ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Noisy signals: do ratings' volatility depend on the length of the consumption span?

David Boto-García
Veronica Leoni

Quaderni - Working Paper DSE N°1183

Noisy signals: do ratings' volatility depend on the length of the consumption span?

David Boto-García[1] Veronica Leoni*[2,3]

Abstract:

This paper investigates the informational content of online reviews. For the case of hotels, we model how the length of the stay shapes the variance of review scores. Grounded on violations of temporal monotonicity, errors in recall and hedonic adaptation theories, we first present a characterization of how the consumption span affects the non-deterministic component of consumer satisfaction. Next, we conduct an empirical analysis using more than 525,000 individual reviews from Booking.com in 5 major European cities. Under a heteroskedastic framework, we document that individual ratings' volatility decreases with the length of the stay. This implies that online ratings from short stayers (short consumption episodes) are noisy signals of the underlying hotel quality. Furthermore, we show that greater volatility in hotel ratings translates into a lower share of useful reviews for subsequent consumers. Our findings offer relevant insights for platform design operators about the sources of ratings' volatility and how this affects social learning.

Keywords: *online reviews; ratings' variance; length of stay; quality uncertainty; heteroskedasticity; Booking.com*

JEL codes: *D12; D83; Z30.*

* Corresponding author: e-mail: veronica.leoni3@unibo.it.

[1] Department of Economics, University of Oviedo (Spain)
[2] Center for Advanced Studies in Tourism, University of Bologna (Italy)
[3] Department of Applied Economics, University of the Balearic Islands (Spain)

Non-technical summary

While online user-generated content has become a valuable tool for ex-ante assessment of service quality, the volatility of scores that different consumers assign to each hotel can affect the usefulness of average scores. This article discusses how the length of a hotel stay can affect the variance of individual ratings. Hotel scores with low variance offer a more consistent and credible signal about expected quality, but the retrospective assessment of hedonic episodes is subject to cognitive and psychological biases. The duration of the episode plays a relevant role, making some dimensions more salient to memory than others.

The paper evaluates the role of guests' length of stay on the variance of individual hotel ratings. The study uses data from 525,000 individual reviews for 1,233 hotels in five major European cities, and the findings suggest that the relative importance of the random component of overall scores (variance) decreases with length of stay. This finding implies greater consistency and lower polarization in ratings among long-stayers. The study also examines heterogeneity by consumer profile and relates the predicted variance from the model to the share of reviews that other consumers deem as 'useful'.

The polarization of reviews is economically meaningful for several reasons. A higher rating variance can reduce demand and sales, and small amounts of misperceptions through polarization can produce important breakdowns when aggregating information in contexts of social learning. Understanding the sources of ratings' polarization thus has important implications for social learning dynamics in online platforms since user learning strongly depends on the size and congruity of information.

This study expands existing literature in two important ways. Firstly, it characterizes how the length of stay affects the non-deterministic component of rating scores. Secondly, it adds to an emerging literature on review helpfulness. The study provides empirical evidence that long-stayers assign less polarized scores, which could impact how hotels should be rated and marketed to potential guests.

Overall, this study highlights the importance of considering the length of stay when evaluating hotel ratings and quality assessments. The findings suggest that hotels may want to consider ways to incentivize longer stays, as they may lead to more consistent and reliable ratings. Furthermore, the study could provide insights into how to design better review platforms that facilitate social learning and decision-making.

1. INTRODUCTION

The provision of experiential services, such as a hotel stay, is inherently characterized by a degree of uncertainty regarding the expected quality. In addition to official hotel star ratings, the proliferation of online platforms such as Booking.com has further facilitated the reduction of information asymmetries and the ‘market for lemons’ problem (Akerlof, 1970). In this sense, online user-generated content is nowadays one of the most relevant informational tools for product search and ex-ante assessment of service quality (Magnani, 2020).

However, the effectiveness of online reviews and ratings as information sources crucially depend on the volatility in the scores that different consumers assign to each hotel. The variance reflects the extent to which individual ratings diverge from the average and can be understood as a measure of consistency that shapes ex-ante expectations. If the variance in ratings for a hotel is high (greater polarization), it may be more difficult for consumers to accurately gauge its overall quality. That is, the usefulness of average scores is diminished when consumers’ report both good and bad experiences. On the contrary, for the same mean rating, hotel scores with low variance offer a more consistent and credible signal about expected quality.

When rating a hotel stay, tourists consider both objective and subjective (affective) dimensions and compare the quality of the service with prior expectations (Oliver, 1980; Engler et al., 2015). However, the retrospective assessment of hedonic episodes (remembered utility) is subject to cognitive and psychological biases (Kahneman et al., 1997). When evaluating temporarily extended outcomes, the duration of the episode plays a relevant role as it makes some dimensions more salient to memory than others. In the context of hotels, ratings made by long-stayers are based on a deeper knowledge about the services provided and therefore more reflective of the underlying hotel quality. At the same time, when consuming heterogeneous goods, people form summary evaluations giving comparatively more weight to positive and negative deviations from expectations during short stays because these shocks (i) tend to dissipate during long consumption spans through hedonic treadmill (Rayo and Becker, 2007), and (ii) are comparatively more salient to memory (Mullainathan, 2002).

This paper evaluates the role of guests' length of the stay on the variance (dispersion from the average) of individual hotel ratings. Previous works have shown that the duration of tourists' stay at destination affects post-trip satisfaction because of changes in destination perceptions during the course of the vacation experience (e.g., Vogt and Andereck, 2003). For the case of hotels, length of stay has been shown to matter for the likelihood of posting numerical and written online reviews (Kim and Han, 2022) and to be negatively associated with mean rating values (Brandes and Dover, 2022; Leoni and Moretti, 2023). However, to our knowledge, there is no evidence to date about how the stay duration affects the volatility (and therefore polarization) of hotel rating scores. This paper precisely aims to address this gap.

Firstly, we characterize the potential mechanisms through which length of stay allegedly influences the variance of ratings from a theoretical viewpoint. We then conduct an empirical analysis using data from 525,000 individual reviews for 1,233 hotels located in 5 major European cities (Madrid, Barcelona, Lisbon, Rome and Milan) and listed on Booking.com platform. Based on a multiplicative heteroskedastic regression (Harvey, 1976) in which both the conditional mean and variance of ratings are modelled, we show that the relative importance of the random component of overall scores (variance) decreases with length of stay. This finding implies a greater consistency and lower polarization in ratings among long-stayers. In doing so, we also examine heterogeneity by consumer profile. Subsequently, we relate the predicted variance from our model to the share of reviews (at the hotel level) that other consumers deem as 'useful'. We provide evidence of a clear negative relationship between polarization and usefulness, which offers important insights for theory and practice.

The polarization of reviews is economically meaningful for several reasons. Firstly, previous works for the case of books (Sun, 2012), restaurants (Wu et al., 2015) and hotels (Ye et al., 2009) have shown that a higher rating variance reduces demand and sales. Secondly, small amounts of misperceptions through polarization can produce important breakdowns when aggregating information in contexts of social learning (Frick et al., 2020). From this viewpoint, understanding the sources of ratings' polarization has important implications for social learning dynamics in online platforms since user learning strongly depends on the size and congruity of information (Acemoglu et al., 2022).

This work expands existing literature in two important ways. On the one hand, we characterize how the length of the stay affects the non-deterministic component of rating scores. Specifically, our framework posits that length of stay as an indicator of the duration of the consumption episode shapes the variance of ratings through a mixture of errors in recall (Mullainathan, 2002), violations of temporal monotonicity (Kahneman and Thaler, 2006) and potential hedonic treadmill (Rayo and Becker, 2007). Based on a large dataset of individual ratings on Booking.com platform, we provide empirical evidence that long-stayers assign less polarized scores. On the other hand, our work adds to an emerging literature on review helpfulness (Lee et al., 2021; Liu et al., 2023; Mudambi and Schuff, 2010; Zhang et al., 2023; Zhao et al., 2013) by showing how greater ratings' volatility dampers the informational content of online reviews, making the signal about expected quality noisier and less useful.

The remainder of the paper is structured as follows. Section 2 reviews the related literature. Section 3 presents the theoretical framework for the analysis. Section 4 describes the data and presents some summary statistics and preliminary evidence together with the econometric modelling. Section 5 reports and discusses the estimation results together with some robustness checks and extensions. Finally, Section 6 concludes with a summary of findings, implications and limitations.

2. LITERATURE REVIEW

2.1. Online reviews as quality cues

Product ratings on online platforms are nowadays a major informational source for consumers when choosing among alternative providers (Wu et al., 2015). Electronic word of mouth reduces search costs and offers up to date information, thereby exerting a strong influence on market demand and revenues for different goods and services (Chevalier and Mayzlin, 2006; Liu, 2006). In presence of asymmetric information, people learn about product quality from both numerical ratings and contextual user-generated comments (Fang, 2022). These ratings can be understood as indicators of remembered experienced utility (hedonic quality) from previous consumers in the sense of Kahneman et al. (1997). Accordingly, agents learn from the public disclosure of information by peers (Amador and Weill, 2012), being the reliance on observational learning more prevalent among infrequent and unexperienced costumers (Cai et al, 2009). Moreover, consumers have been shown to rely more on average ratings than on other quality cues like prices or the number of reviews (de Langue et al., 2016). In the hospitality industry, subjective ratings have displaced objective classification systems as reputation signals, and hotel managers are nowadays monitoring consumer reviews on online platforms (Proserpio and Zervas, 2017).

A large body of literature has investigated the relevance and pervasiveness of online reviews to consumers in many different settings and from different viewpoints. Scholars in marketing have focused on consumers' motives to review whereas economists have been more concerned about the effect of reviews on sales and demand. A review of the state of art can be found in Magnani (2020). A common finding is that reviews and ratings are affected by selection effects since individuals with extreme levels of satisfaction (very low or very high) are more likely to rate products than those with moderate evaluations (Moe and Schweidel, 2012; Schoenmueller et al., 2020).[4] Among them, very satisfied consumers are relatively more prone to review than dissatisfied ones, partly due to manipulation practices (Mayzlin et al., 2014). In this vein, there is some evidence of upward bias in online valuations because scores tend to be J-shaped (Pourfakhimi et al., 2020), with most suppliers receiving very high rates (Zervas et al., 2021).

One explanation for the ratings inflation is that, when acting as reviewers, individuals are subject to *herding behavior* as theoretically conceptualized in Banerjee (1992): they are highly influenced by previously posted ratings by other consumers, with reviewer experience acting as a relevant moderator (Sunder et al., 2019). Moreover, they tend to conform to the average score, which acts as an anchor, leading to the so-called *rating bubbles* phenomenon (Moe and Trusov, 2011). Importantly, herding behaviour is asymmetric: people are comparatively more influenced by excellent rather than low ratings (Cicognani et al., 2022; Moe and Schweidel, 2012).

[4] Another source of selection is differential attrition, by which reviewers with moderate experiences are more likely to exit the pool of active reviewers (Brandes et al., 2022).

2.2.Noisy reviews and heteroskedastic ratings

Despite online reviews are a tool for disclosing quality information for experience goods, the degree of informativeness of such content strongly depends on their polarization. In a recent paper, Acemoglu et al. (2022) characterize learning dynamics from online reviews, showing that more information does not necessarily lead to faster learning; in fact, user learning would strongly depend on the size and congruity of information. Consumers are predicted to purchase a product/good with greater probability under moderate rather than extreme disagreement due to mismatch costs (Lee et al., 2023). In this vein, high-variance reviews are more dampening for expensive products (Kim and Krishnan, 2015) and exert a different influence on consumption propensity depending on subjective prior expectations about the good (West and Broniarczyk, 1998).

Several works have shown that high variance reduces demand because polarization makes the quality signal noisier and, in turn, less helpful (Lee et al., 2021; Mudambi and Schuff, 2010; Park and Park, 2013; Sun, 2012; Ye et al., 2009; Zhao et al., 2013). This stream of literature documents that product type moderates the relationship between review variance and helpfulness, with experience goods being more sensitive to polarization. Most work on this matter has mainly evaluated Amazon products like DVDs, PC video games, cell phones and digital cameras (Lee et al., 2021; Mudambi and Schuff, 2010). For the case of hotels, a growing body of research has analysed the drivers of review usefulness. This literature has shown that informative and readable reviews from consumers with high reputation are deemed as more helpful (Liang et al., 2019; Liu and Park, 2015). Review diversity (different ratings for each item that composes the overall score) improves usefulness for the case of negative reviews (Liu et al., 2023) whereas high arousal makes reviews less helpful (Chatterjee, 2020). Moreover, emoticons enhance usefulness when the review is narrative-based (Huang et al., 2020). Nevertheless, only a few have explicitly considered the role of the variance in ratings scores. Ye et al. (2009) estimate that a 10% increase in review variance decreases the number of bookings by around 2.8%. Lo and Yao (2019) prove experimentally that review credibility is higher when the ratings are consistent with previous reviews. Zhang et al.

(2023) shows that perceived usefulness of reviews increases with the novelty on topics covered but is inversely related to inconsistency, defined as the gap between sentiment scores and the 10 previous reviews on the topic.

Overall, although the abovementioned works have acknowledged that review variance matters for consumers' choices, we still know very little about the sources of ratings' heteroskedasticity. Against this background, we ask the following research question: what explains the differences in ratings' dispersion from the mean?

2.3. Length of stay as a satisfaction shifter

Several studies in tourism and hospitality have investigated the connection between tourists' length of stay and reported satisfaction with visited destinations, showing that length of stay changes destination perceptions while on vacation asymmetrically though information acquisition mechanisms (e.g., Vogt and Andereck, 2003). In the hotel context, though, there is far less evidence on how length of stay impacts rating scores, and empirical findings are inconclusive. Kim and Han (2022) show that length of stay affects review propensity and the content of the written text, with long-stayers offering more analytical information for other consumers. Pokryshevskaya and Antipov (2017) report that length of stay is associated with higher ratings. In contrast, Leoni and Moretti (2023) and Brandes and Dover (2022) find a negative effect of the stay duration on the reported level of satisfaction with hotel facilities. Mariani and Predvoditeleva (2019), instead, do not find a significant relationship. However, its impact on ratings' variance remains underexplored to date.

3. THEORETICAL FRAMEWORK

3.1. A cognitive-affective model for rating scores

The expectation-disconfirmation paradigm (Oliver, 1980) postulates that consumer satisfaction with a service (in our case, a hotel stay) can be conceptualized as the net utility difference between the experienced quality of the service (U_i) and ex-ante expectations ($E(U_i)$). That is, a consumer is satisfied if the utility obtained from the stay equals or surpasses his/her prior expectations. On the one hand, the experienced utility depends on hotel-specific amenities (q), the length of the consumption experience (LOS_i) and guests' idiosyncratic preferences for the amenities (tastes), which can be proxied by his/her sociodemographic profile X_i in the spirit of Pollak and Wales (1981) (e.g., nationality, travel party composition). On the other hand, since a hotel stay is an experience good, there is high uncertainty about the expected quality of the hotel beforehand (Nelson, 1970). As a result, consumers typically infer it based on previous valuations (ratings) made by other consumers in online platforms (denoted by θ). Such ratings act as an anchor (Cicognani et al., 2022; Moe and Trusov, 2011) and affect post-consumption valuation through different mechanisms like confirmatory bias associated to first impressions (Rabin and Schrag, 1999), informational cascades (Anderson and Holt, 2007) or bandwagon effects (Leibenstein, 1950). Acemoglu et al. (2022) show that consumers' learning from previous ratings is subject to selection effects and evolves over time: ex-ante expectations are shaped by the information available from previous users at each time.

Therefore, consumers' latent satisfaction as a measure of experienced utility (${SAT_i}^*$) can be generically expressed as follows:

$$ {SAT_i}^* = U_i - E(U_i) = f(q, LOS_i, X_i, \theta) \quad (1) $$

This definition of satisfaction is merely deterministic and is based on objective factors only. In other words, two consumers staying at the same hotel, with equal stay duration, and with the same profile would be predicted to report the same satisfaction level. However, several authors have posited that satisfaction in experience services also involves 'emotional' components (e.g., San Martín and Rodríguez-del-Bosque, 2008; Brunner-Sperdin and Peters, 2009). Stochastic factors like weather, personal mood, or unexpected breakdowns on one side, or aspects like joy, excitement or the consumption of complementary goods on the other affect people's retrospective evaluations and their corresponding satisfaction (e.g., Brandes and Dover, 2022; Sirakaya et al., 2004).

From a broader perspective, emotions have been largely acknowledged as relevant factors in explaining economic behaviour (Loewenstein, 2000). In a seminal work, Thurstone (1927) introduced the *Law of Comparative Judgement* by which an objective stimulus is perceived with a normal error due to differences in perceptions. Coherently with Random Utility Theory (McFadden, 1974), latent satisfaction can be therefore additively partitioned into a deterministic and a stochastic component in the following way:

$$SAT_i^{*} = \underbrace{f(q, LOS_i, X_i, \theta)}_{deterministic} + \underbrace{\xi_i}_{stochastic} \tag{2}$$

where ξ_i is a stochastic term that follows a certain probability distribution with mean μ and variance σ_ξ and captures the contribution of unobserved dimensions to experienced utility. The magnitude of σ_ξ captures the relative weight of the 'emotional' over the deterministic component in explaining the observed realization of SAT_i^{*}. The greater σ_ξ, the less accurately the observed satisfaction indicator (SAT_i) reflects the underlying quality of hotel services. In other words, measures of satisfaction that are greatly explained by the emotional component will display greater variance and volatility. As a result, these (strongly emotionally driven) scores are noisy signals of the true (objective) quality of the service provided.

3.2.Length of stay as a shifter of the random component of satisfaction

In many instances, a consumption experience is composed of a set of temporally extended individual episodes. A hotel stay is a paradigmatic example, since consumers' overall valuation of the service depends on experienced utility at each period. Research in behavioural economics and psychology has shown that when a consumption experience involves the temporal aggregation of several consumption episodes, overall remembered utility violates temporal monotonicity (Kahneman et al., 1997). That is, consumers do not perfectly aggregate utility at each period; global valuations overweight some parts and underweight others (Kahneman and Thaler, 2006). Different experimental studies have shown that retrospective evaluations follow a peak/end rule by which overall valuations are mostly driven by the hedonic utility at peak moments (Kahneman et al., 1993; Schreiber and Kahneman, 2000). Furthermore, the length of a consumption experience can influence reported measures of satisfaction through memory-based errors in recall (Mullainathan, 2002); the outcomes from long consumption episodes are predicted to be better remembered through greater salience, thereby making valuations from short consumption spans more sensitive to positive and negative shocks.

Inspired by Zimmerman et al. (2018), we postulate that the value of σ_ξ (individual scores' variance) depends on two main dimensions: (i) *experienced quality differences* during the stay caused by potential breakdowns and service failures on the negative side, and enjoyable, pleasant and memorable experiences on the positive one; and (ii) *taste differences* about hotel services associated with subjective valuations of the consumption experience.

Concerning *experienced quality differences*, the longer the stay, the higher the probability that both positive and negative events take place. Since consumers do not perfectly aggregate instant utility over time, positive and negative deviations from expected service quality (anchor) might receive more weight in post-consumption remembered utility. From this perspective, we could expect ratings from long-stayers to be more volatile. However, the *hedonic adaptation* or *hedonic treadmill* framework postulates that, faced with unexpected negative shocks or positive surprises, people quickly adapt to them as time passes (Leoni and Moretti, 2023; Oswald and Powdthavee, 2008; Rayo and Becker, 2007). This means that if positive or negative stimuli take place during the course of a consumption experience, the corresponding utility loss or gain becomes less intense and smooths over the consumption span because of reference-dependent preferences; individuals quickly integrate into their reference points the positive or negative shocks experienced at previous episodes. From this viewpoint, we could expect the opposite pattern: the overall valuation of long-stayers will be smoother and less affected by positive and negative shocks through integration.

Regarding *taste differences* about quality, the longer the length of the consumption span, the more consumers' perceptions about hotel quality will converge to objective criteria. That is, short stayers dispose less precise information about hotel services so that their corresponding ratings might be more volatile. Consistent with the *Law of Comparative Judgement* (Thurstone, 1927), errors in quality perceptions might dissipate as the number of consumption episodes (days) increase.

Overall, we expect hotel ratings' variance (polarization) to decrease with reviewers' length of stay at hotel (intensity of the consumption span).

4. DATA & METHODS

4.1.Dataset

The empirical analysis uses data retrieved from Booking.com, a leading platform for booking accommodation stays. To increase the generalizability of our results and eliminate destination-specific idiosyncrasies, the sample covers five major destinations in three European countries: Italy (Milan and Rome), Spain (Barcelona and Madrid), and Portugal (Lisbon). These three countries are in the top 12 in terms of overnight stays in Europe (Eurostat, 2019). Zooming into individual destinations, Milan and Rome account for 6% of total overnight stays in traditional accommodations in Italy. Similarly, Barcelona and Madrid represent roughly 12% of overnight stays in Spain whereas Lisbon makes up 21% of the nights spent by tourists in Portugal. After data cleaning, the dataset comprises 525,437 valid observations, each of which corresponds to an individual review posted by an actual hotel guest. We work with a total of 1,233 hotels: 191 in Barcelona, 76 in Madrid, 226 in Lisbon, 232 in Milan, and 508 in Rome.

Booking.com is selected because it offers important advantages compared to other platforms. Firstly, it only allows legitimate guests to offer spontaneous feedback, thereby reducing the risk of fake reviews (Mayzlin et al., 2014) and therefore providing more trustworthy information (Figini et al., 2020). Secondly, the review process is easy and convenient; customers are prompted to leave a review after check-out via email and the review form is structured in a few quick steps, increasing the likelihood of review engagement. Furthermore, the opportunity cost of writing reviews is further reduced since guests are not obliged to leave a textual review: they can simply provide a numerical score, which represents the most relevant piece of information for social learning (Acemoglu et al., 2022).

Data spans the interval between November 2018 and August 2019. In September 2019, Booking.com introduced important changes in the review system design (Mellinas and Martín-

Fuentes, 2021). To avoid mixing reviews that would not be directly comparable, the sample period ends in August 2019. The dataset collection was handled via a web-crawler, which gathered all available reviews for the specified period for each hotel. The review details include guest name (if not, we label it as ‘anonymous’), overall rating score (see below), length of stay (in nights), travel party type (distinguishing among solo traveler, couple, family and group/other), country of residence, room type, date of the stay and date of the review.[5]

4.2.Descriptive analysis

Table 1 presents summary statistics of the variables. The overall rating score is our dependent variable and proxies the experienced utility from the hotel stay. It is a continuous variable bounded between 2.5 and 10 and is obtained as the plain average of the ratings assigned to six facets: cleanness, comfort, location, facilities, staff, and value for money. We work with the overall score only because data about scores assigned to each of the six aspects is not available. Nonetheless, the individual overall score is the most important as hotel overall scores are computed as the mean of individual ratings. Indeed, Acemoglu et al. (2022) show that an overload of information does not lead to faster learning and that summary statistics are the most informative measures. Consistent with related-studies (e.g., Mariani and Borghi, 2018), the distribution of overall scores is left-skewed; the average score is 8.65, with a standard deviation of 1.44 points.

From the original dataset, we dropped reviews from guests that stayed for more than 15 days in the hotel (178 observations). In this way, the average length of stay is 2.6 nights, ranging between 1 and 15 with a standard deviation of 1.5 nights. The average length of stay of the sample corresponds with official statistics provided by UNWTO (2019). Tourists in Spain typically stay for 3.2 nights on average in hotels and similar accommodations, while in Italy (Portugal) the average stay is 2.9 (3) nights.

[5] Unfortunately, we lack information about actual rates paid by guests. Monthly and hotel fixed effects together with dummies for room type and travel party composition are used in the empirical analysis to capture quality differences in hotel services. These factors have been shown to impact accommodation prices in previous studies and can be used as proxies for differences in guest profiles and stay purpose (Guizzardi et al., 2022).

We define three types of room categories (*economy*, *standard* and *superior*), which represent 5%, 41%, and 10% of the sample, respectively. The remaining 43% collapse into a fourth category labelled as ‘other’. As regards the country of origin, there is high heterogeneity in nationality (see Appendix, Table A1 for the full list), with only 19% of the total reviews being from domestic travellers. Concerning travel party composition, 40% of reviewers travel in couples, 23% with family members and 17% in groups. The rest (19%) are *solo* travellers. Around 10% of reviews are from anonymous guests, almost 28% of reviews were written on weekends and more than 85% in the same month the stay took place. Because some studies have pointed that the temporal distance between consumption and reviewing affects review positivity (e.g., Brandes and Dover, 2022), we define a dummy variable labelled *Temp. contiguity* that takes value 1 for reviews written close to the stay (same month).

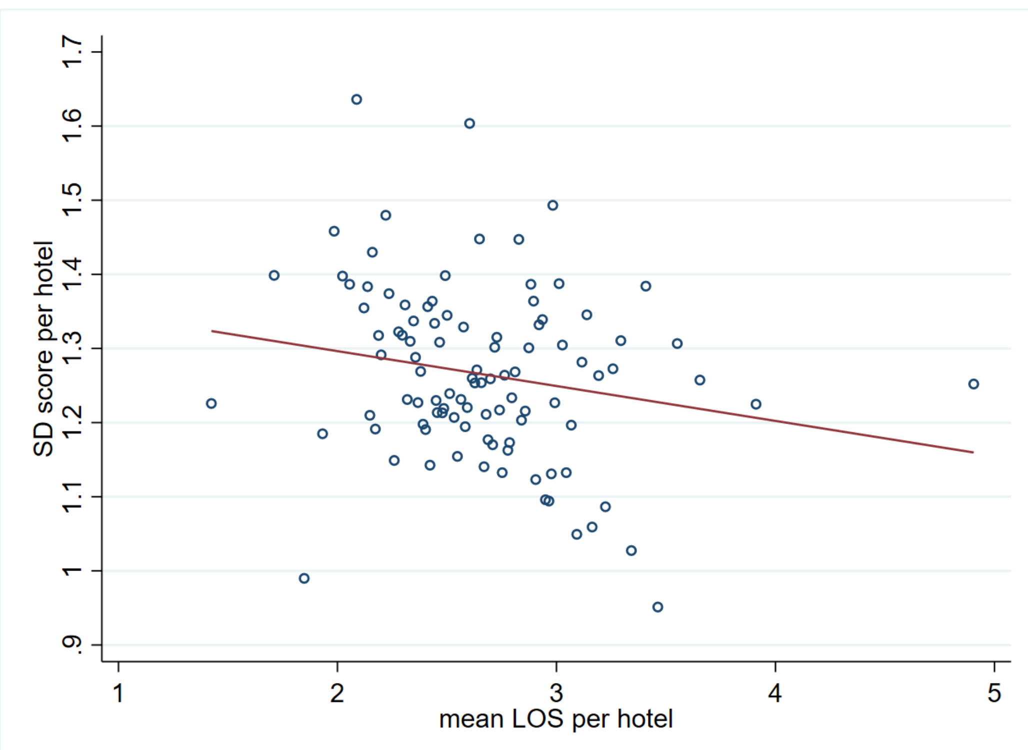
Table 1. Summary statistics of the variables

Label	Description	Mean (%)	SD	Min.	Max.
Score	Rating score	8.648	1.445	2.5	10
Num.reviews (t-1)	Stock of reviews received up to t-1	655.74	928.58	0	7,139
Av.score (t-1)	Average score of stock of reviews up to t-1	8.66	0.52	2.5	10
SDscore (t-1)	Standard deviation of stock of reviews up to t-1	1.30	0.30	0	3.97
LOS	Nights at the hotel (length of stay)	2.62	1.55	1	15
Economy	=1 if room=economy	5.29			
Standard	=1 if room=standard	41.37			
Superior	=1 if room=superior	10.48			
Other	=1 if other type of room	42.86			
Anonymous	=1 if left by anonymous guest	10.18			
Temp. contiguity	=1 if month stay=month review	85.55			
Weekend	=1 if review is left on a weekend	27.55			
Couple	=1 if travel party=couple	40.53			
Family	=1 if travel party= family	22.24			
Group	=1 if travel party= group	17.66			
Solo	=1 if travel party= solo traveler	19.55			
Domestic	=1 if domestic guest	18.90			
	Observations	525,437			

Based on the individual reviews, we computed the stock of reviews received by each hotel in the sample per period. Next, each individual review in the sample was assigned the corresponding hotel review stock up to one-month before the stay. The same was done for the mean and the variance of overall scores. These variables are relevant to control for in the analysis as they capture the quantity, value, and dispersion of ratings at the time of booking, which could influence post-consumption individual ratings through ex-ante expectations and anchoring. These three variables vary over time depending on the date of the stay (and the booking), aiming at capturing learning dynamics and selection effects based on available information in the spirit of Acemoglu et al. (2022). On average, the stock of reviews one-month before the stay is 655, with an overall average score of 8.6 and a standard deviation of 1.30 points.

As a first step, we visually inspect whether there is a link between ratings' variance and the length of the stay. To this end, we compute the standard deviation of rating scores for each hotel in the sample. Subsequently, we calculate the mean length of stay per hotel. Figure 1 presents a binned scatterplot of the pairwise correlation between the two. As illustrated there, it seems there is a negative association: there is lower volatility in ratings' scores in hotels whose customers stay for longer. However, because this analysis is done using aggregate data, we cannot disentangle the influence of length of stay on ratings' variance from hotel quality or other sources of heterogeneity associated with long stays. To offer a more detailed characterization, we move to a formal econometric analysis.

Figure 1. Descriptive binned scatterplot of SD score per hotel on mean LOS per hotel



Note: the bins are defined based on the 100 quantiles of the distribution of the variables

4.3.Econometric modelling

Consistent with the model presented in Section 3, we estimate the following Harvey-type heteroskedastic regression model (Harvey, 1976) at the individual review level:

$$\begin{aligned} score_{ijt} = \alpha + \beta_1 Num.reviews_{j(t-1)} + \beta_2 Av.score_{j(t-1)} + \beta_3 SDscore_{j(t-1)} + \delta LOS_{ijt} \\ + \gamma X_{ijt} + \omega Room.type_{ijt} + H_j + \tau_t + \lambda_1 Temp.Contiguity_{ijt} + \lambda_2 Weekend_{ijt} \\ + \epsilon_{ijt} \end{aligned} \tag{3}$$

where *i* indexes individual reviews, *j* the hotel and *t* the month of the review, $Num.reviews_{jt-1}$, $Av.score_{jt-1}$ and $SDscore_{jt-1}$ capture expected quality through the quantity (stock), level and dispersion of pre-existing ratings (up to one month before the stay); LOS_{ijt} is the guest's length of stay at hotel *j* in period *t*; X_{ijt} are tourist travel party composition and country of origin fixed effects capturing heterogeneity in preferences associated with cultural traits and travel distance (Litvin, 2019; Mariani and Predvoditeleva, 2019); $Room.type_{ijt}$ are dummy indicators for the type of room, capturing quality and price differences within the same hotel that might impact consumer satisfaction; H_j are hotel fixed effects controlling for objective mean quality differences across hotels; τ_t are monthly fixed effects; $Temp.Contiguity_{ijt}$ and $Weekend_{ijt}$ are the dummies defined before; and ϵ_{ijt} is a random error term capturing the emotional component of satisfaction that is allowed to be heteroskedastic as follows:

$$\begin{gathered} \epsilon_{ijt} \sim N(0, {\sigma_\epsilon}^2) \\ {\sigma_\epsilon}^2 = exp\left(\eta + \pi LOS_{ijt}\right) \end{gathered} \tag{4}$$

The additive structure of remembered utility in (2) implies that the value of ${\sigma_\epsilon}^2$ informs about the relative importance of the stochastic over the deterministic components. In the extreme case that ${\sigma_\epsilon}^2 = 0$, ratings would be fully deterministic. As long as ${\sigma_\epsilon}^2$ becomes greater, scores are noisier.

The model in (3)-(4) is a linear regression with heteroskedastic errors that is estimated in one step by Maximum Likelihood. The estimates from the two-step Generalized Least Squares (henceforth GLS) estimator proposed by Harvey (1976) are also presented as a robustness check (see subsection 5.3). The exponential transformation for ${\sigma_\epsilon}^2$ ensures the variance is positive for all possible values of $\eta + \pi LOS_{ijt}$. The key parameter of interest is π, which captures how the variability of scores depends on tourists' length of stay at the hotel.[6]

5. RESULTS

5.1.Main findings

Table 2 reports the coefficient estimates for the model in (3)-(4). Standard errors are clustered at the hotel level to capture potential cross-correlation in individual reviews for the same hotel over time. The country-of-origin fixed effects are visually presented in Figures 2 and 3.[7] The monthly and hotel fixed are omitted to save space but are available upon request.[8]

In the upper part of Table 2, we report the results for the mean equation. Columns 1-3 consider $Num.reviews_{jt-1}$, $Av.score_{jt-1}$ and $SDscore_{jt-1}$ in the regression separately, whereas Column 4 includes the three together. There is high consistency in the estimates across specifications so what follows focuses on the results from the full specification in Column 4. We document that

[6] One could argue that the censored nature of the dependent variable would need a Tobit estimator rather than OLS for the expected value of scores. We prefer to use a linear regression for the mean equation since Tobit is known to suffer from incidental parameter bias under high-dimensional fixed effects (Greene, 2004). Moreover, unlike OLS, Tobit renders inconsistent estimates when the disturbances are non-normal (Arabmazar and Schmidt, 1982).

[7] There is a total of 228 origin countries in the data. We include dummies for 83 origin countries, which represent 98.5% of the sample. The reference category considers all the remaining origins, which have few observations each.

[8] See Figure A1 in Appendix for a histogram of the hotel fixed effects for Column 4 in Table 2.

scores are uncorrelated with the pre-consumption stock volume of reviews. This might be explained by the hotel fixed effects already capturing level differences in reviews across hotels. Nevertheless, we find that, conditional on hotel fixed effects, higher average scores from past guests translate into lower individual ratings. Specifically, a unit increase in the mean score the month before the stay is associated with a 0.112-point decrease in the individual rating. This result can be attributed to feelings of disappointment when expectations are not met (Mitchell et al., 1997) and closely relates to the 'good news-bad news' paradigm. When consumers observe that past evaluations for a hotel are relatively high (low), individuals set up high (low) expectations before consumption. Theory predicts that good news (experienced utility is greater than expected) are under-weighted relative to bad news (Eil and Rao, 2011; Nguyen and Claus, 2013). Therefore, unmet expectations through high benchmarks result in lower experienced utility (and vice versa).

Interestingly, greater volatility in ratings the month before the stay is associated with higher reported level of satisfaction. In particular, a unit increase in SDScore leads to a 0.09 increase in individual scores, everything else being equal. This result is striking, as we would expect the opposite sign direction. Indeed, the raw data suggests an (unconditional) negative association between pre-consumption ratings' volatility and post-consumption individual scores. Since the estimate for SDScore is conditional on hotel fixed effects, it needs to be interpreted as follows: for the same hotel and mean score, greater pre-consumption inconsistency among reviews is associated with higher post-consumption ratings. This result could be potentially explained by people paying more attention to negative than to positive reviews in cases of high variability in scores. This sets lower expectations pre-consumption that might result in better valuations post-consumption given quality through anchoring.

Concerning the role of stay duration on mean scores, an additional night decreases scores by 0.006 points, *ceteris paribus*. Nonetheless, although significant, the effect is quantitatively small. This result falls in line with prior works (Brandes and Dover, 2022; Leoni and Moretti, 2023), suggesting that long stayers exhibit lower reported utility. One argument could be that the longer a consumer stays, the more time he/she has to critically evaluate the services offered (Kim and Han, 2022); for instance, comfort levels might decrease over time in rooms that are generally designed for short stays.

Table 2. Results from heteroskedastic linear regression

	(1)	(2)	(3)	(4)
Dep. Variable: score	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Num.reviews (t-1)	-3.1e-06			-2.4e-06
	(1.3e-05)			(1.2e-05)
Av.Score (t-1)		-0.168***		-0.112***
		(0.031)		(0.041)
SDScore (t-1)			0.163***	0.087**
			(0.027)	(0.037)
LOS	-0.006**	-0.006**	-0.006**	-0.006**
	(0.003)	(0.003)	(0.002)	(0.002)
Travel party: couple	-0.005	-0.005	-0.005	-0.005
	(0.014)	(0.014)	(0.014)	(0.014)
Travel party: family	-0.007	-0.007	-0.007	-0.007
	(0.015)	(0.015)	(0.015)	(0.015)
Travel party: group	0.068***	0.068***	0.068***	0.068***
	(0.013)	(0.013)	(0.013)	(0.013)
Domestic	0.034***	0.034***	0.034***	0.034***
	(0.012)	(0.012)	(0.012)	(0.012)
Room type: economy	0.019	0.018	0.019	0.018
	(0.028)	(0.028)	(0.028)	(0.028)
Room type: standard	-0.012	-0.012	-0.012	-0.012
	(0.013)	(0.013)	(0.013)	(0.013)
Room type: superior	-0.016	-0.016	-0.016	-0.016
	(0.016)	(0.016)	(0.016)	(0.016)
Anonymous	-0.198***	-0.198***	-0.198***	-0.198***
	(0.011)	(0.011)	(0.011)	(0.011)
Temp.Contiguity	0.058***	0.058***	0.058***	0.058***
	(0.008)	(0.008)	(0.008)	(0.008)
Weekend	0.002	0.002	0.002	0.002
	(0.006)	(0.006)	(0.006)	(0.006)
Constant	7.241***	8.638***	7.064***	8.076***
	(0.036)	(0.259)	(0.044)	(0.381)
Hotel fixed effects	YES	YES	YES	YES
Country of origin fixed effects	YES	YES	YES	YES
Monthly fixed effects	YES	YES	YES	YES
Variance equation				
LOS	-0.021***	-0.021***	-0.021***	-0.021***
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.644***	0.644***	0.644***	0.644***
	(0.020)	(0.020)	(0.020)	(0.020)
Observations	525,437	525,437	525,437	525,437

Clustered standard errors at the hotel level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Regarding travel party composition, and compared to *solo* guests, those traveling in groups report higher satisfaction (+0.068). Nonetheless, no significant differences are found for couples and those staying with their family. Interestingly, domestic guests are, on average, less satisfied than foreigners (-0.034). In this vein, the country-of-origin fixed effects (Figures 2 and 3) point to important heterogeneity in reported utility by nationality. This falls in line with previous works documenting that cultural dimensions are relevant predictors of differences in rating scores (Litvin, 2019; Mariani and Predvoditeleva, 2019), plausibly through different quality benchmarks. Surprisingly, there are no significant differences in reported satisfaction associated to room quality. This finding indicates that, conditional on hotel fixed effects, experienced utility is unrelated to quality differences within hotels, on average. A potential explanation for this is that, as explained in Section 4, ratings are derived from the average of six items that capture hotel quality as a whole rather than the specific attributes of individual rooms.

Figure 2. Country-of-origin fixed effects estimates from Column 4 in Table 2 (I)

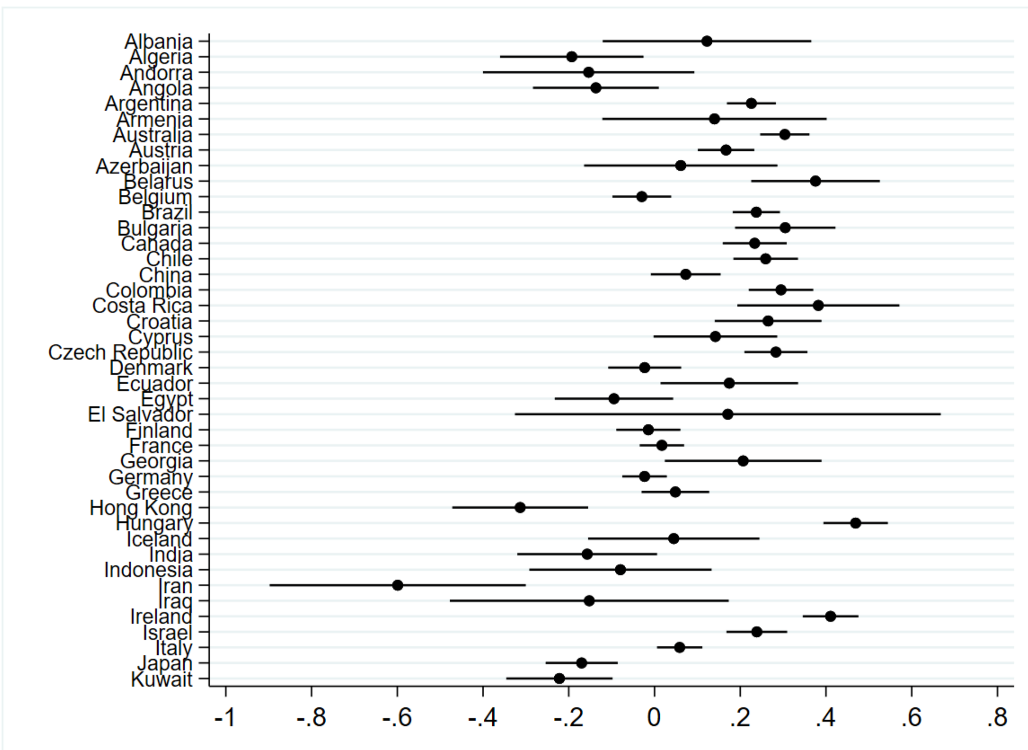
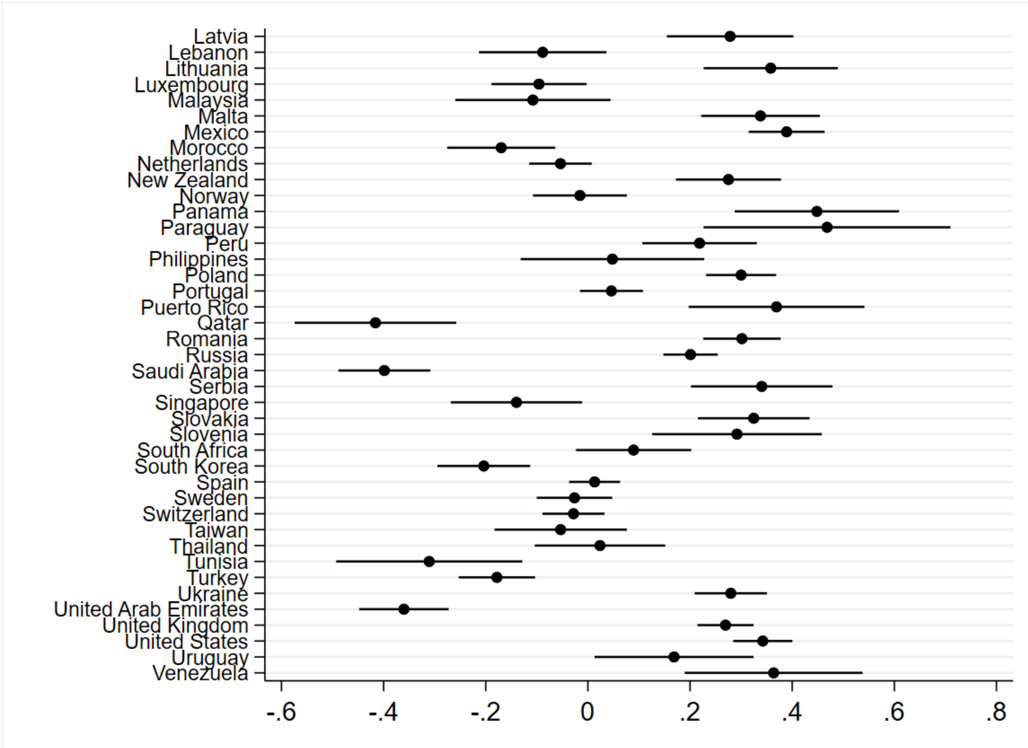


Figure 3. Country-of-origin fixed effects estimates from Column 4 in Table 2 (II)



In line with identity disclosure and deindividualization theories (Deng et al., 2021), we find that anonymous guests tend to provide lower ratings. More precisely, other things being equal, review anonymity is associated with a 0.198-point drop in scores. A plausible mechanism is that hiding their identities allows guests to share their honest opinions without the fear of being identified. Furthermore, the temporal proximity between the stay and the review date (temporal contiguity) is associate with higher scores (+0.058). This suggests that individuals report greater experienced utility for consumption episodes that are temporally close to the time of the review. This relates to Mullainathan (2002) framework of biased memory limitations and rosy retrospection theory (Mitchell et al., 1997): temporal contiguity may contribute to better recall and remembrance of details, which tends to be biased towards good aspects.

Moving to the variance equation, a LR test rejects the null hypothesis of homoscedastic errors (chi2(1)=29.8, p-value<0.001).[9] The coefficient estimate for LOS is negative and statistically significant, indicating that the variance decreases with the length of the stay at the hotel. This result is in line with our theoretical argument, showing that shorter consumption episodes result in more volatile ratings. We interpret it as evidence that ratings by short stayers are more likely to integrate negative and positive shocks. In contrast, evaluations from long-stayers are more deterministic, hence mostly driven by objective dimensions. Accordingly, our findings confirm that the length of stay shapes ratings polarization, with short stays providing noisy signals of the true (objective) hotel quality. Based on the coefficient estimates, we have computed the predicted standard deviation of the error component ($\widehat{\sigma_\epsilon}$) as the square root of $\widehat{\sigma_\epsilon^2}$, with $\widehat{\sigma_\epsilon^2} = exp(\hat{\eta} + \hat{\pi} LOS_{it})$. Table A3 in Appendix presents how $\widehat{\sigma_\epsilon}$ varies with length of stay in the sample. We see that the predicted standard deviation in scores decreases from 1.37 for one-night guests to 1.18 for 15-night guests.

5.2. Heterogeneity by consumer profile

The pooled analysis presented before might mask relevant heterogeneity associated with guests' profile. For instance, the role of length of stay and ex-ante quality signals on individual ratings might vary depending on consumers' nationality (domestic versus international guests) or whether the individual travels alone, in a couple or with the family/in a group. To examine this, Table 3 presents the results from separate regressions.[10]

[9] Table A2 in Appendix presents the results from a baseline linear regression under the assumption of homoscedastic errors. A Breusch-Pagan test rejects the null hypothesis of constant variance (chi(1)=32380.8, p-value<0.0001), suggesting the presence of heteroskedastic errors and the suitability of our modelling approach.

[10] Figures A2 and A3 Appendix present binned scatterplots of the residualized conditional mean relationship between scores and length of stay by nationality and travel party composition. We use Frisch-Waugh-Lovell theorem to visually present the differences in slopes by subsamples conditional on all the remaining controls, hotel, monthly and country-of-origin fixed effects.

Table 3. Results from heteroskedastic linear regression per subsamples

	(1)	(2)	(3)	(4)	(5)
Subsample	Alone	Couple	Family/Group	Domestic	Foreigner
Dep. Variable: score	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Num.reviews (t-1)	1.6e-05	-6.9e-06	-1.5e-05*	-1.0e-05	-2.5e-06
	(1.6e-05)	(2.1e-05)	(8.6e-06)	(2.8e-05)	(1.0e-05)
Av.Score (t-1)	-0.060	-0.044	-0.206***	-0.084	-0.138***
	(0.091)	(0.054)	(0.053)	(0.061)	(0.048)
SDScore (t-1)	0.135*	0.111**	0.042	0.011	0.102**
	(0.079)	(0.050)	(0.054)	(0.060)	(0.045)
LOS	0.002	-0.009***	-0.010***	-0.025***	-0.005*
	(0.004)	(0.003)	(0.004)	(0.007)	(0.003)
Travel party: couple				0.035**	-0.021
				(0.017)	(0.016)
Travel party: family				0.079***	-0.028*
				(0.022)	(0.017)
Travel party: group				0.105***	0.053***
				(0.023)	(0.017)
Domestic	-0.011	0.056***	0.039**		
	(0.031)	(0.016)	(0.017)		
Room type: economy	-0.011	-0.027	-0.001	-0.072*	0.035
	(0.029)	(0.035)	(0.025)	(0.040)	(0.029)
Room type: standard	-0.017	-0.009	-0.004	-0.036*	-0.008
	(0.025)	(0.020)	(0.017)	(0.021)	(0.014)
Room type: superior	-0.111***	-0.007	0.003	-0.054*	-0.009
	(0.033)	(0.019)	(0.022)	(0.029)	(0.016)
Anonymous	-0.265***	-0.183***	-0.174***	-0.259***	-0.184***
	(0.023)	(0.019)	(0.015)	(0.026)	(0.011)
Temp.Contiguity	0.029*	0.064***	0.060***	0.081***	0.052***
	(0.016)	(0.011)	(0.012)	(0.017)	(0.009)
Weekend	0.009	-0.005	0.003	-0.006	0.003
	(0.014)	(0.008)	(0.009)	(0.013)	(0.006)
Constant	5.402***	9.608***	11.264***	5.567***	10.620***
	(0.822)	(0.535)	(0.490)	(0.547)	(0.450)
Hotel fixed effects	YES	YES	YES	YES	YES
Country of origin fixed effects	YES	YES	YES	NO	YES
Monthly fixed effects	YES	YES	YES	YES	YES
Variance equation					
LOS	-0.021**	-0.025***	-0.015***	0.008	-0.019***
	(0.008)	(0.005)	(0.005)	(0.010)	(0.004)
Constant	0.683***	0.590***	0.647***	0.624***	0.627***
	(0.029)	(0.022)	(0.023)	(0.025)	(0.022)
Observations	101,036	212,857	211,544	97,604	427,833

Clustered standard errors at the hotel level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

We find that the negative relationship between length of stay and mean scores is quantitatively greater for domestic guests (-0.025) as compared to the pooled regression in Table 2 (-0.006) but non-significant (at 95% confidence level) for foreign and solo travelers. That is, the deterioration in satisfaction with hotel services as the individual stays for longer only holds for national travelers. For foreigners, the conditional mean of scores is unrelated to the consumption span. On the contrary, the decline in scores' volatility for longer stays applies to all segments except for national guests. For this segment, conditional on the negative trend on the conditional mean, we document that the variance in ratings is uncorrelated to how long the individual has stayed at the hotel. These two results clearly indicate the link between post-consumption valuation of hotel services and the consumption span differs by the origin of the guest.

One potential explanation could be that domestic travelers might have a better knowledge about hotel quality beforehand that makes the gap between experienced utility and ex-ante expectations narrower. This better knowledge might be due to hotel quality being more homogeneous within countries due to common regulatory frameworks and the greater probability of having previous experience, either at the same hotel or at the same hotel chain. In other words, their individual valuations compile more with the average and exhibit less volatility. For foreigners, instead, service quality might depart more from their benchmarks in either direction, making their evaluations noisier.

Also interesting, the negative association between individual scores and the pre-consumption mean value of the stock of reviews only applies to foreign guests and those travelling in groups or with the family. Accordingly, our interpretation that high expectations about quality through a greater ex-ante overall mean result in lower post-consumption individual ratings only applies to foreigners and those travelling in groups. Furthermore, the non-significant influence of room type on scores holds for all segments with the exception of solo travelers, who value superior rooms significantly less (-0.11 points). Overall, these regressions point to relevant heterogeneity in the drivers of remembered utility about hotel stays by guest profile.

5.3. Robustness checks

Possibly the most important threat to identification is the self-selection of long-stayers into hotels for whom available information about expected quality is more precise. Under risk aversion, an individual that is planning to stay for several days in a hotel might pick one for whom quality information is more consistent. To exclude potential selection effects driving our results, we have regressed length of stay on the number, mean value and standard deviation of ratings the month before, customer characteristics and country of origin, hotel and monthly fixed effects. The results are shown in Appendix, Table A4. We find that length of stay is uncorrelated with the volatility in the stock of ratings' score one month before the stay. Accordingly, our findings do not seem to be driven by long-stayers self-selecting into hotels with less polarized expected quality.

Although the model in (3)-(4) controls for hotel fixed effects and therefore exploits the within hotel variation in length of stay across consumers, one could plausibly argue that if long-stayers decide to lodge at high-quality hotels, identification problems could emerge because high ratings imply necessarily high consistency among reviewers (all people agree it is a good quality hotel) and therefore our results concerning ratings' volatility would be driven by a quality mechanism and not by the length of the consumption episode. To explore this, we have regressed the mean LOS per hotel on the hotel fixed effects estimates from Table 2 as quality indicators (Appendix, Table A5). The coefficient estimate is non-significant, implying that average stay duration does not correlate with hotel quality.

A third concern could be that Booking.com changed its review system from the average of six aspects on a 2.5-10 scale to an overall evaluation on a 1-10 scale (Mellinas and Martín-Fuentes, 2021). Since we use data for the old review system, our empirical analysis might have low external validity under the current system. We have chosen ratings' data under the old system because the time span under the new system and before COVID-19 pandemic is pretty short (September 2019-February 2020). In this regard, using data for COVID-19 periods could be problematic as some studies point to changes in satisfaction associated with the pandemic context (e.g., Leoni and Moretti, 2023). Nonetheless, to examine the robustness of our findings to the review system typology, we have redone the analysis using data for the same cities and hotels during the period

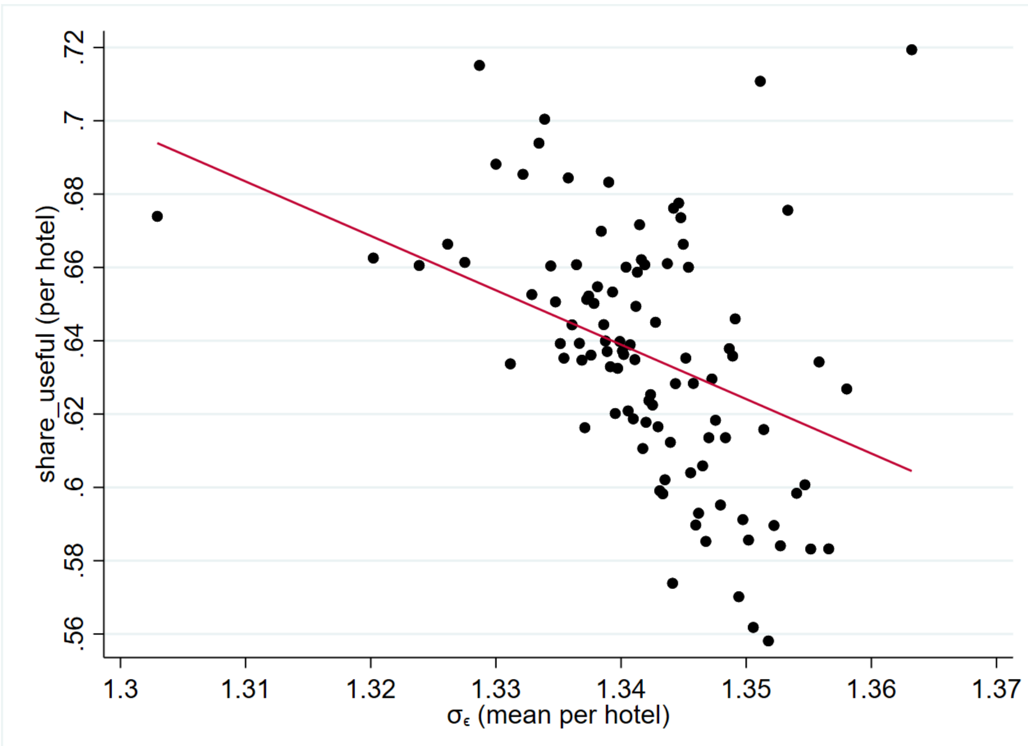
September 2019-February 2020. Table A6 in Appendix reports the estimates. The results are very similar and consistent with the ones in Table 2, with the exception that there are no differences in ratings between domestic and foreign guests. We again document a negative and significant effect of length of stay on the variance of the stochastic term, which is slightly larger in magnitude given that the new scale is wider.

Some additional robustness checks have been performed. First, we have conducted a stepwise estimation where the control variables have been added to the regression sequentially. The estimates are robust across model specifications as compared to Table 2 (Appendix, Table A7). The only exception is the one-month lag of the hotel mean score, which turns to be negatively associated with individual ratings once we control for hotel fixed effects. Second, we have used two-month lags (instead of one) of the stock of reviews, average scores, and standard deviation of scores (Appendix, Table A8). The estimates are very similar, offering great robustness. Third, we have performed separate regressions by city (Appendix, Table A9). Our core findings remain pretty unchanged across cities, although LOS is not significant for explaining scores' variance in Milan and Madrid. Fourth, we have re-estimated our model using a GLS estimator instead of Maximum Likelihood. The results are shown in Appendix (Table A10) and are very similar to the ones presented in Table 2. Finally, we have allowed for a non-linear relationship between length of stay and ratings by including the square of LOS. However, the squared term is not statistically significant, suggesting a linear pattern (available upon request).

5.4.The (low) usefulness of volatile ratings

The preceding section has shown that scores' volatility decreases as the guest stays for longer. As discussed in Section 2, greater polarization in ratings makes the quality signal noisier and therefore average ratings less reliable. To deepen further into this issue, we have calculated the average predicted variance from Column 4 in Table 2 for each hotel in the sample (denoted as $\widehat{\sigma_{\epsilon}}_j$). Next, we have computed the share of individual reviews received by each hotel deemed as useful by other consumers (denoted as $share\_useful_j$). This refers to the percentage of reviews that received "helpful" (thumbs up) flags. Figure 4 below presents a binned scatterplot of the pairwise relationship between the two. We clearly see that the share of useful reviews negatively correlates with the variance of the scores predicted by our model. In line with previous work on technological devices (Lee et al., 2021; Mudambi and Schuff, 2010), a greater variance in hotel scores diminishes the informative power of user-generated content, making it harder for prospective buyers to draw accurate conclusions on the expected quality. This pattern is statistically significant as indicated by an OLS regression of $share\_useful_j$ on $\widehat{\sigma_\epsilon}_j$ shown in Table 4.

Figure 4. Binned scatterplot of the share of useful reviews per hotel on the predicted review variance per hotel



Note: the bins are defined based on the 100 quantiles of the distribution of the variables

Table 4. Coefficient estimates from OLS regression of $share\_useful_j$ on $\widehat{\sigma_\epsilon}_j$ at the hotel level

Dep. Variable: $share\_useful_j$	Coef. (SE)
$\widehat{\sigma}_{\epsilon j}$	-1.484**
	(0.683)
Constant	2.627***
	(0.917)

Observations	1,233

Bootstrapped standard errors after 1,000 replications in parentheses. *** p<0.01, ** p<0.05, * p<0.1

6. DISCUSSION AND CONCLUSIONS

Online platforms nowadays play a crucial role in many markets by facilitating transactions and addressing the problem of asymmetric information for experience goods. Online user generated content is a major source of information for consumers when searching goods and services, contributing to a better ex-ante assessment of uncertain quality through social learning. However, the effectiveness of online reviews in reducing information asymmetries strongly depends on their informational value (Acemoglu et al., 2022). In fact, when choosing among alternative goods, consumers form prior beliefs based on the reported degree of satisfaction by previous consumers, which might disagree in their assessment of the service quality. Greater inconsistency (incongruence) decreases transactions and sales through mismatch costs. On the contrary, if the information provided by online reviews is consistent, it is easier for consumers to make informed decisions. Polarized ratings (high variance) are thus noisy signals of the true quality.

Grounded on the *Law of Comparative Judgement* (Thurstone, 1927) and the model developed by Zimmerman et al. (2018), we posit that agents consuming the same service report different levels of quality because they react differently to (about) the same service quality. Our framework postulates that consumers’ subjective evaluation (individual rating) can be partitioned as the sum of a deterministic and a stochastic part. Using web-scrapped data for more than 225,000 individual reviews from Booking.com platform in Barcelona, Madrid, Milan, Rome and Lisbon and using heteroskedastic regressions, we have found that the volatility of ratings decreases for longer stays at the hotel. That is, spot consumption seems to make ratings highly volatile and mostly driven by emotional factors. This pattern is likely explained by a better knowledge of the type, quality, and variety of hotel characteristics among long-stayers that make their opinions to converge to the average. Since consumers do not linearly aggregate the utilities experienced at each episode of the stay (Kahneman and Thaler, 2006), our results are also potentially explained by hedonic adaptation theories (Rayo and Becker, 2007), according to which unexpected positive and negative shocks in service quality tend to vanish throughout longer consumption spans. Moreover, memory errors in recall *à la* Mullainathan (2002) are likely to play a role, since long stays are more accurately remembered.

Our analysis also reveals that higher overall scores from previous guests, which set higher expectations about service quality beforehand, are associated with lower individual ratings. In line with the good news-bad news paradigm and social learning theories, post-consumption individual ratings are highly affected by ex-ante expectations. Unmet expectations induce consumers to 'punish' the hotel through negative reported satisfaction. Additionally, we have documented substantial heterogeneity in ratings depending on the travel party composition, the nationality of the guest and the latency between the stay and the review time. Our findings are pretty robust, as they remain consistent under the new review system implemented by Booking.com since September 2019.

This study makes two contributions to the existing literature on online reviews. Firstly, we offer the first analysis on how the consumption span shapes ratings' volatility. Existing studies have primarily paid attention to mean ratings, with only a few analysing the role played by the length of the stay at the hotel (e.g., Brandes and Dover, 2022). In doing so, we have shown that long-stayers are more deterministic in their valuations, whereas the ratings by short-stayers are noisier and potentially more affected by emotional (unobserved) factors. Secondly, this study adds to the growing literature on the instrumental value of online reviews as a social learning tool in contexts of quality uncertainty. Existing studies that have analysed reviews' usefulness have mainly looked at the number of words used in the textual comment, the presence of pictures, or the reviewers' expertise. We add novel evidence on this respect, documenting that ratings' consistency in terms of low dispersion is an important metric for prospective consumers at the time of judging the informativeness of user generated content.

Our findings offer some relevant implications, particularly for online platforms. Given the non-neutral effects of online reputation for firms' performance (Chevalier and Mayzlin, 2006) and rating behaviours (Cicognani et al., 2022), platforms should consider the possibility of assigning different weights to reviews based on the length of stay. This could help to mitigate the effects of polarization and improve the accuracy of online evaluations. Stated differently, the weight allotted to each review in calculating the overall score should be based on the guest's level of exposure to the hotel facilities, as this serves as a proxy for their subjective understanding of the intrinsic quality. As an alternative, another valuable possibility is to incorporate length of stay (short vs long stay) as a criterion for filtering reviews. Platforms typically provide users the option to sort reviews based on different criteria, such as the language, overall score, time of year or the type of travel party. Adding this duration filter may enhance the usefulness of ratings to consumers by allowing them to focus on the valuations made by consumers with similar stay duration, which might hence share more similar needs. Furthermore, the negative relationship between length of stay and rating scores indicates that hotels must adapt their services more carefully to the specific needs and requirements of long stayers, who represent an important market segment.

The paper has some limitations that should be acknowledged. First, there is scope for potential bias from reviewers' self-selection: consumers who submit an online review are not a random sample of the population and tend to be those exhibiting extreme opinions (Schoenmueller et al., 2020). This is a common limitation when working with review data. Nonetheless, we believe our results are still highly informative. Even if not truly representative of the population of consumers, it is existing reviews what matters for social learning and subsequent sales. Second, we lack information about actual rates paid by hotel guests, the time in advance with which the booking was made, and the leisure versus business motive of the stay. Future work should expand our analysis by considering the role played by these factors.

REFERENCES

Acemoglu, D., Makhdoumi, A., Malekian, A. and Ozdaglar, A. (2022). Learning from reviews: the selection effect and the speed of learning. *Econometrica*, 90(6), 2857-2899.

Akerlof, G.A. (1970). The market for 'lemons': quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488-500.

Amador, M. and Weill, P.O. (2012). Learning from private and public observations of others' actions. *Journal of Economic Theory*, 147, 910-940.

Anderson, L.R. and Holt, C.A. (1997). Information cascades in the laboratory. *American Economic Review*, 87(5), 847-862.

Arabmazar, A. and Schmidt, P. (1982). An investigation of the robustness of the Tobit estimator to non-normality. *Econometrica*, 50(4), 1055-1063.

Banerjee, A.V. (1992). A simple model of herd behavior. *The Quarterly Journal of Economics*, 107(3), 797-817.

Brandes, L. and Dover, Y. (2022). Offline context affects online reviews: The effect of post-consumption weather. *Journal of Consumer Research*, 49(4), 595-615.

Brandes, L., Godes, D. and Mayzlin, D. (2022). Extremity bias in online reviews: The role of attrition. *Journal of Marketing Research*, 59(4), 675-695.

Brunner-Sperdin, A. and Peters, M. (2009). What influences guests' emotions? The case of high-quality hotels. *International Journal of Tourism Research*, 11, 171-183.

Cai, H., Chen, Y. and Fang, H. (2009). Observational learning: Evidence from a Randomized Natural Field Experiment. *American Economic Review*, 99(3), 864-882.

Chatterjee, S. (2020). Drivers of helpfulness of online hotel reviews: A sentiment and emotion mining approach. *International Journal of Hospitality Management*, 85, 102356.

Chevalier, J.A. and Mayzlin, D. (2006). The effect of word of mouth on sales: online book reviews. *Journal of Marketing Research*, 43(3), 345-454.

Cicognani, S., Figini, P. and Magnani, M. (2022). Social influence bias in ratings: A field experiment in the hospitality sector. *Tourism Economics*, 28(8), 2197-2218.

de Langhe, B., Fernbach, P.M. and Lichtenstein, D.R. (2016). Navigating by the stars: Investigating the actual and perceived validity of online user ratings. *Journal of Consumer Research*, 42(6), 817-833.

Deng, L., Sun, W., Xu, D. and Ye, Q. (2021). Impact of anonymity on consumers' online reviews. *Psychology & Marketing*, 38(12), 2259-2270.

Eil, D. and Rao, J.M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2), 114-138.

Engler, T.H., Winter, P. and Schulz, M. (2015). Understanding online product ratings: A customer satisfaction model. *Journal of Retailing and Consumer Services*, 27, 113-120.

Eurostat (2019). *Annual data on trips of EU residents*. Retrieved from https://ec.europa.eu/eurostat/ on March 9th, 2023.

Fang, L. (2022). The effects of online review platforms on restaurant revenue, consumer learning and welfare. *Management Science*, 68(11), 8116-8143.

Figini, P., Vici, L. and Viglia, G. (2020). A comparison of hotel ratings between verified and non-verified online review platforms. *International Journal of Culture, Tourism and Hospitality Research*, 14(2), 157-171.

Frick, M., Iijima, R. and Ishii, Y. (2020). Misinterpreting others and the fragility of social learning. *Econometrica*, 88(6), 2281-2328.

Galiani, S., Gertler, P.J. and Undurraga, R. (2018). The half-life of happiness: hedonic adaptation in the subjective well-being of poor slum dwellers to the satisfaction of basic housing needs. *Journal of the European Economic Association*, 16(4), 1189-1233.

Greene, W. (2004). The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econometrics Journal*, 7, 98-119.

Guizzardi, A., Mariani, M. M., and Stacchini, A. (2022). A temporal construal theory explanation of the price-quality relationship in online dynamic pricing. *Journal of Business Research*, 146, 32-44.

Harvey, A.C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, 44(3), 461-465.

Huang, S., Chang, C.T., Bilgihan, A. and Okumus, F. (2020). Helpful or harmful? A double-edged sword of emoticons in online review helpfulness. *Tourism Management*, 81, 104135.

Kahneman, D., Fredrickson, B., Schreiber, C.M. and Redelmeir, D. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science*, 4(6), 401–405.

Kahneman, D. and Thaler, R.H. (2006). Utility maximization and experienced utility. *Journal of Economic Perspectives*, 20(1), 221-234.

Kahneman, D., Wakker, P.P. and Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. *The Quarterly Journal of Economics*, 112(2), 375-406.

Kim, J.M. and Han, J. (2022). Impact of the length of stay at hotels on online reviews. *International Journal of Contemporary Hospitality Management*, 34(4), 1249-1269.

Kim, Y. and Krishnan, R. (2015). On product-level uncertainty and online purchase behavior: An empirical analysis. *Management Science*, 61(10), 2449-2467.

Lee, N., Bollinger, B. and Staelin, R. (2023). Vertical versus horizontal variance in online reviews and their impact on demand. *Journal of Marketing Research*, 60(1), 130-154.

Leibenstein, H. (1950). Bandwagon, snob, and Veblen effects in the theory of consumer demand. *The Quarterly Journal of Economics*, 64(2), 183-207.

Leoni, V. and Moretti, A. (2023). Customer satisfaction during COVID-19 phases: the case of the Venetian hospitality system. *Current Issues in Tourism*, https://doi.org/10.1080/13683500.2022.2164709

Liang, S., Schuckert, M. and Law, R. (2019). How to improve the stated helpfulness of hotel reviews? A multilevel approach. *International Journal of Contemporary Hospitality Management*, 31(2), 953-977.

Litvin, S.W. (2019). Hofstede, cultural differences, and TripAdvisor hotel reviews. *International Journal of Tourism Research*, 21, 712-717.

Liu, Z. and Park, S. (2015). What makes a useful online review? Implication for travel product websites. *Tourism Management*, 47, 140-151.

Lo, A.S. and Yao, S.S. (2019). What makes hotel online reviews credible? An investigation of the roles of reviewer expertise, review rating consistency and review valence. *International Journal of Contemporary Hospitality Management*, 131(1), 41-60.

Liu, X., Li, C., Nicolau, J.L. and Han, M. (2023). The value of rating diversity within multidimensional rating system: Evidence from booking platform. *International Journal of Hospitality Management*, 110, 103434.

Liu, Y. (2006). Word of mouth for movies: its dynamic and impact on box office revenue. *Journal of Marketing*, 70, 74-89.

Loewenstein, G. (2000). Emotions in economic theory and economic behavior. *American Economic Review*, 90(2), 426-432.

Magnani, M. (2020). The economic and behavioral consequences of online user reviews. *Journal of Economic Surveys*, 34(2), 263–292.

Mariani, M.M. and Borghi, M. (2018). Effects of the Booking.com rating system: Bringing hotel class into the picture. *Tourism Management*, 66, 47-52.

McFadden, D. (1974). *Conditional Logit Analysis of Qualitative Choice Behavior*. In Paul Zarembka (ed.). Frontiers Econometrics, New York: Academic Press.

Mellinas, J.P. and Martin-Fuentes, E. (2021). Effects of booking. com's new scoring system. *Tourism Management*, 85, 104280.

Mitchell, T.R., Thompson, L., Peterson, E. and Cronk, R. (1997). Temporal adjustments in the evaluation of events: The "rosy view". *Journal of Experimental Social Psychology*, 33(4), 421-448.

Moe, W.W. and Schweidel, D.A. (2012) Online product opinions: incidence, evaluation, and evolution. *Marketing Science*, 31(3), 372–386.

Moe, W.W. and Trusov, M. (2011). The value of social dynamics in online product ratings forums. *Journal of Marketing Research*, 48, 444-456.

Mariani, M. and Predvoditeleva, M. (2019). How do online reviewers' cultural traits and perceived experience influence hotel online ratings? An empirical analysis of the Muscovite hotel sector. *International Journal of Contemporary Hospitality Management*, 31(12), 4543-4573.

Mayzlin, D., Dover, Y. and Chevalier, J. (2014). Promotional reviews: an empirical investigation of online review manipulation. *American Economic Review*, 104(8), 2421-2455.

Mudambi, S. and Schuff, D. (2010). What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Quarterly*, 34, 185–200.

Mullainathan, S. (2002). A memory-based model of bounded rationality. *The Quarterly Journal of Economics*, 117(3), 735-774.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2), 311-329.

Nguyen, V.H. and Claus, E. (2013). Good news, bad news, consumer sentiment and consumption behavior. *Journal of Economic Psychology*, 39, 426-438.

Oliver, R.L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. *Journal of Marketing Research*, 17(4), 460-469.

Oswald, A.J. and Powdthavee, N. (2008). Does happiness adapt? A longitudinal study of disability with implications for economists and judges. *Journal of Public Economics*, 92, 1061-1077.

Park, s. and Park, D. (2013). The effect of low- versus high-variance in product reviews on product evaluation. *Psychology & Marketing*, 30(7), 543-554.

Pollak, R.A. and Wales, T.J. (1981). Demographic variables in demand analysis. *Econometrica*, 49(6), 1533-1551.

Pokryshevskaya, E.B. and Antipov, E.A. (2017). Profiling satisfied and dissatisfied hotel visitors using publicly available data from a booking platform. *International Journal of Hospitality Management*, 67, 1-10.

Pourfakhimi, S., Duncan, T. and Coetzee, W.J.L. (2020). Electronic word of mouth in tourism and hospitality consumer behaviour: state of the art. *Tourism Review*, 75(4), 637-661.

Proserpio, D. and Zervas, G. (2017). Online reputation management: estimating the impact of management responses on consumer reviews. *Marketing Science*, 36(5), 645-665.

Rabin, M. and Schrag, J.L. (1999). First impressions matter: a model of confirmatory bias. *The Quarterly Journal of Economics*, 114(1), 37-82.

Rayo, L. and Becker, G.S. (2007). Evolutionary efficiency and happiness. *Journal of Political Economy*, 115(2), 302-337.

San Martín, H. and Rodríguez-del-Bosque, I.A. (2008). Exploring the cognitive-affective nature of destination image and the role of psychological factors in its formation. *Tourism Management*, 29(2), 263-277.

Schoenmueller, V., Netzer, O. and Stahl, F. (2020). The polarity of online reviews: prevalence, drivers and implications. *Journal of Marketing Research*, 57(5), 853-877.

Schreiber, C. A. and Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General*, 129(1), 27–42.

Sirakaya, E., Petrick, J. and Choi, H. (2004). The role of mood on tourism product evaluations. *Annals of Tourism Research*, 31(3), 517-539.

Sun, M. (2012). How does the variance of product ratings matter? *Management Science*, 58(4), 696-707.

Sunder, S., Kim, K.H. and Yorkston, E.A. (2019). What drives herding behavior in online ratings? The role of rater experience, product portfolio, and diverging opinions. *Journal of Marketing*, 83(6), 93-112.

Thurstone, L.L. (1927). Three psychophysical laws. *Psychological Review*, 34(6), 424–43

UNWTO (2019). *Yearbook of Tourism Statistics*. Retrieved from: https://www.unwto.org/tourism-statistics/key-tourism-statistics. Accessed on March 9, 2023.

Vogt, C.A. and Andereck, K.L. (2003). Destination perceptions across a vacation. *Journal of Travel Research*, 41, 348-354.

West, P.M. and Broniarczyk, S.M. (1998). Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus. *Journal of Consumer Research*, 25, 38–51.

Wu, C., Che, H., Chan, T.Y. and Lu, X. (2015). The economic value of online reviews. *Marketing Science*, 34(5), 739-754.

Xu, X. (2022). A growing or depreciating love? Linking time with customer satisfaction through online reviews. *Information & Management*, 59, 103605.

Ye, Q., Law, R. and Gu, B. (2009). The impact of online user reviews on hotel room sales. *International Journal of Hospitality Management*, 28(1), 180–182.

Zervas, G., Proserpio, D. and Byers, J.W. (2021). A first look at online reputation on Airbnb, where every stay is above average. *Marketing Letters*, 32, 1-16.

Zhang, X., Zhang, X., Liang, S., Yang, Y. and Law, R. (2023). Infusing new insights: How do review novelty and inconsistency shape the usefulness of online travel reviews? *Tourism Management*, 96, 104703.

Zhao, Y., Yang, S., Narayan, V. and Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science*, 32(1), 153-169.

Zimmermann, S., Hernnmann, P., Kundisch, D. and Nault, B.R. (2018). Decomposing the variance of consumer ratings and the impact on price and demand. *Information Systems Research*, 29(4), 984-1002.

ONLINE APPENDIX

Table A1. Descriptive statistics of number of reviews per country of origin of the guest

Country	**Obs**
Abkhazia, Georgia	121
Afghanistan	8
Albania	362
Algeria	1114
Andorra	302
Angola	891
Argentina	15517
Armenia	172
Aruba	13
Australia	10680
Austria	6193
Azerbaijan	322
Bahamas	10
Bahrain	306
Bangladesh	76
Barbados	11
Belarus	723
Belgium	9645
Belize	6
Benin	9
Bermuda	17
Bolivia	180
Bonaire St Eustatius and Saba	2
Bosnia and Herzegovina	144
Botswana	15
Brazil	25076
Brunei Darussalam	13
Bulgaria	1525
Burkina Faso	6
Cambodia	41
Cameroon	9
Canada	5919
Cape Verde	74
Cayman Islands	16
Chad	5
Chile	4226
China	5333
Colombia	3403
Costa Rica	590
Croatia	1137
Cuba	6
Curaçao	17
Cyprus	695
Czech Republic	3500
Côte d'Ivoire	64
Democratic Republic of Congo	16
Denmark	2866
Djibouti	5
Dominican Republic	251
East Timor	9
Ecuador	554
Egypt	1174
El Salvador	142

Equatorial Guinea	19
Estonia	792
Ethiopia	16
Faroe Islands	14
Fiji	8
Finland	2807
France	40286
French Guiana	35
French Polynesia	31
French Southern Territories	6
Gabon	24
Georgia	477
Germany	27443
Ghana	34
Gibraltar	95
Greece	4451
Greenland	6
Grenada	6
Guadeloupe	60
Guam	5
Guatemala	305
Guernsey	32
Guinea	9
Guinea-Bissau	19
Haiti	15
Honduras	71
Hong Kong	1200
Hungary	3901
Iceland	466
India	1582
Indonesia	523
Iran	303
Iraq	179
Ireland	6498
Isle of Man	32
Israel	10593
Italy	76829
Jamaica	14
Japan	4638
Jersey	77
Jordan	364
Kazakhstan	726
Kenya	71
Kosovo	38
Kuwait	1477
Kyrgyzstan	42
Laos	11
Latvia	909
Lebanon	1001
Lesotho	7
Libya	169
Liechtenstein	53
Lithuania	1123
Luxembourg	1671
Macao	170
Madagascar	10
Malaysia	626
Maldives	25
Mali	10
Malta	1096

Martinique	38
Mauritania	20
Mauritius	92
Mayotte	8
Mexico	3324
Moldova	205
Monaco	141
Mongolia	10
Montenegro	205
Morocco	2654
Mozambique	265
Myanmar	20
Namibia	22
Nepal	23
Netherlands	12914
New Caledonia	65
New Zealand	1536
Nicaragua	19
Nigeria	110
North Macedonia	230
Norway	2873
Oman	313
Pakistan	397
Palestinian Territory	46
Panama	420
Papua New Guinea	5
Paraguay	205
Peru	1237
Philippines	579
Poland	6350
Portugal	23028
Puerto Rico	421
Qatar	1005
Reunion	226
Romania	4299
Russia	22077
Rwanda	10
San Marino	46
Saudi Arabia	4280
Senegal	62
Serbia	825
Seychelles	27
Singapore	1029
Slovakia	1435
Slovenia	585
South Africa	1625
South Korea	4370
Spain	66954
Sri Lanka	112
Sudan	18
Sweden	5657
Switzerland	13768
Syria	23
São Tomé and Príncipe	11
Taiwan	1579
Tajikistan	17
Tanzania	34
Thailand	819
Trinidad and Tobago	25
Tunisia	447

Turkey	5714
Turkmenistan	9
Turks & Caicos Islands	7
UK Virgin Islands	5
US Virgin Islands	6
Uganda	10
Ukraine	4679
United Arab Emirates	3272
United Kingdom	34537
United States	23914
United States Minor Outlying Islands	29
Uruguay	1675
Uzbekistan	85
Vatican City	10
Venezuela	586
Vietnam	216
Zambia	13
Zimbabwe	16
Other	81
Total	561015

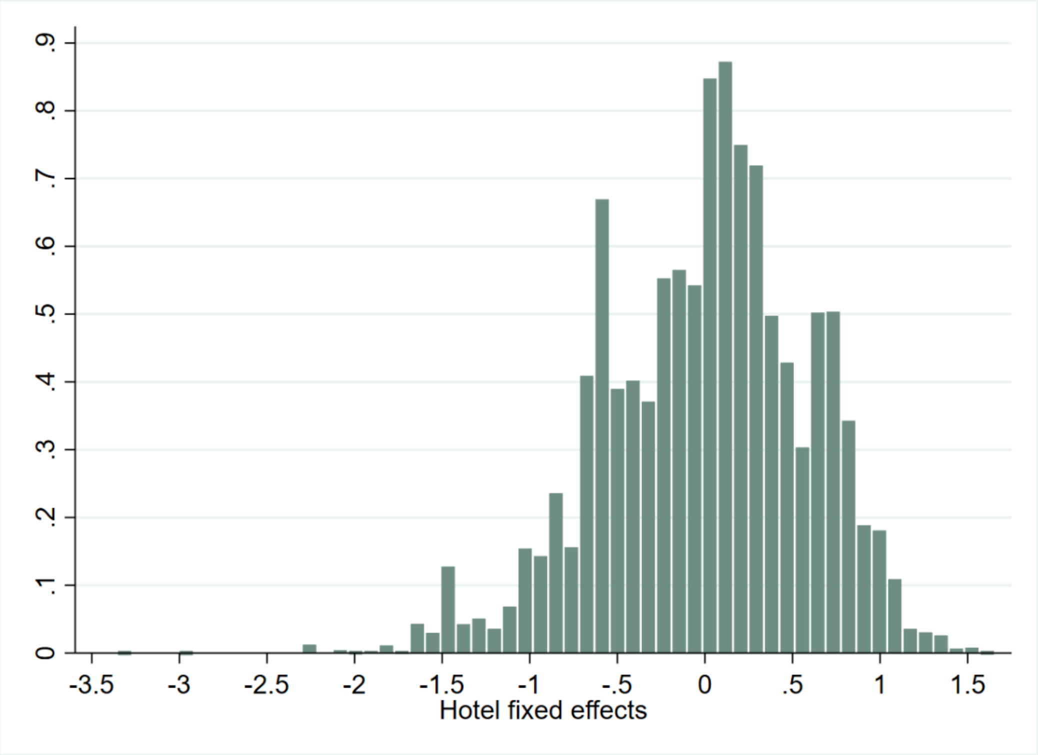


Figure A1. Histogram of hotel fixed effects estimates from Column 4 in Table 2

Table A2. Baseline OLS regression results assuming homoscedastic variance

	(1)	(2)	(3)	(4)
Dep. Variable: score	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Num reviews (t-1)	-3.6e-06			-8.4e-08
	(1.2e-05)			(1.3e-05)
Av. Score (t-1)		-0.168***		-0.113***
		(0.031)		(0.041)
SD Score (t-1)			0.162***	0.086**
			(0.027)	(0.037)
LOS	-0.006**	-0.006**	-0.006**	-0.006**
	(0.003)	(0.003)	(0.003)	(0.003)
Travel party: couple	-0.004	-0.004	-0.004	-0.004
	(0.014)	(0.014)	(0.014)	(0.014)
Travel party: family	-0.006	-0.006	-0.006	-0.006
	(0.015)	(0.015)	(0.015)	(0.015)
Travel party: group	0.069***	0.069***	0.069***	0.069***
	(0.013)	(0.013)	(0.013)	(0.013)
Domestic	0.035***	0.034***	0.034***	0.034***
	(0.012)	(0.012)	(0.012)	(0.012)
Room type: economy	0.019	0.019	0.020	0.019
	(0.027)	(0.028)	(0.028)	(0.028)
Room type: standard	-0.012	-0.012	-0.012	-0.012
	(0.013)	(0.013)	(0.013)	(0.013)
Room type: superior	-0.016	-0.016	-0.016	-0.016
	(0.016)	(0.016)	(0.016)	(0.016)
Room type: anonymous	-0.199***	-0.199***	-0.199***	-0.199***
	(0.011)	(0.011)	(0.011)	(0.011)
Temporal contiguity	0.058***	0.058***	0.058***	0.058***
	(0.008)	(0.008)	(0.008)	(0.008)
Weekend	0.001	0.001	0.001	0.001
	(0.006)	(0.006)	(0.006)	(0.006)
Constant	7.189***	8.586***	7.016***	8.033***
	(0.034)	(0.258)	(0.043)	(0.379)
Hotel fixed effects	YES	YES	YES	YES
Country of origin fixed effects	YES	YES	YES	YES
Monthly fixed effects	YES	YES	YES	YES
Observations	525,437	525,437	525,437	525,437
R-squared	0.138	0.138	0.138	0.138

Clustered standard errors at the hotel level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A3. Predicted standard deviation of random component vs LOS in the sample

LOS (days)	Predicted standard deviation of random component
1	1.365
2	1.351
3	1.337
4	1.323
5	1.309
6	1.295
7	1.282
8	1.268
9	1.255
10	1.242
11	1.229
12	1.216
13	1.204
14	1.191
15	1.179

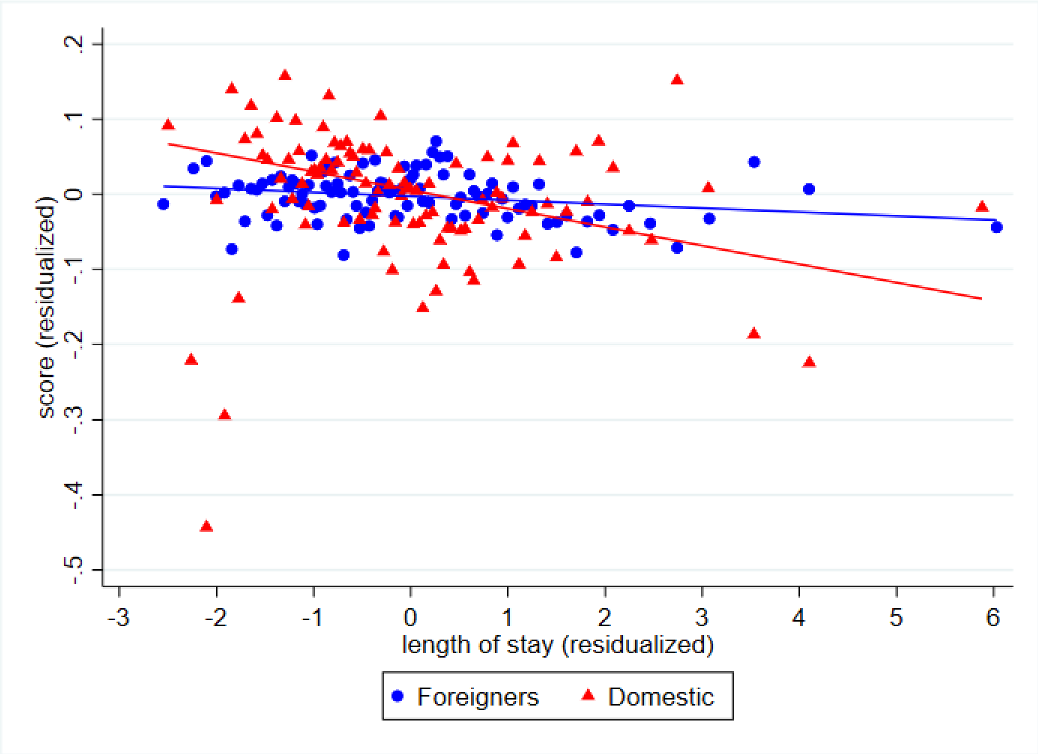


Figure A2. Binned scatterplot of the residualized relationship between score and length of stay (conditional mean) by origin of the guest

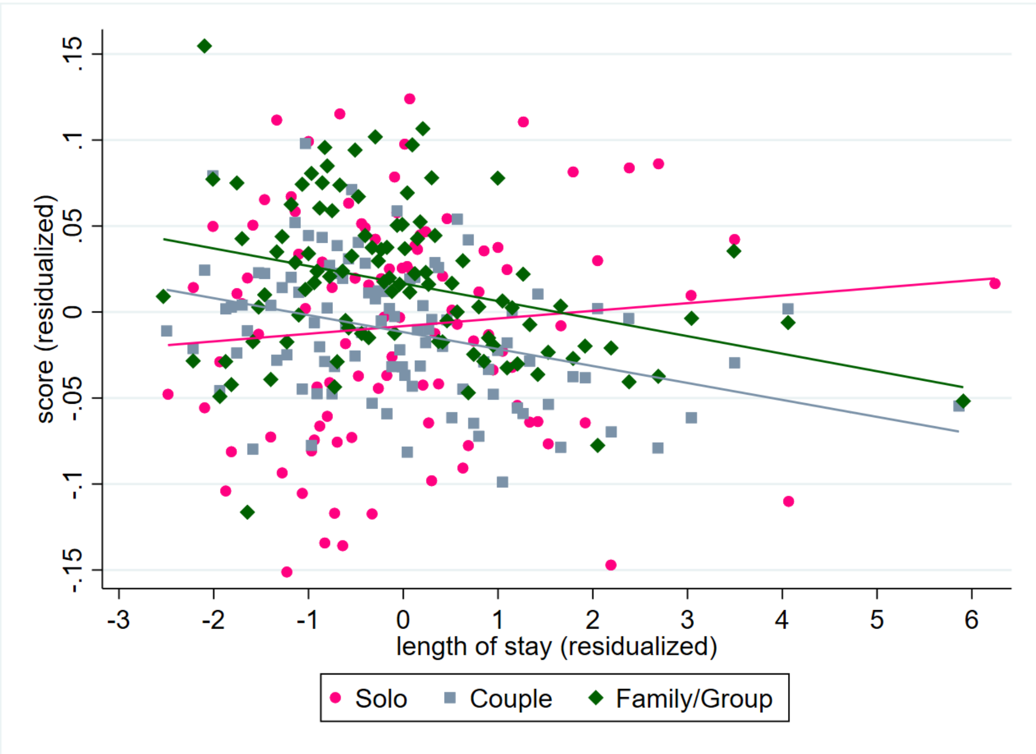


Figure A2. Binned scatterplot of the residualized relationship between score and length of stay (conditional mean) by travel party composition

Table A4. OLS regression of length of stay on explanatory variables

Dep. Variable: LOS	Coef. (SE)
Num reviews (t-1)	5.0e-05***
	(4.5e-06)
Av. Score (t-1)	-0.045**
	(0.022)
SD Score (t-1)	0.004
	(0.022)
Travel party: couple	0.245***
	(0.006)
Travel party: family	0.235***
	(0.007)
Travel party: group	0.086***
	(0.007)
Domestic	-0.978***
	(0.008)
Room type: economy	-0.018
	(0.011)
Room type: standard	-0.030***
	(0.007)
Room type: superior	0.097***
	(0.009)
Room type: anonymous	0.032***
	(0.008)
Temporal contiguity	0.059***
	(0.006)
Weekend	0.020***
	(0.004)
Constant	2.938***
	(0.209)
Hotel fixed effects	YES
Country of origin fixed effects	YES
Monthly fixed effects	YES
Observations	525,422

Clustered standard errors at the hotel level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A5. Coefficient estimates from OLS regression of mean LOS at the hotel level on hotel FE estimates from Table 2

Dep. Variable: mean LOS	Coef. (SE)
$Hotel\ FE$	0.016
	(0.027)
Constant	2.658***
	(0.015)
Observations	1,233

Bootstrapped standard errors after 1,000 replications in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A6. Coefficient estimates from heteroskedastic linear regression using data for the period September 2019-February 2020 (Booking.com new rating system)

	(1)	(2)	(3)	(4)
Dep. Variable: score	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Num.reviews (t-1)	-2.4e-06			7.2e-06
	(3.1e-06)			(3.0e-06)
Av.Score (t-1)		-1.097***		-1.813***
		(0.342)		(0.284)
SDScore (t-1)			0.898***	-0.260
			(0.144)	(0.168)
LOS	-0.013***	-0.013***	-0.013***	-0.013***
	(0.003)	(0.003)	(0.003)	(0.003)
Travel party: couple	0.035***	0.034***	0.034***	0.034***
	(0.013)	(0.013)	(0.013)	(0.013)
Travel party: family	0.027*	0.026*	0.026*	0.025*
	(0.015)	(0.015)	(0.015)	(0.015)
Travel party: group	0.068***	0.067***	0.068***	0.067***
	(0.015)	(0.015)	(0.015)	(0.015)
Domestic	-0.023	-0.024	-0.024	-0.025
	(0.016)	(0.016)	(0.016)	(0.016)
Room type: economy	0.005	0.004	0.005	0.004
	(0.023)	(0.023)	(0.023)	(0.023)
Room type: standard	-0.031**	-0.031**	-0.031**	-0.030**
	(0.014)	(0.014)	(0.014)	(0.014)
Room type: superior	-0.032	-0.032	-0.032	-0.032
	(0.023)	(0.022)	(0.023)	(0.023)
Room type: anonymous	-0.193***	-0.193***	-0.193***	-0.194***
	(0.015)	(0.015)	(0.015)	(0.015)
Temp.Contiguity	0.067***	0.067***	0.067***	0.067***
	(0.009)	(0.009)	(0.009)	(0.009)
Weekend	0.007	0.007	0.007	0.007
	(0.007)	(0.007)	(0.007)	(0.007)
Constant	8.504***	17.549***	6.860***	23.932***
	(0.034)	(2.821)	(0.268)	(2.581)
Hotel fixed effects	YES	YES	YES	YES
Country of origin fixed effects	YES	YES	YES	YES
Monthly fixed effects	YES	YES	YES	YES
Variance equation				
LOS	-0.025***	-0.025***	-0.025***	-0.025***
	(0.005)	(0.005)	(0.005)	(0.005)
Constant	0.794***	0.794***	0.794***	0.794***
	(0.022)	(0.022)	(0.022)	(0.022)
Observations	335,470	334,766	334,640	334,640

Clustered standard errors at the hotel level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A7. Coefficient estimates from heteroskedastic linear regression with sequential addition of covariates

	(1)	(2)	(3)	(4)	(5)
Dep. Variable: score	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Num.reviews (t-1)		-8.8e-06	-1.0e-05*	-1.3e-05**	-2.4e-06
		(6.6e-06)	(6.0e-06)	(5.3e-06)	(1.2e-05)
Av.Score (t-1)		0.997***	0.992***	1.001***	-0.112***
		(0.019)	(0.019)	(0.019)	(0.041)
SDScore (t-1)		0.179***	0.177***	0.192***	0.087**
		(0.031)	(0.031)	(0.030)	(0.037)
LOS	0.006	-0.009***	-0.010***	-0.009***	-0.006**
	(0.005)	(0.003)	(0.003)	(0.003)	(0.002)
Travel party: couple			0.005	-0.005	-0.005
			(0.012)	(0.012)	(0.014)
Travel party: family			-0.001	-0.014	-0.007
			(0.014)	(0.014)	(0.015)
Travel party: group			0.050***	0.059***	0.068***
			(0.012)	(0.012)	(0.013)
Domestic			-0.019**	0.052***	0.034***
			(0.010)	(0.012)	(0.012)
Room type: economy			0.029	0.018	0.018
			(0.023)	(0.023)	(0.028)
Room type: standard			0.001	-0.002	-0.012
			(0.009)	(0.009)	(0.013)
Room type: superior			0.011	0.016	-0.016
			(0.013)	(0.012)	(0.016)
Room type: anonymous			-0.200***	-0.192***	-0.198***
			(0.010)	(0.010)	(0.011)
Temp.Contiguity			0.042***	0.060***	0.058***
			(0.008)	(0.008)	(0.008)
Weekend			-0.001	0.003	0.002
			(0.006)	(0.006)	(0.006)
Constant	8.631***	-0.195	-0.172	-0.384*	8.076***
	(0.026)	(0.198)	(0.204)	(0.201)	(0.381)
Hotel fixed effects	NO	NO	NO	NO	YES
Country of origin fixed effects	NO	NO	NO	YES	YES
Monthly fixed effects	NO	NO	NO	YES	YES
Variance equation					
LOS	-0.020***	-0.019***	-0.019***	-0.020***	-0.021***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.790***	0.673***	0.671***	0.659***	0.644***
	(0.023)	(0.020)	(0.020)	(0.020)	(0.020)
Observations	526,130	526,130	526,130	525,437	525,437

Clustered standard errors at the hotel level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A8. OLS regression results considering two-period lags of stock of reviews, average score and standard deviation of score

	(1)	(2)	(3)	(4)
Dep. Variable: score	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Num reviews (t-2)	-7.9e-06			-6.1e-07
	(1.2e-05)			(1.2e-05)
Av. Score (t-2)		-0.220***		-0.181***
		(0.032)		(0.043)
SD Score (t-2)			0.178***	0.059
			(0.032)	(0.041)
LOS	-0.006**	-0.006**	-0.006**	-0.006**
	(0.003)	(0.003)	(0.003)	(0.003)
Travel party: couple	-0.005	-0.006	-0.006	-0.006
	(0.014)	(0.015)	(0.015)	(0.015)
Travel party: family	-0.007	-0.008	-0.008	-0.008
	(0.015)	(0.016)	(0.016)	(0.016)
Travel party: group	0.068***	0.069***	0.069***	0.069***
	(0.013)	(0.014)	(0.014)	(0.014)
Domestic	0.034***	0.032**	0.032**	0.032**
	(0.012)	(0.013)	(0.013)	(0.013)
Room type: economy	0.019	0.022	0.022	0.022
	(0.028)	(0.027)	(0.027)	(0.027)
Room type: standard	-0.012	-0.012	-0.012	-0.012
	(0.013)	(0.014)	(0.014)	(0.014)
Room type: superior	-0.016	-0.019	-0.019	-0.019
	(0.016)	(0.017)	(0.017)	(0.017)
Room type: anonymous	-0.198***	-0.201***	-0.201***	-0.201***
	(0.011)	(0.012)	(0.012)	(0.012)
Temporal contiguity	0.058***	0.052***	0.052***	0.052***
	(0.008)	(0.009)	(0.009)	(0.009)
Weekend	0.002	0.001	0.001	0.001
	(0.006)	(0.006)	(0.006)	(0.006)
Constant	7.237***	9.129***	7.191***	8.766***
	(0.035)	(0.281)	(0.037)	(0.382)
Hotel fixed effects	YES	YES	YES	YES
Country of origin fixed effects	YES	YES	YES	YES
Monthly fixed effects	YES	YES	YES	YES
Variance equation				
LOS	-0.021***	-0.021***	-0.021***	-0.021***
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	0.644***	0.653***	0.654***	0.653***
	(0.020)	(0.021)	(0.021)	(0.021)
Observations	525,437	468,109	467,692	467,692

Clustered standard errors at the hotel level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A9. Heteroskedastic regression results per city

	(1)	(2)	(3)	(4)	(5)
City	Barcelona	Madrid	Milan	Rome	Lisbon
Dep. Variable: score	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Num reviews (t-1)	4.0e-05***	-5.4e-06	-1.4e-04***	-1.3e-05	2.0e-05
	(1.1e-05)	(4.4e-05)	(4.6e-05)	(3.9e-05)	(1.3e-05)
Av. Score (t-1)	-0.213***	0.022	-0.482***	-0.053	-0.090
	(0.081)	(0.126)	(0.101)	(0.056)	(0.086)
SD Score (t-1)	0.037	0.129	-0.027	0.133***	0.214**
	(0.079)	(0.179)	(0.109)	(0.047)	(0.092)
LOS	-0.004	-0.010	-0.020**	-0.006	-0.003
	(0.005)	(0.007)	(0.008)	(0.004)	(0.004)
Travel party: couple	0.024	0.037	-0.002	-0.029*	-0.057***
	(0.042)	(0.023)	(0.030)	(0.017)	(0.016)
Travel party: family	-0.009	0.072**	-0.006	-0.026	-0.043**
	(0.046)	(0.031)	(0.031)	(0.019)	(0.018)
Travel party: group	0.114***	0.094***	0.048	0.003	0.035**
	(0.039)	(0.025)	(0.035)	(0.018)	(0.018)
Domestic	0.049	-0.120**	0.157**	0.103**	0.081*
	(0.053)	(0.060)	(0.065)	(0.048)	(0.046)
Room type: economy	0.010	0.000	0.094	0.011	0.084*
	(0.054)	(0.041)	(0.123)	(0.036)	(0.046)
Room type: standard	-0.000	-0.021	-0.016	-0.002	-0.027
	(0.035)	(0.022)	(0.035)	(0.022)	(0.023)
Room type: superior	-0.000	0.012	0.014	-0.033	-0.036
	(0.053)	(0.039)	(0.039)	(0.024)	(0.024)
Room type: anonymous	-0.204***	-0.170***	-0.125***	-0.251***	-0.189***
	(0.023)	(0.023)	(0.029)	(0.021)	(0.021)
Temporal contiguity	0.032**	0.143***	0.004	0.077***	0.056***
	(0.014)	(0.016)	(0.025)	(0.018)	(0.014)
Weekend	-0.005	0.013	0.032	0.003	-0.015*
	(0.012)	(0.009)	(0.024)	(0.011)	(0.009)
Constant	9.963***	6.942***	11.948***	7.539***	10.231***
	(0.786)	(1.131)	(0.939)	(0.511)	(0.881)
Hotel fixed effects	YES	YES	YES	YES	YES
Country of origin fixed effects	YES	YES	YES	YES	YES
Monthly fixed effects	YES	YES	YES	YES	YES
Variance equation					
LOS	-0.027***	-0.019*	0.014	-0.020***	-0.031***
	(0.007)	(0.011)	(0.010)	(0.006)	(0.007)
Constant	0.717***	0.539***	0.641***	0.654***	0.572***
	(0.048)	(0.047)	(0.044)	(0.031)	(0.037)
Observations	157,737	64,616	64,607	121,493	116,984

Clustered standard errors at the hotel level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A10. Heteroskedastic regression results from GLS estimator

	(1)	(2)	(3)	(4)
Dep. Variable: score	Coef. (SE)	Coef. (SE)	Coef. (SE)	Coef. (SE)
Num reviews (t-1)	-2.8e-06			-2.2e-06
	(4.3e-06)			(4.3e-06)
Av. Score (t-1)		-0.168***		-0.112***
		(0.016)		(0.021)
SD Score (t-1)			0.164***	0.088***
			(0.016)	(0.021)
LOS	-0.006***	-0.006***	-0.006***	-0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Travel party: couple	-0.005	-0.005	-0.005	-0.005
	(0.006)	(0.006)	(0.006)	(0.006)
Travel party: family	-0.007	-0.007	-0.007	-0.007
	(0.006)	(0.006)	(0.006)	(0.006)
Travel party: group	0.068***	0.068***	0.068***	0.068***
	(0.006)	(0.006)	(0.006)	(0.006)
Domestic	0.034***	0.033***	0.033***	0.033***
	(0.008)	(0.008)	(0.008)	(0.008)
Room type: economy	0.018	0.018	0.018*	0.018
	(0.011)	(0.011)	(0.011)	(0.011)
Room type: standard	-0.012*	-0.012*	-0.012*	-0.012*
	(0.006)	(0.006)	(0.006)	(0.006)
Room type: superior	-0.016*	-0.016*	-0.016*	-0.016*
	(0.009)	(0.009)	(0.009)	(0.009)
Room type: anonymous	-0.198***	-0.198***	-0.198***	-0.198***
	(0.007)	(0.007)	(0.007)	(0.007)
Temporal contiguity	0.058***	0.058***	0.058***	0.058***
	(0.006)	(0.006)	(0.006)	(0.006)
Weekend	0.002	0.002	0.002	0.002
	(0.004)	(0.004)	(0.004)	(0.004)
Constant	7.268***	8.664***	7.090***	8.097***
	(0.945)	(0.954)	(0.945)	(0.963)
Hotel fixed effects	YES	YES	YES	YES
Country of origin fixed effects	YES	YES	YES	YES
Monthly fixed effects	YES	YES	YES	YES
Variance equation				
LOS	-0.032***	-0.031***	-0.032***	-0.031***
	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.610***	0.609***	0.611***	0.608***
	(0.006)	(0.006)	(0.006)	(0.006)
Observations	525,437	525,437	525,437	525,437

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it