, and maintenance exhibit serially uncorrelated complaints, suggesting that these issues are typically addressed effectively and do not persist over time. Finally, we find that unexpected customer experience problems are less likely to be resolved during peak hospitality seasons, emphasizing the role of seasonality in service responsiveness.

Our paper claims to make two main contributions. First, it develops a novel measure of serial correlation in customer reviews that, unlike existing approaches, explicitly accounts for the potential heterogeneity across individual categories or dimensions of customer experience. Managers can use this measure to better forecast the nature and intensity of future customer complaints based on past complaints, helping them

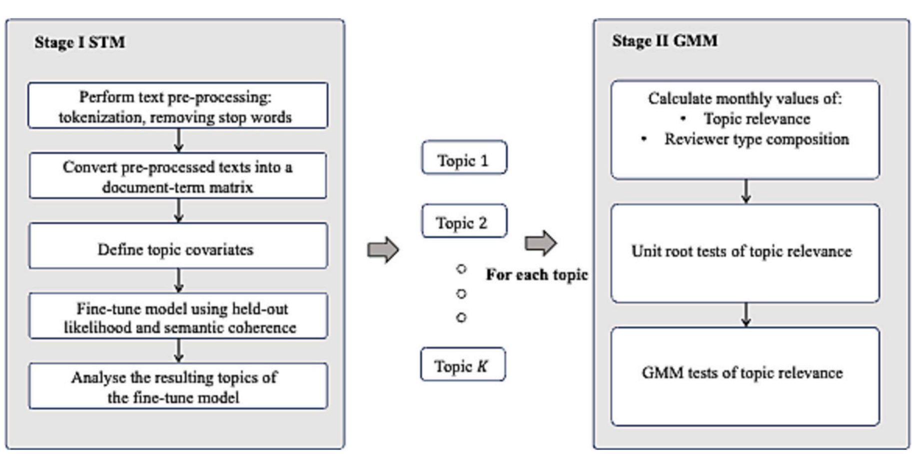

Fig. 1. Schematic view of the methodology.

prioritize corrective actions. For instance, managers that observe a recent spike in complaints can choose to allocate resources to areas that are likely to be affected by persistent, rather than transient (serially uncorrelated), problems.

Our methodology also enables business owners to assess the pace and effectiveness of managerial corrective actions, thereby enhancing strategic planning, managerial performance evaluation, and overall firm performance. Importantly, it can be readily customized to utilize both publicly available online reviews and the customer feedback collected in-house, improving the detection of persistent problems that need close attention.

Secondly, our empirical results show that the size of the time series effects in hospitality can be substantial and important in practice. Platforms such as *booking.com* could use our method to develop additional functionality based on the temporal component of negative reviews and promote this functionality to service providers eager to boost competitiveness. Digital information collected from online reviews may be particularly useful where the managers of branded chains may pool, compare and contrast review data across both units and time.

2. Literature review

The hotel accommodation industry is endowed with a mixture of physical facilities and personal service provision (Hu et al., 2019). Hotel services (and other tourism and hospitality services) are often built on interpersonal encounters between customers and employees, and service assessment must depend upon customers' subjective evaluations (Lee & Hu, 2004).

Compared with sellers' quality signals (e.g. advertising), customers' signals (e.g. reviews) have several distinguishing characteristics, e.g. an individual customer's information may be biased, as the evaluation of service quality is based largely on personal experience and perceptions (Blodgett, Granbois, & Walters, 1993; Conlon & Murray, 1996). As individual customers' experiences and perceptions vary, their evaluations on service quality may diverge (Parasuraman, Zeithaml, & Berry, 1988). In these circumstances, service providers may often need to attend to conflicting individual signals (e.g. complaints) from different customers simultaneously. However, large numbers of online reviewers pool together a multiplicity of evaluations (Wang, Wezel, & Forgues, 2016), and researchers have been able to study the volume of online reviews (Chevalier & Mayzlin, 2006), their valence (Chevalier & Mayzlin, 2006), dispersion (Godes & Mayzlin, 2004), and content (Yang, Ren, & Adomavicius, 2019).

Recent tourism and hospitality studies have increasingly applied text mining and topic modeling algorithms, which typically interpret customer reviews as a combination of multiple topics (e.g. Hu et al., 2019; Kolomoyets & Dickinger, 2023; Le et al., 2025; Mensah et al., 2025; Tsai et al., 2020; Yucel et al., 2022; also see papers reviewed in Chang et al., 2020). These computer-based methods are particularly valuable for analysing large text corpora, while facilitating transparency and replicability of findings. For recent reviews of modern topic modeling methods, see Egger and Yu (2022) and Kirilenko and Stepchenkova (2025).

Online hotel reviews reveal various customer and hotel characteristics, signal service quality and directly affect subsequent customer decisions (Gao, Li, Liu, & Fang, 2018; Liu, Teichert, Rossi, Li, & Hu, 2017; Pantano, Priporas, & Stylos, 2017; Sparks & Browning, 2011). At the same time, their visibility and searchability provide valuable feedback to sellers (firms, hotels). This is especially important for managers in large, hierarchically-organized firms (e.g. branded chains of hotels or restaurants), who may be removed from day-to-day service provisions. In this context, the explosion of digital communications, social media and online reviews has emerged as a valuable tool for both potential customers and sellers of personal services.

Our research focuses on serial correlation in customer experience that emerges across consecutive customer cohorts. This form of serial correlation differs from the within-customer dynamics typically analyzed in longitudinal studies. Longitudinal designs usually rely on repeatedly surveying the same customers across multiple touchpoints or transactions (Bolton, Lemon, & Bramlett, 2006; Hogreve, Bilstein, & Mandl, 2017; McColl-Kennedy, Zaki, Lemon, Urmetzer, & Neely, 2019; Sriram, Chintagunta, & Manchanda, 2015; Voorhees, Beck, Randhawa, DeTienne, & Bone, 2021). These designs are well suited to capturing the within-customer dynamics, such as the evolution of complaints, commendations and termination behaviour. By contrast, an analysis of managerial corrective actions requires examining experiences among different customers who enter the service relationship in successive periods. Despite the established contributions of longitudinal studies, the serial correlation of customer experiences across consecutive customer cohorts remains under-researched.

While managerial corrective actions at facility level are usually not directly observable, a related literature investigates managers' narrative responses to negative online reviews that could, in some cases, outline the firms' *intended* corrective actions (e.g. Chang et al., 2020; Kim et al., 2015; Proserpio & Zervas, 2017; Ravichandran & Deng, 2023; Tran et al., 2025). A weakness of this approach, however, is that it may reproduce the longstanding distinction between managers' intended and realized changes (Mintzberg & Waters, 1985). We see our work as complementary to this literature, in that it can help predict future customer complaints and identify service areas that may require further corrective action.

3. Data and sample

The hotel industry has been acutely concerned with the problem of fake reviews, with unscrupulous hotels offering incentives to customers or even writing positive reviews themselves (Ananthakrishnan, Li, & Smith, 2020; Lappas, Sabnis, & Valkanas, 2016; Tuomi, 2021). However, certain booking platforms have taken steps to authenticate them, and in any case we focus on negative reviews. For authenticated reviews, we chose *booking.com* for its verified, paid-purchase reviews and secure storage (Rodriguez-Diaz, Rodriguez-Diaz, & Espino-Rodriguez, 2018). Its review form also includes a dedicated section for negative aspects, eliminating the need to classify reviews manually into positive and negative. For example, a customer might write 'good location' in the section dedicated to positive reviews, while the same customer could also write that '…there were some maintenance repairs going on, but this didn't affect me much' in the section dedicated to negative reviews.

Our initial dataset was obtained from *Kaggle*, a well-established crowdsourcing platform. The data contained 515,000 *booking.com* reviews across 1493 hotels in six European cities (Amsterdam, Barcelona, London, Milan, Paris and Vienna), covering 25 consecutive months, August 2015 to August 2017. The dataset contained the text of negative reviews (i.e. responses left in the feedback-form section dedicated to negative aspects of customer stay), review date, hotel location and customer characteristics (i.e. single occupants, couples, families with young children, families with older children, groups). Hotel brand and group information was verified against STR's list of registered hotels. *STR* is the leading provider of performance benchmarking and comparative analytics for the global hotel industry, widely used in both academic research and industry practice. After discarding hotels without identifiable brand or group information, the final sample comprised 80,521 negative reviews across 271 hotels, which we used for the first step of our methodology—the textual analysis of negative reviews.

The hospitality and tourism industries, characterized by seasonal demand and labour intensity (e.g., Jang, 2004), may exhibit varying degree of customer experience serial correlation across hospitality seasons. We therefore set the *High Season* dummy equal to 1 for April through August in the northern hemisphere, reflecting the months with the highest average hotel review volumes in our sample, and 0 otherwise. Table 1 shows descriptive statistics for our sample.

Table 1
Descriptive statistics, individual customer reviews.

Variable	N	Mean	Std. Dev.	Min	Max
Time (days since 3 August 2015)	80,521	382.49	208.08	1	731
HighSeason	80,521	0.469	0.499	0	1
Customer Type:					
Couple	80,521	0.4673	0.4989	0	1
Family with older children	80,521	0.0557	0.2294	0	1
Family with younger children	80,521	0.1424	0.3495	0	1
Group	80,521	0.1370	0.3439	0	1
Solo traveler	80,521	0.1929	0.3946	0	1

4. Description of methods

Online reviews are a free, publicly available source of high frequency, consumer-generated information where each review has a digital 'time stamp'. (In contrast, traditional, offline, word-of-mouth reviews often cannot be reliably linked to a particular date or month, which also limits their applications). This digital 'time stamp' allowed us to use inferentially well-established time series and panel data techniques to investigate serial correlation in reviews.

Before online review data can be used for inference, researchers must address four key theoretical and methodological challenges. Firstly, customer experience is often complex and multi-dimensional. Secondly, online reviews may suffer from self-selection bias (Hu, Pavlou, & Zhang, 2017), where customers who write reviews are not selected at random but rather self-select based on observed and unobserved personal characteristics. Thirdly, customers might have unique preferences, leading to idiosyncratic reviews and 'noisy' data. Finally, time series review data could be non-stationary, which could potentially lead to cases of spurious regression in applied research (e.g., Wooldridge, 2012, p. 644).

The first step in our methodology, described in sections 4.1 and 4.2, tackled the issue of multi-dimensionality by probabilistically classifying negative reviews into those related to distinct issues or topics, while ensuring consistency, transparency and replicability of our results. The second step, based on dynamic panel data tests described in sections 4.3–4.5, was designed to take into account self-selection bias, noise, and potential non-stationarity considerations.

4.1. *The structural topic model*

Kirilenko and Stepchenkova (2025) compare different topic models in tourism research, noting that while transformer-based models such as BERT and GPT perform well with noisy or short texts, they face challenges of explainability and replicability. By contrast, LDA is widely used in tourism for its interpretability, highlighting the value of probabilistic topic models. Our study employs the STM, a state-of-the-art generalization of LDA that incorporates metadata and allows explicit control over the number of topics – features essential for our panel regression design. These insights from Kirilenko and Stepchenkova (2025) support our methodological choice.

The mathematics details of model estimation and validation are explained by Roberts et al. (2014), while Hu et al. (2019) provides further overview. Consider a collection of $j = 1, \ldots, J$ negative reviews. Review d_j contains L_j words, which are a mixture over K topics. As a part of its output, the Structural Topic Model identifies a set of weights $0 < \rho_{j,k} < 1$, that represent the proportion (i.e., relevance) of topic k in review d_j. For each individual review, the total sum of topic weights across all topics equals to 1 by construction, i.e., $\sum_{k=1}^{K} \rho_{j,k} = 1$. Unlike the widely used Latent Dirichlet Allocation (LDA), the Structural Topic Model offers a way to make topic selection dependent on a set of predefined covariates.

In this study, we deploy the Structural Topic Model by using the stm R package (Roberts, Stewart, & Tingley, 2019) with covariates brand, group, room, trip and customer type, consistent with prior applications of STM in hospitality research (Hu et al., 2019). It should be noted that standard natural language preprocessing steps were applied, including tokenization, removal of punctuation, numbers, stopwords, stemming, and conversion of all terms to lowercase. These procedures helped standardize the vocabulary prior to estimating the STM.

4.2. *Choosing a suitable number of topics (K): goodness of fit vs. interpretability*

While the topic-generating procedure in the Structural Topic Model can be run for any number of topics K, the choice of the 'right' value of K involves a trade-off between the model's predictive power (as measured by the held-out likelihood) and the average semantic coherence of the model's topics (Mimno, Wallach, Talley, Leenders, & McCallum, 2011; Wallach, Murray, Salakhutdinov, & Mimno, 2009). Semantic coherence of an individual topic is high if the most frequent words in this topic frequently co-occur together. Models that produce topics with high semantic coherence are desirable as it has been shown to aid topic interpretability by human domain experts (Mimno et al., 2011).

Figure 2 presents the values of held-out likelihood and semantic coherence for a range of K values from 5 to 50, consistent with the topic modeling settings in Chen, Anker, and Liang (2025). The results show that more complex models (i.e., higher K) achieve a better statistical fit but, on average, generate topics with lower semantic coherence, making them more difficult to interpret reliably by researchers and managers (Mimno et al., 2011; Wallach et al., 2009).

Figure 3 further illustrates the resulting trade-off between held-out likelihood and semantic coherence. Fig. 3(a) scatterplots the held-out likelihood versus negative semantic coherence for models with different Ks. While no single Structural Topic Model can simultaneously generate the highest goodness of fit and semantic coherence, Fig. 3(b) highlights a sub-set of models that forms the Pareto frontier (i.e., models that produce the highest values of held-out likelihood for a fixed level of semantic coherence). In particular, Fig. 3(b) shows that the Pareto frontier includes models fitted for K = 27 and K = 32. Importantly, while the Pareto frontier is relatively steep for lower values of K (K < 27), it seems to flatten out for higher values of K (K > 32). This means that a further increase in K beyond K = 32 does not add much explanatory power per unit loss in model semantic coherence (interpretability). Based on this analysis, the rest of the paper uses values K = 27 (main results) and K = 32 (robustness tests, available upon request) to generate and interpret the topics.

Table 2 lists these topics and their descriptive labels, ranked by average weights. The identified topics are broadly consistent with Hu et al. (2019). The most highly weighted topics are 'Room size', 'Food and dining', 'Location & pool', and 'Bathroom experience, cleanliness'.

4.3. *Dynamic panel data tests*

For the second step of our methodology – the analysis of serial correlation in customer signals using dynamic panel regressions – we aggregated individual review scores and Structural Topic Model generated review topics into monthly series as follows. For each topic, let $\rho_{i,t,j}$ denote the topic weight in individual reviews j received by hotel i in month t. The monthly average *Topic Relevance*$_{i,t}$ is defined as $\frac{1}{J}\sum_{j=1}^{J} Log(\rho_{i,t,j})$. While individual values of raw weights (ρ) are constrained within the range of 0 to 1, they typically exhibit positive skewness. Applying a logarithm transformation to the topic weights ensured that *Topic Relevance*$_{i,t}$ had a symmetric bell-shaped distribution. To ensure that our results were not affected by outliers, panel regressions only include hotel/month pairs that report ten or more reviews. The final unbalanced panel dataset comprises 2588 hotel-month observations across 199 hotels, after removing hotels with fewer than 10

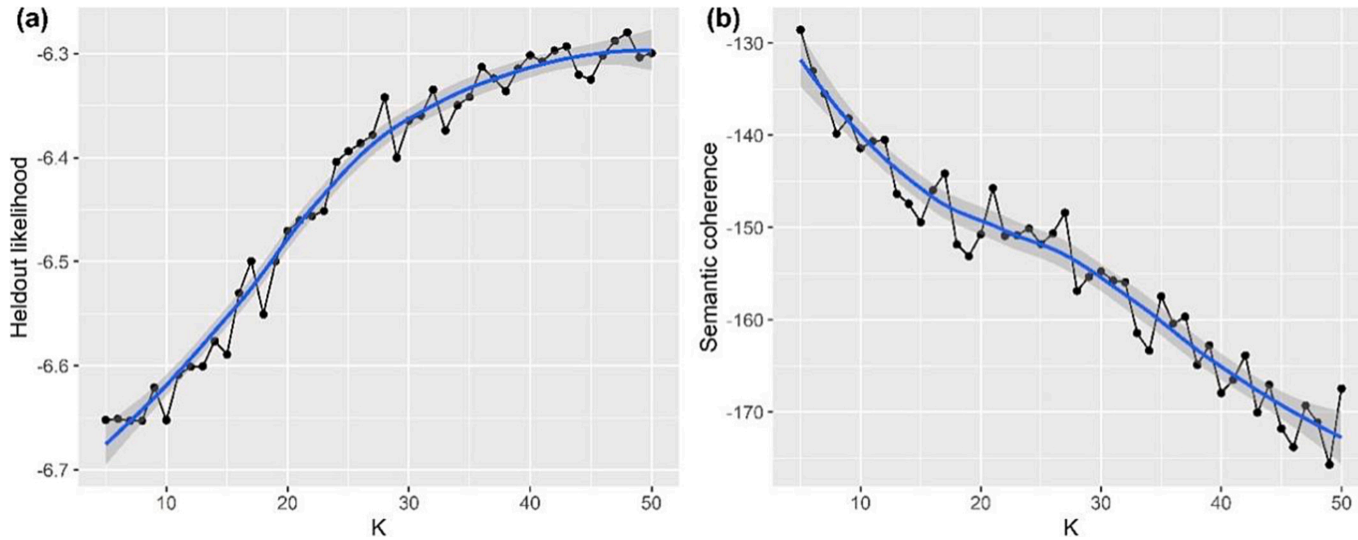

Fig. 2. Evaluation metrics vs. K: (a) held-out likelihood and (b) semantic coherence.

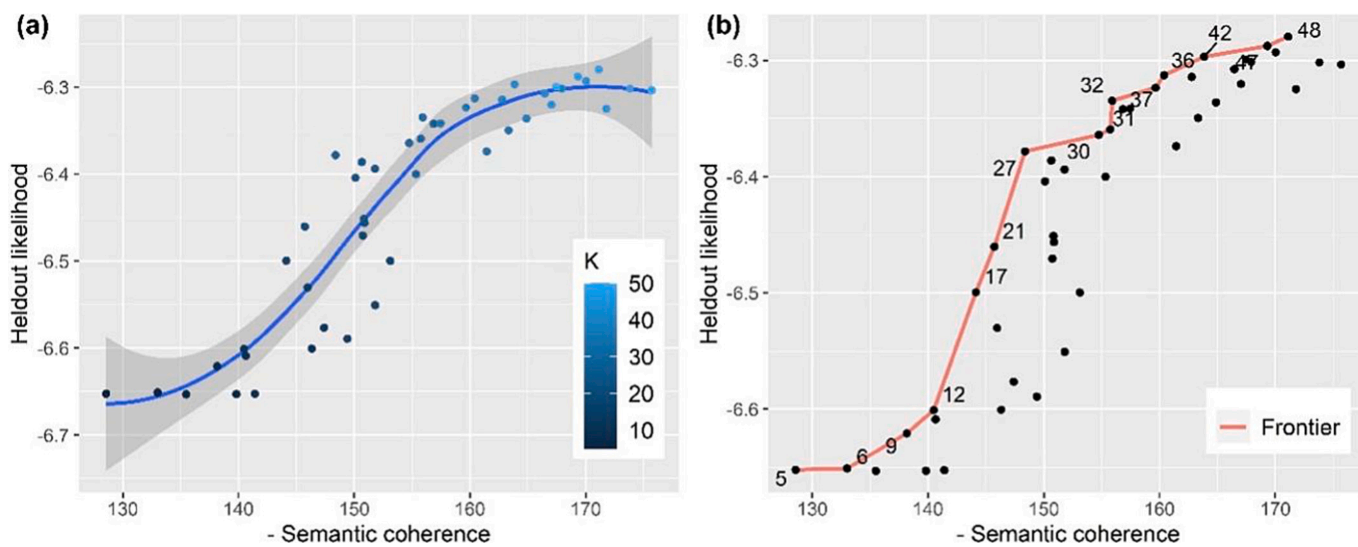

Fig. 3. Held-out likelihood vs. negative semantic coherence: (a) scatter plot with smoothing trend line and confidence interval; and (b) scatter plot with Pareto frontier.

observations from the analysis.

We first tested for evidence of unit roots and non-stationarity in *Topic Relevance*. This helped detect highly persistent signals, such as a random walk, where the correlation between present and future signals does not decay with time. More specifically, we used the modified Fisher-type panel unit-root tests[5] (based on the augmented Dickey-Fuller and, separately, Phillips-Perron tests), which are designed to handle unbalanced datasets where we have more panels (hotels, $N = 199$) than time periods ($T \leq 25$). The tests follow the null hypothesis of unit roots in all panels, and the alternative is that at least one unit in the panel is stationary.

For stationary customer signals where the unit-root tests rejected the presence of unit roots, we estimated the impulse response functions using the following linear dynamic panel data model:

$$\text{Topic Relevance}_{\text{i,t}} = \sum_{\text{j}=1}^{\text{p}} \alpha_{\text{j}} \text{Topic Relevance}_{\text{i,t}-\text{j}} + \sum_{\text{j}=1}^{\text{q}} \widehat{\alpha}_{\text{j}} \text{High Season}_{\text{t}-\text{j}} \times \text{Topic Relevance}_{\text{i,t}-\text{j}} + \sum_{\text{j}=0}^{\text{l}} \beta_{\text{j}} \text{x}_{\text{i,t}-\text{j}} + \text{Month}_{\text{t}} \gamma + \text{u}_{\text{i}} + \text{e}_{\text{i,t}}, \tag{1}$$

where $i = 1, \ldots, N$ indexes hotels, $t = 1, \ldots, T$ denotes months, p is the number of autoregressive lags in low seasons, while q lags were used to test whether *Topic Relevance* had stronger or weaker serial correlation in high seasons compared to low seasons.

Eq. (1) focuses on the sign and significance of autocorrelation coefficients α_j and $\widehat{\alpha}_j$. Endogenous covariates, x, include monthly review shares by customer types (e.g., Couples, Families, Groups) and their lags. The dummy vector $Month_t$ accounts for time effects, while unobserved, hotel-specific heterogeneity, u_i, may correlate with explanatory

[5] We used *Stata* command *xtunitroot* with *dfuller drift* and *pperron* options. Option *demean* is used to mitigate the impact of cross-sectional dependence in panels (Levin, Lin, & Chu, 2002).

Table 2
Topics and tests for serial correlation in monthly average Topic Relevance.

Topic	Label	Top words	*p*, *q*, *l* – autoregressive lags selected based on the Model Selection Criterion using the Hannan-Quinn Information Criterion	Low season Chi-square	High season Chi-square
1	Room size	Room, small, bit, little, tire, basic, smoke, extrem, bigger, smaller	4, 4, 0	***	***
2	Food and dining	Breakfast, expens, poor, food, restaur, includ, tabl, quality, limit, option	1, 1, 0		*
3	Location & pool	Pool, locat, can, area, close, far, walk, bit, citi, centranc	5, 1, 0	***	***
4	Bathroom experience, cleanliness	Bathroom, shower, need, smell, toilet, bath, door, carpet, date	3, 1, 0		
5	Neutral/ positive	Noth, like, reali, old, everyth, think, anyth, perfect, say, didnt	1, 1, 0		
6	Noise	Floor, noisi, nois, door, next, peopl, stay, outsid, lot, earli	1, 1, 0		
7	Bed quality	Bed, doubl, hard, two, size, pillow, comfort, uncomfot, single, bedroom	1, 1, 0		
8	Air quality & temperature	Work, window, air, hot, open, cold, condit, heat, proper, turn	5, 2, 0		***
9	General experience and value	Hotel, price, bad, great, star, renov, londo, recommend, consid, visit	7, 6, 0	***	***
10	Reception and check-in times	Time, recept, checkin, wait, will, call, front, never, desk, long	1, 1, 0		
11	Coffee & tea-making facility	Coffee, light, tea, make, enough, dirti, provid, lack, dark, bottle	4, 1, 0		**
12	Payment issues	Stay, pay, check, told, money, want, said, card, know, return	1, 1, 0		
13	Reservation issues	Book, ask, charg, arriv, hour, paid, per, offer, com, differ	1, 1, 0		
14	Internet and parking	Just, wifi, park, free, find, issu, feel, car, leav, well	1, 1, 0	*	**
15	Unhelpful or unfriendly staff	Staff, help, problem, friend, rude, late, experi, found, member, weren	5, 3, 0	**	***
16	Unresolved requests	One, day, first, lift, request, another, given, second, end, two	1, 2, 0		***
17	Room change and upgrade requests	Room, clean, change, avail, standard, upgrade, towel, new, done, iron	7, 1, 0	***	***
18	Maintenance issues	Night, water, morn, thing, everi, come, terribl, kept, fix, either	1, 3, 0		
19	Spatial aspects	View, wall, tini, move, space, side, see, around, face, insid	8, 6, 0	***	***
20	Guest information	Use, guest, Euro, busi, person, travel, inform, luggag, entrance, possible	1, 1, 0		
21	Slow and overpriced services	Service, good, slow, order, English, value, fine, overpr, otherwis, horribl,	1, 3, 0		**
22	Misc	Get, back, drink, took, wen, although, came, keep, best, overall	1, 1, 0		
23	Gym facilities	Look, nice, facil, place, rather, mani, gym, picture, slight, run	1, 6, 0		

(*continued on next page*)

Table 2 (*continued*)

Topic	Label	Top words	p, q, l – autoregressive lags selected based on the Model Selection Criterion using the Hannan-Quinn Information Criterion	Low season Chi-square	High season Chi-square
24	Unmet expectations	Better, much, expect, high, especi, furniture, rate, mayb, less, definit	1, 1, 0		
25	Misc	Even, also, got, though, made, always, didn, someone, sure, away	1, 1, 0		
26	Misc. facilities	Bar, quit, seem, felt, lobbi, top, love, roof, seat, rest	1, 2, 0		
27	Misc	Extra, disappoint, tri, cost, howev, custom, someth, almost, concierge, complet	2, 1, 0		

a) We implemented the Structural Topic Model uses stemming for word reduction in text processing. Words like "locate" and "location" share the same root "locat". Stemming ensures that such variations are treated as the same word, improving the accuracy and focus of topics. b) * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

variables. It is well-recognized (Nickell, 1981) that Ordinary Least Squares, fixed-effect estimators, and random-effect estimators produce biased results when the estimated equation contains both lagged dependent variables and correlated unobserved heterogeneity, as does Eq. (1). We therefore estimated Eq. (1) using a system Generalized Method of Moments estimator that produces consistent and efficient results by using lagged values and differences in lagged values as instruments for the endogenous effects (Arellano & Bond, 1991; Blundell & Bond, 1998; Roodman, 2009). More specifically, we used the Generalized Method of Moments estimator implemented in a user-written *Stata* package *xtdpdgmm* (Kripfganz, 2019).

4.4. Endogenous covariates and self-selection Bias

In this section, we discuss ways in which our methodology took into account self-selection bias in the reviewer type composition. In testing for serial correlation in customer experience, it is important that we control for exogenous sources of heterogeneity. To this end, regression covariates x included the share of monthly reviews written by *Couples*, *Family with older children*, *Family with younger children*, *Groups* and *Solo* and, potentially, their lags of up to degree l. If some customer types are more sensitive to customer experience issues than others, an exogenous variation in the composition of customer types is likely to affect review scores. However, since customers self-select to write reviews, we cannot expect the observed composition of customer types that end up writing reviews to fully reflect the true composition among all customers that used the hotel's services in a period. This resulting self-selection bias in the reviewer type composition meant that both fixed-effect estimators and random-effect estimators, if applied to Eq. (1), would produce biased results. The Generalized Method of Moments approach to dynamic panel data uses instrumental variable techniques based on the past realization of endogenous covariates to estimate the true unbiased effects (Arellano & Bond, 1991; Blundell & Bond, 1998). In this paper, we instrumented endogenous covariates by using time lags 2 to 11.

4.5. Generalized method of moments model selection

Linear dynamic panel data models such as (1), when estimated by Generalized Method of Moments, are powerful and flexible tools that are designed to interrogate data plagued by unobserved heterogeneity correlated with included covariates. For this reason, valid Generalized Method of Moments applications require careful implementations of specification checks. First, researchers have to test for the AR(2) autocorrelation in the first-differenced residuals, as such correlation could make the model invalid (Roodman, 2009). Secondly, researchers have to check the adequacy of the instrumental variables employed using the Sargan-Hansen tests of the overidentifying restrictions (Roodman, 2009). Invalid instruments should be identified and discarded. An alternative approach, instead of discarding ostensibly invalid instruments outright, involves expanding the regression model by incorporating additional (longer) lags or previously excluded variables. This is motivated by the recognition that omitted variables, including higher-order lags of already incorporated variables and other excluded factors, can introduce correlation between the instruments and the error term (Kripfganz, 2019).

An important consideration in the implementation of dynamic panel data model (1) is the choice of lag lengths p, q and l. Our results are based on the following 6-step general-to-specific model selection procedure suggested by Kiviet, Pleus, and Poldermans (2017) and Kripfganz (2019): (1) we start with lag lengths $p = q = 8$ and $l = 1$, representing the most general model we can feasibly estimate (higher values of p and q reduce the sample size and risk increasingly unreliable results); (2) for the pre-specified values of p, q and l, calculate the iterated Generalized Method of Moments estimator with cluster-robust standard errors and collapsed instruments designed to keep the number of instruments down; (3) deploy the autocorrelation and the Sargan-Hansen specifications tests described above; (4) reduce the value of p, q or l by 1 by removing the longest lags, and then repeat steps (2) and (3) until we reach the most restrictive model with $p = q = 1$ and $l = 0$; (5) for all candidate models that pass the specification tests, calculate the model and moment selection criteria (MMSC) suggested by Andrews and Lu (2001) based on Hannan and Quinn (1979) (MMSC-HQIC) and, alternatively, on Bayesian (MMSC-BIC) statistics (note that Andrews and Lu (2001) recommend against using MMSC-AIC criteria in Generalized Method of Moments settings); (6) finally, select the model (i.e. select values p, q and l) that corresponds to the lowest value of the MMSC statistics. With our data and sample, the two alternative module selection criteria, MMSC-BIC and MMSC-HQIC, produced consistent results.

5. Analysis of results

The augmented Dickey-Fuller and Phillips-Perron tests confirmed stationarity of *Topic Relevance* for all topics, rejecting the unit root hypothesis across all lag lengths (inverse normal p-value: 0.000; modified inverse chi-squared p-value: 0.000). This allowed us to use eq. (1) to model serial correlation in *Topic Relevance*.

Table 2, column four, presents the results of the model selection procedure with respect to lags p, q and l described in the previous section. Columns five and six report the results of the Chi-square tests for serial correlation across seasons. Specifically, we tested the joint significance of all selected lags in low and, separately, high seasons. Our results suggest that the following six topics exhibited significant serial correlation in both low and high seasons: 'Room size', 'Location and pool', 'General experience and value', 'Unhelpful or unfriendly staff', 'Room change and upgrade requests', and 'Spatial aspects'. An additional five topics, including 'Air quality and temperature' and 'Internet and parking', showed significant correlation in high seasons only. Overall, 11 topics revealed predictive patterns for future complaints, while 16 topics lacked significant serial correlation.

For illustrative purposes, Table 3 presents the full details of the Generalized Method of Moments analysis for two individual topics – Topic 1 (Room size) and Topic 4 (Bathroom experience, cleanliness). For Topic 1, the model selection procedure retained lags $p = 4$, $q = 4$, and $l = 0$. Statistically significant regression results suggested the presence of a mean-reverting process in *Topic Relevance* over a period extending up to four months, both in low and high seasons. Furthermore, we did not find that the serial correlation in customer complaints was stronger or weaker in high seasons than in low seasons. This was supported by the joint test of the four lags of *Topic Relevance*High Season* showing statistically insignificant results (Chi2(4) = 4.16, *P*-value = 0.38). Table 3 further reports that the models passed the required specification tests. We thus classify Topic 1 as serially correlated.

For Topic 4, the model selection procedure retained lags $p = 3$, $q = 1$, and $l = 0$. Statistically insignificant regression results suggested that this topic was not serially correlated. This was supported by joint tests of lag terms both in low and, separately, in high seasons. Table 3 further reports that the models passed the required specification tests. We thus classify Topic 4 as serially uncorrelated.

Table 3
Monthly Topic Relevance, Generalized Method of Moments results[a,b] (robust standard errors in parentheses).

Dependent variable: *Topic Relevance*	Topic 1 (Room Size)	Topic 4 (Bathroom experience, cleanliness)
p, q, l – AR lags retained based on the MMSC-HQIC statistics	4, 4, 0	3, 1, 0
Topic relevance (t-1)	0.376*** (0.073)	−0.089 (0.177)
Topic relevance (t-2)	0.279*** (0.071)	−0.018 (0.154)
Topic relevance (t-3)	0.143*** (0.041)	−0.047 (0.105)
Topic relevance (t-4)	0.191*** (0.039)	
Topic relevance (t-1) * High Season (t-1)	−0.049 (0.043)	0.172* (0.104)
Topic relevance (t-2) * High Season (t-2)	0.036 (0.048)	
Topic relevance (t-3) * High Season (t-3)	−0.033 (0.050)	
Topic relevance (t-4) * High Season (t-4)	−0.064 (0.045)	
Share of monthly reviews written by:		
Couples (t)	−0.28 (0.315)	0.221 (0.508)
Families with older children (t)	−0.473 (0.479)	0.449 (2.02)
Families with young children (t)	0.35 (0.269)	1.31** (0.552)
Groups (t)	0.013 (0.402)	0.441 (0.441)
Solo	Base	Base
Const	−0.047 (0.589)	−4.08 (1.83)
Time (month) dummies	YES	YES
N obs.	2145	2252
N groups (hotels)	188	194
N instruments	97	98
Test for AR(2) autocorrelation in the first-differenced residuals	Prob>z = 0.807	Prob>z = 0.181
Sargan-Hansen test of the overidentifying restrictions	Chi2(64) = 59.43 Prob>chi2 = 0.639	Chi2(64) = 69.39 Prob>chi2 = 0.43

a) All variables except period dummies are assumed endogenous, instrumented using time lags 2 to 11. b) * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

5.1. Is the size of time series effects important in practice?

Figures 4 and 5 depict impulse response functions for serially correlated topics during high and low seasons, as calculated based on Eq. (1). The figures illustrate the impact of a 100 % unexpected increase in topic weights at month 0 on subsequent *Topic Relevance*. While the effect size varied across individual topics and seasons, the results showed that the overall time series effect can be substantial. Within the first two months following the initial shock, complaints about 'Room size' and 'General experience and value' could remain up to 30 % above their trend. Similarly, complaints about 'Spatial aspects' and 'Air quality' could stay up to 25 % higher, while complaints about 'Location and pool', 'Unhelpful or unfriendly staff', and 'Room change and upgrade' could stay up to 20 % higher. Additionally, complaints related to 'Slow and overpriced services', 'Internet and parking', and 'Unresolved requests' could stay up to 15 % higher. Interestingly, Figs. 4 and 5 indicate that 'Coffee & tea-making facility' was the only topic that showed negative serial correlation, likely due to managerial over-correction. Overall, our estimates suggest that the size of time series effects in customer complaints could be substantial and informative to the managers and potential customers.

5.2. Alternative specifications and robustness

A potential limitation of the methodology outlined so far is that it requires multiple testing across a relatively large number of K = 27 topics. In individual tests, we usually reject the null hypothesis of no serial correlation if the corresponding *p*-value is below 5 or 10 % cut-off level. With K = 27 tests however, we could expect one or two tests to fall into the rejection area purely by chance. Since our choice of the value K was motivated by the trade-off between goodness of fit and semantic coherence of the Structural Topic Model generated topics, we advocate running the analysis for a few alternative values of K from the Pareto frontier (Fig. 3b). This can guard against the risks of getting chance results.

The supplementary material reports results of alternative specifications using K = 32 topics (see Table S1). The two specifications, K = 32 and K = 27, yield complementary results. Both highlight potential inefficiencies related to 'Unhelpful and unfriendly staff'. Additionally, topic 'Mismanaged orders', identified as serially correlated for K = 32, mirrors the findings of serially correlated 'Unresolved requests over multiple days' and 'Slow and overpriced services' for K = 27, confirming the robustness of our results. However, the two specifications differ in their findings regarding complaints about reservation and payments. Specifically, while K = 27 identifies 'Reservation' and 'Payment' as two distinct topics and finds them individually transient, K = 32 combines them into a single 'Reservation and Payment' topic, which exhibits serial correlation.

This suggests that our methodology contributed to a more nuanced understanding of the temporal characteristics of customer signals when applied across an appropriate selection of alternative values of K. Serial correlation results that are robust to alternative specifications are generally preferred.

6. Discussion and implications

Using a combination of unsupervised machine learning for pattern discovery (Choudhury, Allen, & Endres, 2021) and dynamic panel data techniques for inference, this study advances our understanding of online customer reviews by uncovering serial correlation patterns across specific dimensions of the hotel experience. The results are largely consistent with the likely responsibilities of local managers, with a few exceptions. Topics we perceived to be resource-constrained (e.g. room size, location and pool, special aspects) often persist over time, while others (e.g. maintenance issues and food quality) seem to be remediable and are typically resolved more quickly. By moving beyond aggregate

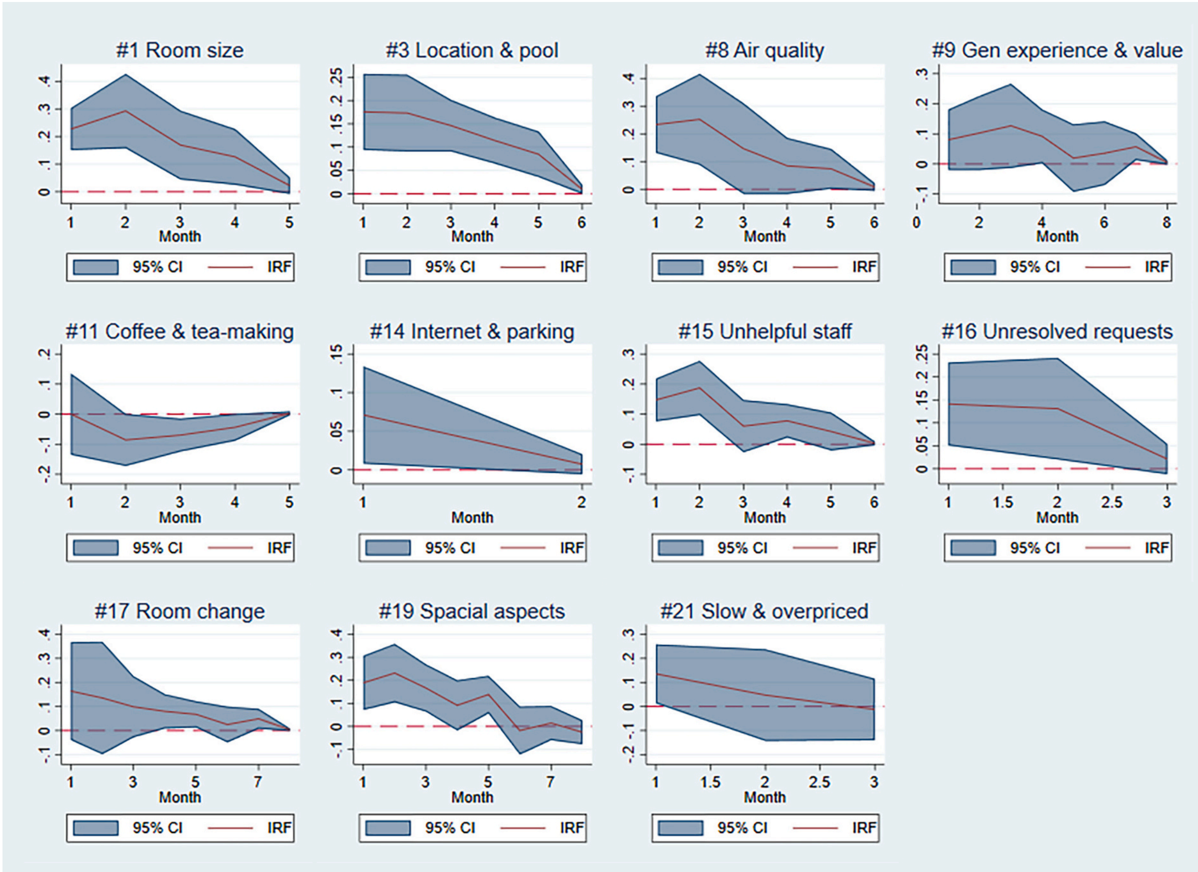

Fig. 4. Impulse response functions (IRFs) of Topic Relevance in high season, where the dashed horizontal line indicating the long-term trend.

review scores and sentiment measures, this study offers a more fine-grained understanding of how customer dissatisfaction unfolds over consecutive cohorts.

From a managerial perspective, the substantial market-wide effects presented in this paper can help shape practical service management strategies. For example, the persistence of complaints about staff behaviour and rudeness implies that ad hoc responses are generally insufficient; instead, systematic investment in training, or recruitment practices may be required. Similarly, persistent complaints about facilities during high-demand periods highlight the need for proactive capacity planning.

Perhaps more importantly, our research contributes to the literature on service management and consumer-generated content by positioning online reviews as a proxy for realized managerial corrective actions. Previous work has largely emphasized the contemporaneous role of reviews as signals of service quality or focused on manager's narrative responses to negative feedback. In contrast, our analysis demonstrates that serial correlation provides an additional, dynamic dimension through which managerial effectiveness can be inferred indirectly. In doing so, this study complements and extends prior work in agency theory and service management by providing an observable lens into hidden managerial behaviours, particularly in contexts where corrective actions are difficult to observe directly.

This approach is particularly valuable for owners of large branded chains, who may be distant from day-to-day service provisions and may face challenges in distinguishing transient service failures from persistent ones. Owners seeking to assess the performance of their own properties, or comparing it with that of competitors, can apply our methodology to data partitioned by individual hotel brands. By identifying serially correlated complaint categories, our method provides a systematic approach to performance evaluation by hotel brand.

At the time of writing, there are still limits on how far text analytics (or large-language models) can go in their contributions to managerial decision-making, without the need for human judgements. In this paper, the term 'capacity constraint' has been carefully used in relation to customer experience topics. However, even capacity constraints may require a judgement as to which topics are under the control and responsibility of local managers, i.e. are susceptible to managerial control, and which are not. Sooner or later, senior managers in hotels or on platforms may need to make the required human judgements, and take corrective action when customer complaints are persistent.

The findings also carry implications for digital platforms such as Booking.com. Platforms could integrate temporal analytics into their review dashboards, allowing hotels to track not only the volume and valence of reviews but also the persistence of complaint categories over time (see Fig. 6).

The relative strengths of our approach based on authenticated online reviews can be claimed only by comparison with alternative techniques for measuring customer experience. For example, customer experience for smaller samples could be identified through web-based customer panels but these are costly and comprise those registrants willing to participate (Porter, Outlaw, Gale, Thomas, & Cho, 2019). Online or paper-and-pencil questionnaires are also very costly and time-consuming in different ways. While care is needed to take seasonal factors into consideration, online reviews provide richer information in less time and at a lower cost (Qi, Zhang, Jeon, & Zhou, 2016: 951). In non-docile markets, the complexity and speed with which a firm needs to acquire and analyze information must increase (Xu, Frankwick, & Ramirez, 2015: 1565), and it seems likely that in highly competitive international hotel chains, for example, one may be justified in assuming that big data analytics predominate over traditional methods. Managers and potential customers can now access negative online reviews at

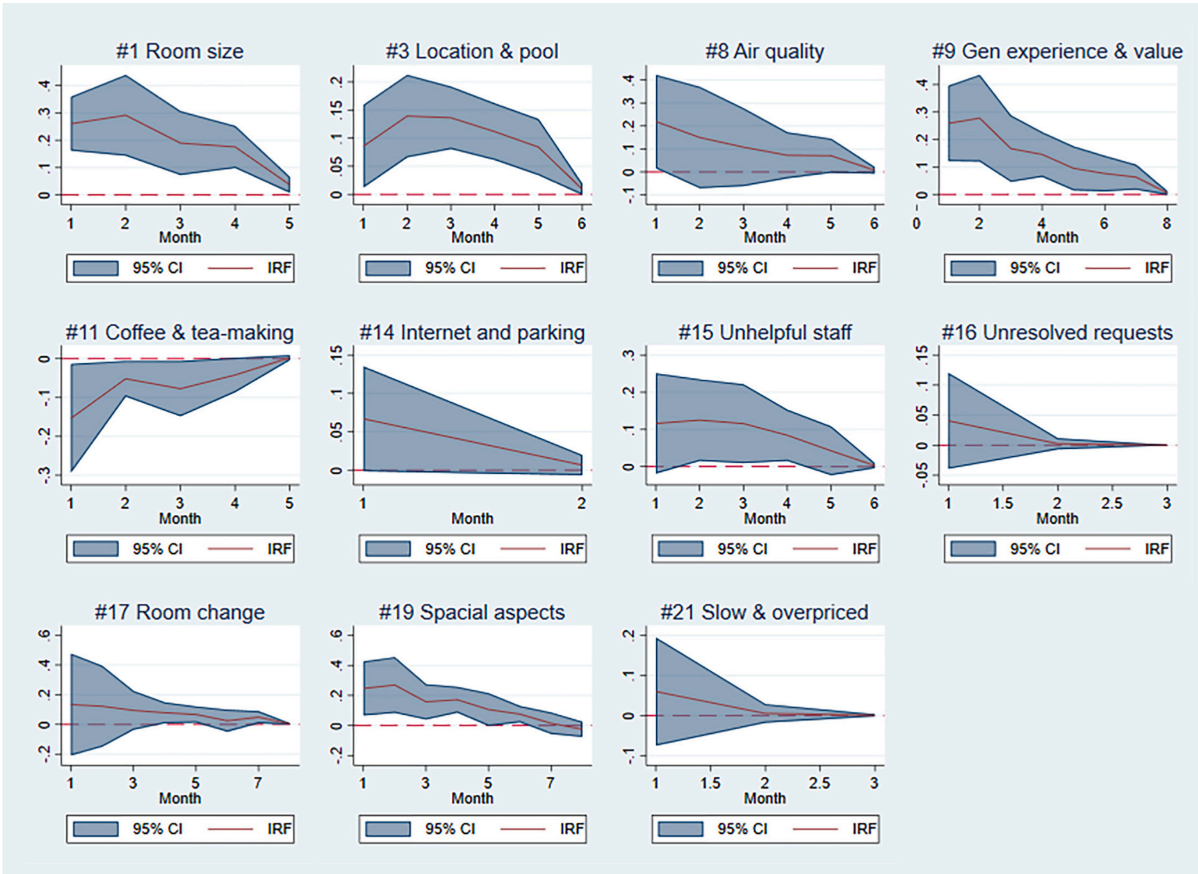

Fig. 5. Impulse response functions (IRFs) of Topic Relevance in low season, with the dashed horizontal line indicating the long-term trend.

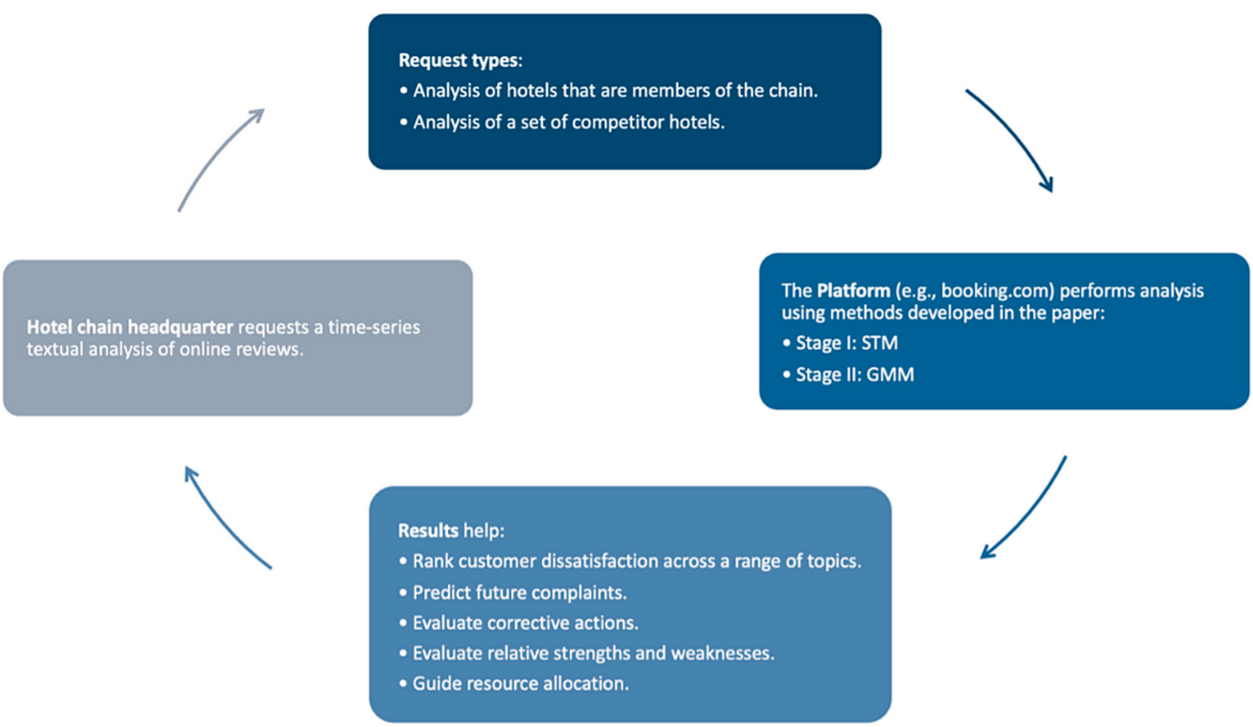

Fig. 6. Potential applications for platforms and hotel chains.

negligible cost (Xu et al., 2015). In these circumstances, big data and traditional marketing analytics may be used synergistically, where the big data analysis of serially correlated negative reviews may alert managers to the need for deeper, more traditional investigation.

Finally, this study opens several avenues for future research. First, extending the approach to other service industries such as restaurants or airlines could test the generalizability of our findings. Second, future studies could explore the interaction between managerial narrative responses and their realized corrective actions, examining whether communication strategies differ between topics characterized by persistent complaints and those dominated by transient complaints.

CRediT authorship contribution statement

Rodion Skovoroda: Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Wei Yang:** Writing – review & editing, Formal analysis, Conceptualization. **Bowei Chen:** Writing – review & editing, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Trevor Buck:** Writing – review & editing, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that there are no conflicts of interest related to this work.

Acknowledgements

The authors would like to thank the Open University and the University of Glasgow for their support in this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.annale.2025.100199.

References

Ananthakrishnan, U. M., Li, B., & Smith, M. D. (2020). A tangled web: Should online review portals display fraudulent reviews? *Information Systems Research, 31*, 950–971.
Andrews, D. W., & Lu, B. (2001). Consistent model and moment selection procedures for GMM estimation with application to dynamic panel data models. *Journal of Econometrics, 101*, 123–164.
Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies, 58*, 277–297.
Bauman, K., & Tuzhilin, A. (2022). Know thy context: Parsing contextual information from user reviews for recommendation purposes. *Information Systems Research, 33*(1), 179–202.
Blodgett, J. G., Granbois, D. H., & Walters, R. G. (1993). The effects of perceived justice on complainants' negative word-of-mouth behavior and patronage intentions. *Journal of Retailing, 69*, 399–428.
Blundell, R. W., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics, 87*, 115–143.
Bolton, R. N., Lemon, K. N., & Bramlett, M. D. (2006). The effect of service experiences over time on a supplier's retention of business customers. *Management Science, 52*, 1811–1823.
Bowen, D. E., Schneider, B., & Cummings, L. L. (1988). Services marketing and management: Implications for organizational behavior. In B. M. Staw, & B.M (Eds.), *Vol. 10. Research in organizational behavior* (pp. 43–80). Greenwich, CT: JAI.
Bowen, J., & Ford, R. C. (2002). Managing service organizations: Does having a "thing" make a difference? *Journal of Management, 28*(3), 447–469.
Chang, Y. C., Ku, C. H., & Chen, C. H. (2020). Using deep learning and visual analytics to explore hotel reviews and responses. *Tourism Management, 80*, Article 104129.
Chen, B., Anker, T. B., & Liang, X. (2025). Business continuity management in the sharing economy: insights from Airbnb online reviews. *Tourism Management, 107*, Article 105067.
Chevalier, J., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*, 345–354.
Choudhury, P., Allen, R. T., & Endres, M. G. (2021). Machine learning for pattern discovery in management research. *Strategic Management Journal, 42*, 30–57.
Conlon, D. E., & Murray, N. M. (1996). Customer perceptions of corporate responses to product complaints: The role of explanations. *Academy of Management Journal, 39*, 1040–1056.
Daw, A., Hampshire, R. C., & Pender, J. J. (2025). How to staff when customers arrive in batches. *Management Science, 71*(8), 6580–6601.
Dellarocas, C. (2003). The digitization of word-of-mouth: Promise and challenges of online feedback mechanisms. *Management Science, 49*, 1407–1424.
Drucker, P. F. (1954). *The practice of management*. New York: HarperCollins.
Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in Sociology, 7*, Article 886498.
Fang, Z., Luo, X., & Jiang, M. (2013). Quantifying the dynamic effects of service recovery on customer satisfaction: Evidence from Chinese mobile phone markets. *Journal of Service Research, 16*, 341–355.
Feng, J., Li, X., & Zhang, X. (2019). Online product reviews-triggered dynamic pricing: Theory and evidence. *Information Systems Research, 30*, 1107–1123.
Gao, B., Li, X., Liu, S., & Fang, D. (2018). How power distance affects online hotel ratings: The positive moderating roles of hotel chain and reviewers' travel experience. *Tourism Management, 65*, 176–186.
Godes, D., & Mayzlin, D. (2004). Using online conversations to study word of mouth communication. *Marketing Science, 23*, 545–560.
Gu, B., Park, J., & Konana, P. (2012). The impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research, 23*, 182–196.
Hannan, E. J., & Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society: Series B: Methodological, 41*, 190–195.
Hogreve, J., Bilstein, N., & Mandl, L. (2017). Unveiling the recovery time zone of tolerance: When time matters in service recovery. *Journal of the Academy of Marketing Science, 45*, 866–883.
Hu, N., Pavlou, P. A., & Zhang, J. J. (2017). On self-selection biases in online product reviews. *MIS Quarterly, 41*, 449–471.
Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management, 72*, 417–426.
Jang, S. S. (2004). Mitigating tourism seasonality: A quantitative approach. *Annals of Tourism Research, 31*, 819–836.
Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*, 305–360.
Kim, W. G., Lim, H., & Brymer, R. A. (2015). The effectiveness of managing social media on hotel performance. *International Journal of Hospitality Management, 44*, 165–171.
Kirilenko, A. P., & Stepchenkova, S. (2025). Facilitating topic modeling in tourism research: Comprehensive comparison of new AI technologies. *Tourism Management, 106*, Article 105007.
Kiviet, J., Pleus, M., & Poldermans, R. (2017). Accuracy and efficiency of various GMM inference techniques in dynamic micro panel data models. *Econometrics, 5*, 14.
Kolomoyets, Y., & Dickinger, A. (2023). Understanding value perceptions and propositions: A machine learning approach. *Journal of Business Research, 154*, Article 113355.
Kripfganz, S. (2019). Generalized methods of moments estimation of linear dynamic panel data models. In *Proceedings of the 2019 London Stata Conference*.
Lappas, T., Sabnis, G., & Valkanas, G. (2016). The impact of fake reviews on online visibility: A vulnerability assessment of the hotel industry. *Information Systems Research, 27*, 940–961.
Le, H. T. M., Phan-Thi, T. A., Nguyen, B. T., & Nguyen, T. Q. (2025). Mining online hotel reviews using big data and machine learning: An empirical study from an emerging country. *Annals of Tourism Research Empirical Insights, 6*, Article 100170.
Lee, C. C., & Hu, C. (2004). Analyzing hotel customers' e-complaints from an internet complaint forum. *Journal of Travel & Tourism Marketing, 17*, 167–181.
Levin, A., Lin, C.-F., & Chu, C.-S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties. *Journal of Econometrics, 108*, 1–24.
Liu, Y., Teichert, T., Rossi, M., Li, H., & Hu, F. (2017). Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews. *Tourism Management, 59*, 554–563.
McColl-Kennedy, J. R., Zaki, M., Lemon, K. N., Urmetzer, F., & Neely, A. (2019). Gaining customer experience insights that matter. *Journal of Service Research, 22*, 8–26.
Mensah, E. A., Odame, D. N. A., Ankrah, I., Obuobisa-Darko, T., & Hinson, R. E. (2025). From reviews to reflections: Understanding tourist sentiments and satisfaction in African destinations through user-generated content. *Annals of Tourism Research Empirical Insights, 6*, Article 100174.
Mimno, D., Wallach, H., Talley, E., Leenders, M., & McCallum, A. (2011). Optimizing semantic coherence in topic models. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing, Edinburgh, 27-31 July* (pp. 262–272).
Mintzberg, H., & Waters, J. A. (1985). Of strategies, deliberate and emergent. *Strategic Management Journal, 6*, 257–272.
Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica, 49*, 1417–1426.
Pantano, E., Priporas, C. V., & Stylos, N. (2017). 'You will like it!' using open data to predict tourists' response to a tourist attraction. *Tourism Management, 60*, 430–438.
Papathanassis, A., & Knolle, F. (2011). Exploring the adoption and processing of online holiday reviews: A grounded theory approach. *Tourism Management, 32*(2), 215–224.
Parasuraman, A., Zeithaml, V., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perception of service quality. *Journal of Retailing, 64*, 12–41.
Pelsmacker, P. D., Tilburg, S. V., & Holthof, C. (2018). Digital marketing strategies, online reviews and hotel performance. *International Journal of Hospitality Management, 72*, 47–55.

Porter, C. O. L. H., Outlaw, R., Gale, J. P., Thomas, S., & Cho, T. S. (2019). The use of online panel data in management research: A review and recommendations. *Journal of Management, 45*, 319–344.

Proserpio, D., & Zervas, G. (2017). Online reputation management: Estimating the impact of management responses on consumer reviews. *Marketing Science, 36*, 645–665.

Qi, J., Zhang, Z., Jeon, S., & Zhou, Y. (2016). Mining customer requirements from online reviews: A product improvement perspective. *Information & Management, 53*, 951–963.

Ravichandran, T., & Deng, C. (2023). Effects of managerial response to negative reviews on future review valence and complaints. *Information Systems Research, 34*, 319–341.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software, 91*, 1–40.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., … Rand, D. G. (2014). Structural topic models for open-ended survey responses. *American Journal of Political Science, 58*, 1064–1082.

Rodriguez-Diaz, M., Rodriguez-Diaz, R., & Espino-Rodriguez, T. F. (2018). Analysis of the online reputation based on customer ratings of lodgings in tourism destinations. *Administrative Sciences, 8*, 1–18.

Roodman, D. (2009). How to do xtabond2: An introduction to "difference" and "system" GMM in Stata. *Stata Journal, 9*, 86–136.

Sparks, B. A., & Browning, V. (2011). The impact of online reviews on hotel booking intentions and perception of trust. *Tourism Management, 32*, 1310–1323.

Sriram, S., Chintagunta, P. K., & Manchanda, P. (2015). Service quality variability and termination behavior. *Management Science, 61*, 2739–2759.

Susskind, A., Kacmar, K. M., & Borchgrevinkm, C. P. (2003). Consumer service providers' attitudes relating to customer service and customer satisfaction in the customer-server exchange. *Journal of Applied Psychology, 88*, 179–187.

Tran, D. T., Nguyen, K. T., Van Huynh, D., & Stangl, B. (2025). Satisfaction with response: The impact on potential customers' perceived service quality and intent to stay. *Annals of Tourism Research Empirical Insights, 6*, Article 100179.

Tsai, C. F., Chen, K., Hu, Y. H., & Chen, W. K. (2020). Improving text summarization of online hotel reviews with review helpfulness and sentiment. *Tourism Management, 80*, Article 104122.

Tuomi, A. (2021). Deepfake consumer reviews in tourism: Preliminary findings. *Annals of Tourism Research Empirical Insights, 2*(2), Article 100027.

Voorhees, C. M., Beck, J. M., Randhawa, P., DeTienne, K. B., & Bone, S. A. (2021). Assessing the effects of service variability on consumer confidence and behavior. *Journal of Service Research, 24*, 405–420.

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009). Evaluation methods for topic models. In *Proceedings of the 26th Annual International Conference on Machine Learning, Montreal, 14-18 June* (pp. 1105–1112).

Wang, T., Wezel, F. C., & Forgues, B. (2016). Protecting market identity: When and how do organizations respond to consumers' devaluations? *Academy of Management Journal, 59*, 135–162.

Williamson, O. E. (1975). *Markets and hierarchies: Analysis and antitrust implications*. London: Collier-Macmillan.

Wooldridge, J. M. (2012). *Introductory econometrics: A modern approach*. South Western: Cengage Learning.

Xu, Z., Frankwick, G. L., & Ramirez, E. (2015). Effects of big data analytics and traditional marketing analytics on new product success: a knowledge fusion perspective. *Journal of Business Research, 69*, 1562–1566.

Yang, M., Ren, Y., & Adomavicius, G. (2019). Understanding word-of-mouth and customer engagement on Facebook business pages. *Information Systems Research, 30*, 839–855.

Ye, F., Xia, Q., Zhang, M., Zhan, Y., & Li, Y. (2022). Harvesting online reviews to identify the competitor set in a service business: evidence from the hotel industry. *Journal of Service Research, 25*, 301–327.

You, Y., Vadakkepatt, G. G., & Joshi, A. M. (2015). A meta-analysis of electronic word-of-mouth elastisity. *Journal of Marketing, 79*, 19–39.

Yucel, A., Dag, A., Oztekin, A., & Carpenter, M. (2022). A novel text analytic methodology for classification of product and service reviews. *Journal of Business Research, 151*, 287–297.

Zhang, Z., Wang, B., Law, R., & Han, Y. (2024). Public health emergencies and travelers' review efforts. *Annals of Tourism Research, 106*, Article 10376.