How Gig Economy Suppliers Learn from Customer Reviews:

Evidence from Text Analysis Using Large Language Models

Mariko Sakakibara and Yiyang Zeng

August 2026

Abstract

This study examines how suppliers in the gig economy learn from customer text reviews. Online text reviews convey consumer sentiment and highlight salient attributes that influence purchasing decisions, providing suppliers with valuable learning signals. Suppliers can improve performance by addressing concerns raised in reviews, but learning effectiveness varies across suppliers. Those with prior experience can respond to feedback more effectively than de novo entrants. Moreover, learning does not always translate into performance gains when feedback targets attributes that are difficult to change. Using Airbnb listings data and large language models to analyze approximately 1.9 million text reviews, we find evidence broadly consistent with these arguments. Overall, the study demonstrates how customer text reviews enable entrepreneurial learning and shape supplier performance in platform-based markets.

Keywords: entrepreneurial learning, gig economy, text review, large language models, sentiment analysis

Sakakibara: UCLA Anderson School of Management, 110 Westwood Plaza, Los Angeles, CA 90095. mariko.sakakibara@anderson.ucla.edu
Zeng: Scheller College of Business, Georgia Institute of Technology, 800 W Peachtree St NW, Atlanta, GA 30308. yiyang.zeng@scheller.gatech.edu

We would like to thank the participants and reviewers at the annual meetings of the Strategic Management Society and Strategy Science for helpful suggestions. We gratefully acknowledge the financial support from the Academic Senate of the University of California, Los Angeles.

1. INTRODUCTION

Learning is a critical process that enables firms to acquire, accumulate, and integrate knowledge in ways that develop and sustain competitive advantage (Teece, Pisano, & Shuen, 1997). For entrepreneurs, learning is especially important because early feedback informs strategic adaptation (Miller, O'Mahony & Cohen, 2024). Entrepreneurs enter markets uncertain about their ventures' prospects. They update their expectations based on realized performance and remain in the market only when the resulting outlook warrants continued operation (Jovanovic, 1982). Subsequent ventures should therefore exhibit longer survival and stronger performance, even after time-invariant differences in founder ability are held fixed, and evidence on serial entrepreneurship is consistent with this implication (Gompers, Kovner, Lerner, & Scharfstein, 2010; Lafontaine & Shaw, 2016). Research accordingly conceptualizes entrepreneurship as experimentation (Kerr, Nanda, & Rhodes-Kropf, 2014; Manso, 2016). Founders who treat market feedback as tests of explicit hypotheses achieve stronger performance (Camuffo, Cordova, Gambardella, & Spina, 2020). Yet far less is known about the specific information from which entrepreneurs learn after entry. The gig economy has lowered the cost of new venture formation (Barrios, Hochberg, & Yi, 2022) and expanded the population of first-time micro-entrepreneurs who receive customer reviews. These reviews provide entrepreneurs with granular information about the attributes customers reward and those they criticize.

Customers are an important channel for firms to learn from. In the context of innovation, von Hippel (1986) characterizes customers as a source of innovation, while Bower & Christensen (1996) warn of the danger of following current customers too closely. In marketing, Jaworski & Kohli (1993) demonstrate that learning from consumers enhances business

performance and employee outcomes, and Day (1994) conceptualizes learning from customers as a distinct organizational capability central to building a sustainable competitive advantage.

Customer feedback has become increasingly abundant and accessible with the rise of digital platforms that host online reviews. These reviews play a central role in customer purchase decisions by allowing customers to share evaluations (Chen & Xie, 2008). When quality is uncertain, reviews enable social learning by helping customers update their beliefs on quality based on others' experiences (Anderson & Magruder, 2012; Mayzlin, Dover, & Chevalier, 2014). Available on platforms such as Amazon, Yelp, and TripAdvisor, customer reviews facilitate low-cost, large-scale peer learning and complement firms' advertising efforts (Dellarocas, 2003).

Online reviews are particularly important for suppliers in the gig economy, where participants share goods and services via online peer-to-peer platforms on a part-time or temporary basis, because these suppliers are micro-entrepreneurs and often lack the resources to advertise their credibility due to the temporary and uncertain nature of supply (Garud, Kumaraswamy, Roberts, & Xu, 2022). As a result, both sides of the platform (e.g., hosts and guests for Airbnb, drivers and passengers for Uber) actively seek high ratings to signal reputation and facilitate transactions (Proserpio, Xu, & Zervas, 2018). Platforms encourage these efforts by promoting review systems designed to build trust and enhance platform value (Tadelis, 2016).

As online reviews have grown in importance, research in business and economics has largely focused on numerical ratings, documenting their effects on business performance (Luca, 2016; Chevalier & Mayzlin, 2006; Zhu & Zhang, 2010). Far less is known about the effects of text reviews, despite their richer and more nuanced informational content. Studying reviews is further complicated by the prevalence of fake reviews, which can distort reputational signals (He, Hollenbeck, & Proserpio, 2022) and hinder inference. An important exception is the gig

economy, where review systems—such as Airbnb's—restrict review submission to verified, paying users, resulting in substantially higher authenticity.

In the gig economy, individuals leverage their underutilized time and resources to pursue entrepreneurial opportunities, but heterogeneity in prior experience and capabilities leads to substantial variation in how effectively these resources are deployed and, consequently, in performance. Because effective resource use requires ongoing skill development, new participants face strong incentives to learn over time. Although research has documented experience-based performance gains, most studies rely on coarse productivity measures and give limited attention to individual performance. Notable exceptions include studies of learning among Uber drivers (Cook, Diamond, Hall, List, & Oyer, 2021) and New York City taxi drivers (Haggag, McManus, & Paci, 2017).

In this paper, we examine how suppliers in the gig economy learn from customers through text reviews. Specifically, we identify the attributes of services that customers emphasize, assess whether suppliers treat these attributes as learning signals and improve along them to achieve greater performance, and which types of suppliers learn most effectively. We also investigate when and why learning fails. We argue that customer sentiment embedded in review text highlights salient attributes that shape demand and provide firms with learning signals. However, suppliers differ in learning capacity: those with prior experience respond more effectively to customer feedback, achieving superior performance. Furthermore, learning may not translate into performance gains if the feedback concerns attributes that are difficult to change or beyond the supplier's control.

We focus on the gig economy, where reviews play a central role and are likely to be highly authentic. Our empirical setting is Airbnb in Los Angeles from May 2015 to early March

2020, prior to COVID-19–related disruptions, with a focus on new listings to capture learning dynamics. To analyze approximately 1.9 million text reviews, we employ large language models (LLMs) to extract attribute-level information and sentiment from review text and construct panel data for regression analysis. While prior research has used machine learning and topic modeling to analyze text (e.g., Miric, Jia, & Huang, 2022), LLMs enable context-sensitive text analysis at scale with minimal manual preprocessing (Carlson & Burbano, 2025). We find that suppliers achieve performance gains by improving customer sentiment on attributes, and experienced suppliers respond to customer feedback more effectively. We also find that although suppliers can improve on all attributes after receiving negative sentiments, improvements in some attributes do not necessarily translate to realized performance gains.

This study makes two primary contributions. First, we contribute to research on entrepreneurial learning by examining heterogeneity among new entrants in their ability to learn and translate feedback into performance. Prior knowledge shapes the opportunities founders identify (Shane, 2000), and pre-entry knowledge and post-entry learning jointly affect new-firm performance (Dencker, Gruber, & Shah, 2009). Experience also transfers more readily when acquired in a related setting (Eggers & Song, 2015). We extend this literature by studying learning from attribute-level signals in review text, rather than inferring learning from coarse measures of survival or earnings. Using rich data from Airbnb, a platform characterized by highly authentic reviews, we show how LLMs can be used to extract meaningful signals from customer feedback and identify which types of micro-entrepreneurs learn most effectively from text reviews. We also demonstrate that experienced entrepreneurs capture larger performance gains from the same feedback and improve along criticized attributes rather than raising quality in general.

Second, we add to the growing literature on the gig economy by showing how individual participants become more effective at deploying underutilized resources after entering a platform market. Much of this literature emphasizes how platforms lower entry barriers and enable individuals to monetize idle assets (Einav, Farronato, & Levin, 2016; Filippas, Horton, & Zeckhauser, 2020); we instead illustrate the post-entry process through which participants refine how those assets are offered to customers. By accounting for participants' prior experience while controlling for competition, listing characteristics, and individual attributes, our study provides novel evidence on the development of micro-entrepreneurial capabilities within the gig economy.

We organize the rest of this paper as follows. In section 2, we review relevant literature and develop testable hypotheses. Section 3 introduces the data and the overarching empirical approach. Section 4 explains the details of our empirical analyses and presents our findings. Section 5 discusses the insights and limitations and concludes.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

2.1 Customer Reviews and Firm Performance

Advances in digital technologies have made customer reviews increasingly influential in shaping consumer purchasing decisions (Luca, 2016; Reimers & Waldfogel, 2021) and firms' strategic choices regarding product offerings (Chen & Xie, 2005, 2008; Zhu & Zhang, 2010). Prior research emphasizes that reviews affect business performance primarily by signaling product quality and facilitating social learning, allowing consumers to update beliefs based on peer feedback (Chevalier & Mayzlin, 2006). Reviews also serve as a feedback mechanism for firms by revealing customer evaluations that can inform quality improvement (Dellarocas, 2003).

Customer reviews typically take two forms: numerical ratings and text reviews. Research in business and economics documents the impact of numerical ratings on firm reputation and

performance across a range of markets, including restaurants (Luca, 2016; Greenberg, Sands, Cattani, & Porac, 2024), books (Chevalier & Mayzlin, 2006; Reimers & Waldfogel, 2021), games (Zhu & Zhang, 2010), and healthcare services (Gao, Greenwood, McCullough, & Agarwal, 2012; Lu & Rui, 2018). Across these settings, studies consistently show that numerical ratings shape consumer demand and, consequently, firm outcomes.

In contrast, the role of text reviews remains less well understood. Although text reviews contain richer and more nuanced information than numerical scores (Ganu, Elhadad, & Marian, 2009), their unstructured format makes them more difficult to analyze than conventional archival datasets. Early studies rely on coarse proxies, such as review length, to infer the informational value of text and show that consumers review text rather than relying solely on summary statistics (Chevalier & Mayzlin, 2006). More recent work leverages machine learning techniques to analyze review text and examine consumer choice behavior on platforms such as Yelp (Farronato & Zervas, 2022; Chakraborty, Kim, & Sudhir, 2022). Related studies analyze employer text reviews to infer worker preferences (Tambe, Ye, & Cappelli, 2020) and healthcare text reviews to assess patient choices and physician quality (Xu, Armony, & Ghose, 2021). Collectively, this literature suggests that text reviews convey meaningful information and represent a potentially important source of firm learning. Yet significant challenges remain in extracting and operationalizing the rich, unstructured information embedded in text reviews.

A persistent challenge in studying reviews, however, is the prevalence of fake reviews, which can distort reputational signals and undermine empirical inference. Because reviews strongly influence firm outcomes, firms may engage in both positive review fraud for their own offerings and negative review fraud targeting competitors (Mayzlin, Dover, & Chevalier, 2014; Luca & Zervas, 2016). Empirical evidence of review manipulation has been documented on

platforms such as Amazon (He, Hollenbeck, & Proserpio, 2022) and Yelp (Luca & Zervas, 2016). Although various machine-learning approaches have been developed to detect fake reviews (Wu et al., 2020), their effectiveness remains limited (Heydari et al., 2015; Vidanagama, Silva, & Karunananda, 2020). Consequently, access to authentic review data is critical for credible analysis of the informational and performance effects of customer reviews.

2.2 Gig Economy and the Role of Reviews

Gig economy platforms facilitate peer-to-peer exchange among large numbers of fragmented buyers and sellers (Einav, Farronato, & Levin, 2016). Because suppliers of goods and services are often temporary and heterogeneous, these platforms lack the inherent credibility of established incumbents (Garud et al., 2022). To address this credibility gap, platforms embed review and feedback systems as a trust-building mechanism (Xu, Lu, Ghose, Dai, & Zhou, 2023). Review submission is typically restricted to customers who have completed paid transactions, which enhances the authenticity of reviews (Bucher, Fieseler, Fleck, & Lutz, 2018).

A second defining feature of the gig economy is that it enables individuals to act as micro-entrepreneurs by monetizing underutilized assets and time. Effectively deploying these resources requires ongoing learning and skill development. Given the centrality of reviews in this setting, customer feedback provides micro-entrepreneurs with a salient and credible source of information for evaluating their practices and identifying opportunities for improvement.

2.3 Customer Sentiment and Attention

Consumer sentiment—customers' satisfaction with a firm's offerings—has been shown to influence purchasing intentions and actual spending (Carroll, Fuhrer, & Wilcox, 1994; Gillitzer & Prasad, 2018). Early research in marketing and consumer behavior developed scale-based measures to assess how sentiment shapes consumer behavior (Barksdale & Darden, 1972;

Lundstrom & Lamont, 1976; Gaski & Etzel, 1986), consistently finding that positive sentiment encourages spending while negative sentiment discourages it (Bougie, Pieters, & Zeelenberg, 2003). As consumer-generated digital text has proliferated, extracting sentiment from text has become a managerial priority (Villarroel Ordenes, Ludwig, De Ruyter, Grewal, & Wetzels, 2017; Humphreys & Wang, 2018). Recent evidence further shows that sentiment expressed in text reviews can outperform traditional measures in predicting and explaining consumer decisions (Berger, Humphreys, Ludwig, Moe, Netzer, & Schweidel, 2020).

Importantly, customers do not write text reviews at random. Instead, they focus on salient attributes, problems, and usage contexts that are relevant to purchasing decisions and that help them gain peer recognition (Jarvenpaa & Leidner, 1999). For example, Archak, Ghose, & Ipeirotis (2011) decompose Amazon review text into feature-level components and show that consumers systematically discuss attributes that drive purchase choices, enabling estimation of the pricing power of individual features. Similarly, Ghose & Ipeirotis (2011) demonstrate that review text contains concentrated information about the aspects consumers care about most.

Taken together, the attributes emphasized in customer reviews and the sentiments expressed toward those attributes constitute a valuable source of information through which suppliers can learn about customer preferences that directly shape demand. This information can guide suppliers' attention toward areas where improvement is most consequential, enhancing alignment with customer expectations and improving performance. Accordingly, we hypothesize:

Hypothesis 1. Suppliers that improve on attributes mentioned in customer text reviews, as reflected in improvements in customer sentiment, achieve greater performance gains.

2.4 Heterogeneity among Suppliers in Learning from Text Reviews

Learning-by-doing refers to performance improvements arising from accumulated experience

through repeated production or practice. Economists and management scholars have long documented such experience premium in both manufacturing (Lieberman, 1984, 1989; Argote, Beckman, & Epple, 1990; Egelman, Epple, Argote, & Fuchs, 2017) and services (Darr, Argote, & Epple, 1995). Prior research has also examined the relationship between learning and firm diversification (Ryu, McCann, & Wan, 2022), human resources (Song, Almeida, & Wu, 2003; Jain, 2016), supply chains (Alcacer & Oxley, 2014), and alliances (Kale & Singh, 2007). Importantly, suppliers and entrepreneurs differ in their learning rates due to variations in prior experience, intrinsic capabilities, and environmental conditions (Balasubramanian & Lieberman, 2010, 2011; Balasubramanian, 2011).

Empirical evidence on such heterogeneity, however, remains limited, largely because fine-grained data on individual-level activities are difficult to obtain. As a result, most studies rely on broad productivity measures, offering limited insight into how individual suppliers improve and through which learning channels. Two recent exceptions come close to addressing this gap. Haggag, McManus, and Paci (2017) show that New York City taxi drivers learn through experience, with performance gaps between novice and experienced drivers widening under challenging conditions. Cook, Diamond, Hall, List, and Oyer (2021) find that cumulative driving experience explains a substantial portion of gender differences in performance among Uber drivers. These studies underscore the importance of experience in shaping learning outcomes.

Building on this insight, we argue that experienced and inexperienced suppliers differ in their ability to act on customer feedback. Experienced suppliers possess tacit, context-specific knowledge that facilitates faster and more effective responses to learning signals (Schneider, 2010). Although inexperienced suppliers can accumulate experience over time, learning-by-doing is often slow. As customer text reviews provide immediate, attribute-specific guidance on

where improvement is needed, suppliers with prior experience are better positioned to translate such feedback into action, whereas de novo entrants are likely to respond less effectively. Thus:

Hypothesis 2. New suppliers with prior experience respond to customer feedback more effectively than de novo entrants, resulting in superior performance gains.

2.5 Impediments for Learning

Even when suppliers learn from customer reviews, translating that learning into performance improvements is not always feasible. In some cases, negative sentiment is tied to attributes that are difficult or impossible to change, limiting suppliers' ability to respond. Additionally, early negative reviews can inflict substantial reputational damage, deterring future customers and rendering subsequent improvements ineffective. Such "killer comments" are particularly harmful because, under quality uncertainty, consumers update beliefs based heavily on early peer feedback (Moretti, 2011). As a result, initial negative reviews require many consistently positive subsequent reviews to be offset, leading to persistent sales penalties (Luca, 2016). Consistent with this logic, prior studies show that early negative numerical ratings have lasting effects on demand (Chevalier & Mayzlin, 2006; Luca, 2016), and even small early rating differences can substantially alter consumer behavior (Anderson & Magruder, 2012). Under these conditions, learning may not readily translate into performance gains. Thus:

Hypothesis 3. When negative sentiment is tied to certain attributes, such as those that are difficult to address, suppliers are unable to realize performance gains even when they respond to customer feedback.

3. DATA AND EMPIRICAL APPROACH

3.1 Data and Methodology

We draw on Airbnb listing data obtained from Inside Airbnb, an organization that scrapes

information from the Airbnb platform on a roughly monthly basis. Our dataset for this analysis includes listings, calendar pricing, and customer reviews for Los Angeles from May 2015 to March 12, 2020. We end the sample prior to the onset of COVID-19–related disruptions to travel, which began with the declaration of a national emergency on March 13, 2020. The data allow us to observe sales performance, listing characteristics, host attributes, geographic clustering, and information from both numerical ratings and textual reviews.

Because nationwide Airbnb data are computationally intensive, we focus on Los Angeles to maintain tractability. Los Angeles represents a large and active marketplace, ranking as the largest city in the United States by listings—exceeding New York and Orlando—making it a particularly informative empirical setting. To study learning dynamics, we restrict the sample to listings that entered the platform during the observation period, yielding 81,249 new listings. Hosts associated with these listings may have operated other listings prior to the focal entry.

Our analysis centers on customer review text, from which we extract attribute-specific information. Prior strategy research has employed machine learning and topic modeling to classify unstructured text (e.g., Miric et al., 2022), but these approaches typically require extensive custom classifier development. Recent advances in LLMs enable context-sensitive classification of unstructured text with substantially lower implementation costs (Carlson & Burbano, 2025). We therefore rely on LLMs to process review text and restrict the sample to English-language reviews to facilitate analysis, resulting in 1,929,340 usable observations.

We construct datasets from customer text reviews in three steps. First, we identify the listing attributes most frequently discussed by Airbnb guests. We apply BERTopic to the full corpus, supplemented by n-gram frequency analysis and cluster analysis, to uncover key themes. Then we validate category coverage and structure subcategories for multi-dimensional attributes.

After reviewing and refining these outputs, we retain 11 main categories and 16 subcategories.

Second, we extract review substrings corresponding to each category and subcategory and conduct sentiment analysis on the extracted text, coding sentiment as positive (1), neutral (0), or negative (–1). We implement this step using a series of few-shot prompts embedded in a Python program and executed with GPT-4o-mini due to Microsoft Azure batch-processing constraints. An important advantage of LLMs is their robustness to linguistic variation; for example, references to a host may appear under different names or titles, such as "host", "the host", "Charles", "Chas", or "Rev. Dr. Louise-Diana", yet appropriately designed prompts allow the model to consistently identify and analyze these mentions. We then validate the sentiment outputs. Table 1 reports the categories, subcategories, and their frequencies. Overall evaluation is the most frequently mentioned category (96.0%), followed by property (65.2%) and location (57.5%). Overall, reviews are positive, with negative sentiment appearing in 2.0%–97.7% of attribute mentions. Mentions of certain attributes are predominantly negative in sentiment, including cancellations (97.7%) and noise (49.1%).

Third, we aggregate the review-level data to the listing–3-month period level for empirical analysis. A key consideration in constructing sentiment measures is how to weight reviews over time. Because our focus is learning—specifically, how hosts respond to and improve upon recent feedback—we construct a 3-month moving average measure for each attribute. This approach focuses on recent reviews, better capturing the feedback signals that hosts are most likely to attend to and act upon. For each listing-month, the sentiment score represents the average sentiment from reviews received in the current month and the preceding two months. In our regression analyses, we restrict the sample to listing-months where a valid 3-month moving average can be calculated. This restriction ensures that our sentiment measures

reflect actual guest feedback rather than imputed values.

[Insert Table 1 here]

Because Airbnb data are scraped at an approximately monthly frequency, we use 30-day revenue as our primary performance measure. We construct this measure using the calendar data, which report listing availability and prices up to 30 days ahead. Specifically, we compute the weighted average price of available days and multiply it by the number of unavailable days to proxy for revenue from bookings over the subsequent 30-day period. The dataset also includes detailed listing features and host characteristics. Table 2 presents summary statistics for key variables at the listing–month level.

[Insert Table 2 here]

3.2 Empirical Specification

3.2.1 Baseline specification

Our first hypothesis (H1) posits that suppliers that improve on attributes mentioned in customer text reviews achieve greater performance gains. To test this prediction, we estimate the following baseline specification:

$$\ln(Y_{it}) = \beta \times \text{Sentiment}_{i,t-1} + \gamma \times X_{it} + \alpha_i + \tau_t + \varepsilon_{it} \quad (1)$$

where Y_{it} represents performance outcomes for listing i in month t. We examine three outcomes: the number of booked nights, the average nightly price, and total 30-day revenue (the product of bookings and price). All outcomes are measured in natural logarithms to facilitate the interpretation of coefficients as approximate percentage changes. $\text{Sentiment}_{i,t-1}$ is our key explanatory variable—the one-month lagged 3-month moving average of overall evaluation sentiment, scaled from −1 (entirely negative) to +1 (entirely positive). The use of lagged sentiment serves two purposes: it addresses reverse causality concerns (that better-performing

listings might simply attract more favorable reviews rather than learning from feedback) and reflects the realistic timing of host responses to guest feedback.

The vector X_{it} includes time-varying controls, most notably the log of the number of nearby active listings to account for local competitive conditions. In our preferred specification, we include listing fixed effects α_i to absorb all time-invariant heterogeneity across listings—including neighborhood, property and host characteristics, and other unobserved factors that might confound the sentiment-performance relationship. Time fixed effects τ_t capture common shocks affecting all listings in a given month, such as seasonality, macroeconomic conditions, or platform-wide policy changes. Standard errors are clustered at the listing level to account for serial correlation within listings over time.

The coefficient β captures the within-listing effect of sentiment improvement on performance. A positive and significant β indicates that when a listing's sentiment score improves relative to its own historical average, it subsequently achieves improved performance outcomes. We estimate Equation (1) with progressively richer specifications to assess the robustness of this relationship: (i) OLS with no controls or fixed effects, providing a descriptive baseline; (ii) OLS with controls and time and neighborhood fixed effects, accounting for geographic heterogeneity; and (iii) the preferred specification with listing fixed effects that identifies effects from purely within-listing variation over time.

3.2.2 Robustness and causality tests of baseline specification

Although the combination of lagged sentiment and listing fixed effects addresses many identification concerns, we employ several additional tests to probe the robustness of our findings and strengthen causal interpretations.

Persistence of effects. If sentiment improvements reflect genuine learning and quality

enhancement rather than transitory fluctuations or spurious correlation, we would expect their effects to persist over time. We therefore extend Equation (1) to include both one-month and two-month lagged sentiment:

$$\ln(Y_{it}) = \beta_1 \times \text{Sentiment}_{i,t-1} + \beta_2 \times \text{Sentiment}_{i,t-2} + \gamma \times X_{it} + \alpha_i + \tau_t + \varepsilon_{it} \quad (2)$$

Significant effects at both lags would suggest that sentiment improvements have lasting benefits rather than merely transitory effects. The combined effect $\beta_1 + \beta_2$ represents the total performance impact of a sustained sentiment improvement.

First-difference specification. As an alternative approach that differences out all listing-level unobserved heterogeneity, we estimate:

$$\Delta \ln(Y_{it}) = \beta \times \Delta\text{Sentiment}_{i,t-1} + \tau_t + \varepsilon_{it} \quad (3)$$

where Δ denotes the first difference (change from t−1 to t). This specification provides an alternative identification strategy that differences out all listing-level heterogeneity. By focusing on month-to-month changes in the 3-month moving average sentiment rather than levels, this approach tests whether immediate changes in recent sentiment trends predict immediate changes in performance, offering a complementary test that is robust to any time-invariant confounds that might remain after controlling for listing fixed effects.

Additional robustness tests. We conduct several supplementary analyses reported in the appendix. First, we test another alternative specification where two-month lagged sentiment is used instead of one-month lagged sentiment. Second, we examine whether effects are stronger among listings with initially poor sentiment scores, which would be consistent with a learning interpretation where there is more scope for improvement. Third, we examine cumulative improvement measures (current minus initial sentiment) as an alternative operationalization of the learning construct. Finally, we employ propensity score matching with inverse probability

weighting to address potential selection into sentiment improvement, weighting observations by the inverse probability of improvement based on listing characteristics and lagged performance.

3.2.3 Experience heterogeneity

Our second hypothesis (H2) predicts that suppliers with prior experience respond to customer feedback more effectively than de novo entrants. To test this prediction, we estimate Equation (1) separately for different host experience groups and compare the magnitude of sentiment effects.

$$\ln(Y_{it}) = \beta^{k} \times \text{Sentiment}_{i,t-1} + \gamma \times X_{it} + \alpha_i + \tau_t + \varepsilon_{it} \quad [\text{for experience group } k] \qquad (4)$$

We operationalize host experience using two complementary measures that capture different dimensions of platform-specific knowledge and capabilities. The first measure distinguishes between repeat hosts—those who have operated previous Airbnb listings before the focal listing's entry—and first-time hosts who are new to the platform. Repeat hosts have accumulated experience of managing guest expectations, optimizing pricing, and responding to feedback through their prior listings. The second measure distinguishes between multi-listing hosts—those currently operating two or more listings simultaneously—and single-listing hosts. Multi-listing hosts not only have more experience but also may benefit from economies of scale and scope in addressing guest concerns.

H2 predicts that $\beta^{experienced} > \beta^{inexperienced}$, that is, hosts with prior platform experience or larger current portfolios translate sentiment improvements into larger performance gains. This differential responsiveness would suggest that accumulated knowledge and capabilities condition the returns to learning from customer feedback.

3.2.4 Category-specific learning and performance effects

Our third hypothesis (H3) examines whether learning effectiveness varies across attribute categories. Even if hosts can and do respond to feedback across all dimensions of the guest

experience, the ability to translate that response into measurable performance gains may depend on whether the criticized attribute is actionable—that is, within the host's control to meaningfully change. We investigate this heterogeneity through three complementary analyses.

Own-category learning. We first test whether hosts actually improve in categories where they receive negative feedback. For each category c, we estimate:

$$\text{Improvement}_{c,i,t+k} = \beta_c \times \text{NegativeFeedback}_{c,it} + \alpha_i + \tau_t + \varepsilon_{it} \quad (5)$$

where $\text{Improvement}_{c,i,t+k}$ measures the maximum sentiment improvement in category c over the subsequent 1–3 months, and $\text{NegativeFeedback}_{c,it}$ is an indicator for whether the listing experienced a sentiment decline greater than 0.1 points in category c during month t. A positive and significant β_c indicates that hosts improve in category c after receiving negative feedback in that category, consistent with targeted learning.

Cross-category learning spillovers. To assess whether learning is truly category-specific or whether it spills over across dimensions, we estimate Equation (5) separately for all 121 combinations of shock category (where negative feedback occurs) and improvement category (where improvement is measured). We run one regression for each combination, using indicator for negative feedback in the shock category as the independent variable and sentiment improvement in the improvement category as the dependent variable. The coefficients from these 121 regressions are organized into an 11×11 spillover matrix, where each cell represents the estimated effect of negative feedback in the row category on improvement in the column category. If learning is targeted, we would expect strong diagonal elements (own-category learning) and weak off-diagonal elements (limited spillover). Conversely, if negative feedback triggers a general effort to improve, we would observe similar effects across all columns regardless of which category received the shock.

Category improvement effects on performance. Finally, and most critically for H3, we test which category improvements translate into measurable performance gains:

$$\ln(\text{Revenue}_{it}) = \beta_c \times \text{Improvement}_{c,it} + \gamma \times X_{it} + \alpha_i + \tau_t + \varepsilon_{it} \quad (6)$$

for each category c, where $\text{Improvement}_{c,it}$ is the difference between current and initial sentiment in category c, capturing cumulative improvement since listing entry. Categories with larger and significant β_c coefficients represent areas where host improvements yield tangible performance benefits, while categories with small or insignificant coefficients suggest that improvement does not translate into better outcomes—either because guests do not value these improvements or because the improvements themselves are superficial.

H3 predicts heterogeneity in these performance effects: actionable attributes that hosts can genuinely improve (such as cleanliness, property amenities, and host responsiveness) should show larger returns, while attributes that are more difficult to change or more subjective in nature (such as value perceptions, location characteristics, or accuracy of expectations) should show smaller or no returns to improvement.

4. EMPIRICAL RESULTS

4.1 Customer Sentiment Improvements and Performance Gains

We begin by examining whether improvements in customer sentiment predict better performance outcomes, as posited in H1. Table 3 presents results from our baseline specification, estimated with progressively richer controls and fixed effects across three performance measures: bookings, price, and revenue.

[Insert Table 3 here]

The results provide strong and consistent support for H1. Consider first the revenue results in Panel C, which capture the combined effects on both quantity and price. The simple

OLS estimate in Column (1) indicates that a one-point improvement in sentiment—equivalent to moving from entirely negative to neutral reviews, or from neutral to entirely positive reviews—is associated with 52.3% higher revenue. This large coefficient reflects both the true effect of sentiment and the contribution of omitted variables that correlate with both good reviews and good performance. Adding time and neighborhood fixed effects in Column (2) reduces the coefficient to 40.6%, suggesting that some of the cross-sectional relationship is attributable to locational advantages or temporal trends. The preferred listing fixed effects specification in Column (3) yields a coefficient of 0.134, indicating that a one-point within-listing improvement in sentiment is associated with approximately 13.4% higher revenue.

The pattern across specifications is informative. The substantial attenuation from Column (1) to Column (3) indicates that much of the cross-sectional correlation between sentiment and performance reflects time-invariant listing characteristics rather than the causal effect of sentiment improvement. This finding underscores the importance of controlling for unobserved heterogeneity when studying the returns to reputation and learning. At the same time, the persistence of significant effects in the most demanding specification suggests that within-listing sentiment improvements do translate into meaningful performance gains.

Decomposing the revenue effect, the booking results in Panel A show that sentiment improvements primarily operate through increased demand: a one-point improvement increases bookings by 10.0% in the listing fixed effects specification. In contrast, the price effects in Panel B are economically small and statistically insignificant once listing fixed effects are included (coefficient of 0.002), suggesting that hosts do not—or cannot—raise prices in response to improved reviews. This asymmetry is consistent with a competitive market structure where prices are largely determined by market conditions and listing characteristics, while demand

responds to quality signals conveyed through reviews.

To put these magnitudes in perspective, it is useful to consider realistic scenarios. A one-point sentiment improvement represents an overwhelmingly strong shift in guest evaluations. A more plausible improvement—say, from neutral to reasonably positive sentiment, or a 0.5-point increase—would translate to approximately 5% higher bookings and 6.7% higher revenue based on the listing fixed effects estimates. As the sample average of a listing earning is approximately $2,000 in monthly revenue, this represents roughly $135 in additional monthly earnings.

4.1.1 Persistence of effects

Table 4 examines whether sentiment effects persist over time by including both one-month and two-month lagged sentiment. The results confirm that sentiment improvements have lasting benefits. In the revenue specification (Column 6), the one-month lag coefficient is 0.096 and the two-month lag coefficient is 0.082, both statistically significant at the 1% level. The combined effect of 0.178 implies that a sustained one-point sentiment improvement is associated with approximately 17.8% higher revenue—substantially larger than the contemporaneous effect alone. This persistence is consistent with a genuine learning interpretation: hosts that improve their service quality enjoy sustained benefits as their improved reputation compounds over time.

[Insert Table 4 here]

4.1.2 Robustness tests

Appendix Tables A1–A2 present additional robustness checks that reinforce our main findings. The first-difference specification (Table A1), which differences out all listing-level heterogeneity, confirms that changes in sentiment predict changes in performance: a one-point increase in sentiment change is associated with a 10.1% increase in revenue growth. Table A2 shows four alternative specifications. Panel A uses two-month lagged sentiment instead of one-month lag,

confirming that effects persist over longer horizons. Panel B restricts the sample to listings with low initial sentiment (bottom quartile) and shows that effects are stronger among these listings, consistent with a learning interpretation where there is more scope for improvement among lower-quality providers. Note that when we run the same regression for the high initial sentiment subsamples (above the bottom quartile), we still obtain positive and significant coefficients, but the magnitude is smaller. Panel C uses cumulative sentiment improvement over the initial sentiment as an alternative operationalization of the learning construct, which yields consistent results. Panel D employs propensity matching through inverse probability weighting to show that selection into improvement does not drive our findings.

Collectively, the evidence strongly supports H1: suppliers that achieve better customer sentiment subsequently realize meaningful performance gains.

4.2 Supplier Experience and Learning Outcomes

We next examine whether the returns to sentiment improvement vary with host experience, as predicted by H2. If experienced hosts possess tacit knowledge, established processes, or resource advantages that facilitate more effective responses to customer feedback, we would expect them to translate sentiment improvements into larger performance gains.

Table 5 presents split-sample regressions comparing first-time hosts—those operating their first Airbnb listing—to repeat hosts who have operated previous listings on the platform. The results reveal a consistent pattern across all three performance measures: repeat hosts realize larger gains from sentiment improvement than first-time hosts.

[Insert Table 5 here]

Focusing on the revenue results, the coefficient for repeat hosts (Column 6) is 0.135, compared to 0.115 for first-time hosts (Column 5). This difference implies that repeat hosts derive

approximately 17% larger revenue gains from equivalent sentiment improvements ((0.135 − 0.115) / 0.115 = 0.17). While both groups benefit significantly from improved sentiment, the premium for experienced hosts suggests that prior platform experience enhances the ability to capitalize on positive feedback—whether through better service delivery, more effective response to concerns, superior understanding of what guests value, or other channels.

Table 6 presents an alternative experience measure based on current portfolio size: single-listing versus multi-listing hosts operating two or more properties simultaneously. The pattern is strikingly similar. Multi-listing hosts show a revenue coefficient of 0.132 compared to 0.115 for single-listing hosts—again, approximately 15% larger effects for the more experienced group.

[Insert Table 6 here]

The consistency across both experience measures strengthens confidence in the finding. Repeat host status captures accumulated platform-specific human capital from managing previous listings, while multi-listing status captures current scale and potentially ongoing learning across a portfolio of properties. That both measures yield similar patterns suggests a robust relationship between experience and responsiveness to feedback.

We acknowledge that these experience advantages may reflect multiple mechanisms beyond pure learning effects. Multi-listing hosts, for instance, may benefit from economies of scale in addressing guest concerns (e.g., bulk purchasing of amenities, dedicated cleaning staff) or from diversification that allows them to experiment with improvements across properties. Similarly, repeat hosts may have selected into continued platform participation precisely because they possess underlying capabilities that predict both experience accumulation and effective feedback response. While we cannot fully disentangle these mechanisms, the consistent pattern across specifications and measures provides solid support for H2: hosts with more platform

experience more effectively translate sentiment improvements into business performance.

4.3 Heterogeneity in Learning Outcomes Across Attributes

Our final hypothesis examines the boundaries of learning from customer feedback. Even if hosts attend to and respond to criticism across all dimensions of the guest experience, not all improvements may translate equally into performance gains. Some attributes may be difficult to genuinely improve, others may reflect guest expectations rather than objective quality, and still others may concern aspects that guests do not heavily weight in subsequent booking decisions. We investigate these possibilities through three complementary analyses.

4.3.1 Category-specific learning

Table 7 presents results from our own-category learning analysis, testing whether negative feedback in a specific category predicts subsequent improvement in that same category. Equation (5) is run for each category. The results show hosts improve in every single category after receiving negative feedback. All eleven coefficients are positive and statistically significant ($p < 0.01$). The magnitudes indicate that hosts respond substantively to negative feedback. For example, when a listing receives negative overall evaluation (sentiment decline > 0.1 points), it improves by 0.060 points on the sentiment scale over the subsequent 1-3 months.

[Insert Table 7 here]

These results indicate that hosts do attend to and respond to criticism regardless of the category. Whether the complaint concerns location information, host communication, property amenities, or any other dimension, hosts on average show improved sentiment in subsequent reviews. This finding is consistent with a motivated learning interpretation: hosts recognize negative feedback as a problem requiring attention and take actions to address guest concerns.

4.3.2 Cross-category learning spillovers

Table 8 presents the cross-category spillover matrix, showing how negative feedback in one category (rows) affects improvement in other categories (columns). Each cell in the matrix represents the coefficient from a separate regression of Equation (5), where negative feedback in the row category serves as the independent variable and improvement in the column category serves as the dependent variable. The diagonal elements—representing own-category learning—are consistently the largest in each row, indicating that learning is primarily targeted. For example, hosts criticized for host behavior improve most in host-related aspects; hosts who receive negative feedback about cleanliness improve most in cleanliness; and so on.

[Insert Table 8 here]

The off-diagonal elements are generally small, suggesting limited spillover effects. This pattern has important implications for understanding the learning process. Rather than responding to criticism with generalized effort across all dimensions—which might reflect either a "try harder everywhere" response or measurement error in the feedback—hosts appear to make targeted investments in the specific areas where guests have expressed dissatisfaction. This targeted response is consistent with the principles of targeted attention and efficient learning: hosts direct their limited time and resources toward the dimensions most in need of improvement based on the specific feedback received.

4.3.3 Category improvements and performance gains

The critical question for H3 is whether all category improvements generate equal performance returns. Table 9 presents category-specific estimates of the effect of sentiment improvement on revenue. Equation (6) is run with all categories' improvement measures. Categories are ordered by coefficient magnitude. The results reveal substantial heterogeneity.

[Insert Table 9 here]

A clear divide emerges between categories where improvement pays off and those where it does not. Improvements in overall evaluation (0.254), location (0.189), property (0.180), cleanliness (0.152), and host (0.146) all show large, positive, and statistically significant effects on revenue. These represent the actionable, host-controllable dimensions of the guest experience: hosts can improve cleanliness by more thorough cleaning, enhance property appeal through better amenities or maintenance, and provide better service and communication.

In contrast, improvements in parking (0.049), noise (0.027), sleep quality (−0.015), and value (−0.023) show small and statistically insignificant effects. Most strikingly, check-in (−0.144) and accuracy (−0.211) show negative and significant coefficients—listings that improve on these dimensions actually experience lower revenue.

What explains this heterogeneity? The categories with positive returns generally reflect dimensions where hosts have substantial control and where improvements represent genuine quality enhancements that guests value. Cleanliness, for instance, is directly within the host's control and is consistently cited as a top priority for guests. Property improvements—better furniture, upgraded amenities, more attractive décor—represent tangible investments that enhance the guest experience. Even for location where a listing cannot move, hosts can improve through clearer directions or neighborhood guides.

The categories with null or negative returns share a common feature: they tend to reflect guest expectations or external factors rather than objective, improvable quality. Value sentiment, for example, depends heavily on guests' price expectations rather than the property itself; a host cannot easily change whether guests perceive the listing as worth the price. Similarly, accuracy sentiment often reflects whether the listing matched guests' expectations based on photos and descriptions—negative accuracy reviews may indicate a fundamental mismatch between

marketing and reality that surface-level improvements cannot resolve. Noise and sleep quality also fall outside the host's direct control in many cases—external noise from traffic, neighbors, or construction cannot be remedied by host actions.

Taken together, the evidence provides nuanced support for H3. On one hand, hosts demonstrate the ability to learn from feedback across all evaluation categories—criticism triggers improvement regardless of the dimension. On the other hand, only improvements in actionable, host-controllable attributes translate into measurable performance gains. Theoretically, it suggests that the boundaries of learning from customer feedback are defined by the actionability of the feedback itself. Practically, it implies that hosts should prioritize investments in improvable dimensions—cleanliness, property amenities, service quality—rather than attempting to address concerns about subjective guest expectations.

5. DISCUSSION AND CONCLUSION

5.1 Summary of Findings

We examine how suppliers in the gig economy learn from customers through text reviews. We argue that customer sentiment embedded in review text highlights salient attributes that shape demand and provide firms with learning signals that suppliers can use to improve their performance. However, suppliers differ in learning ability: those with prior experience can respond to customer feedback more quickly and effectively than de novo entrants, leading to better performance. Moreover, learning may fail when feedback targets certain attributes, such as those that are hard-to-change.

We examine new Airbnb hosts to test our predictions. Consistent with this argument, we find that suppliers achieve performance gains by improving customer sentiment on attributes, and experienced suppliers respond to customer feedback more effectively. We also find that

while suppliers can improve on all attributes after receiving negative sentiments, improvements in some attributes do not necessarily translate to realized performance gains. These results are robust to alternative specifications.

5.2 Discussion and Conclusion

Our study has several implications for strategy research. First, it advances research on entrepreneurial learning by showing that learning from customers is attribute-specific, sentiment-driven, and uneven over time. Rather than viewing learning as a function of accumulated experience or abstract routines, our findings highlight that learning occurs through managerial attention (Ocasio & Joseph, 2018) to salient attributes emphasized in customer feedback, thereby clarifying the mechanisms through which learning unfolds. Second, the study contributes to the microfoundations of strategy by demonstrating that prior experience conditions learning effectiveness. Even when firms receive the same customer feedback on the same platform, their ability to interpret and act on that information varies systematically with experience. Learning effectiveness, therefore, depends not only on organizational processes but also on the experience of decision-makers, particularly in settings such as the gig economy where individuals drive strategic action. Third, the results temper optimistic views of learning by identifying clear limits. Learning does not translate into improved performance when feedback concerns attributes that are difficult to change or when early negative reviews create lasting reputational effects, underscoring the role of path dependence and early signals in shaping competitive outcomes.

This study also has implications for managers. First, it demonstrates the usefulness of customer review text as a tool for identifying learning opportunities and guiding performance improvement. However, the findings also highlight that not all customer feedback is equally actionable. Managers should prioritize learning from reviews that target attributes within their

control and recognize when feedback reflects structural constraints unlikely to be resolved through incremental improvements. Second, the results underscore the strategic importance of early-stage performance and reputation. Early negative reviews can impose persistent performance penalties, suggesting that managers should invest disproportionately in quality control and customer experience at entry. Finally, the study suggests that experience shapes responsiveness. Firms and platforms should support less experienced suppliers—through training, templates, or feedback interpretation tools—to help them translate customer feedback into effective action more quickly and avoid early missteps that are difficult to reverse.

This study has some limitations. First, we argue that experienced suppliers can respond to customer feedback more effectively than de novo entrants because past experience shapes learning outcomes. While our measures of experience identify hosts with prior or multiple listings, we recognize that the associated advantage may reflect both learning experience and scale economies (e.g., spreading fixed maintenance costs across properties). This interpretation is consistent with the entrepreneurship literature on incumbent new establishments (e.g., Balasubramanian & Sakakibara, 2023), where prior industry experience provides resource advantages beyond pure learning from managing Airbnb listings. While our two measures of experience offer some insight into this issue, additional examination is warranted. Second, our results depend on how accurately LLMs extract relevant categories and subcategories from text. We find that later versions of GPT models perform better in sequential processing; however, at this time, we use the best available version for batch processing, which may have constrained the accuracy of categorization. Third, our sample is limited to Airbnb listings in Los Angeles. Although this choice is driven by computational considerations, learning patterns may differ across locations. We consider these issues as future research agenda for researchers interested in

this subject.

Despite these limitations, we demonstrate the importance of unpacking the entrepreneurial learning process by observing both learning inputs and outputs. We hope that this research stimulates further inquiry into the mechanism through which learning unfolds.

REFERENCES

Alcacer, J., & Oxley, J. (2014). Learning by supplying. *Strategic Management Journal*, *35*(2), 204–223.

Anderson, M., & Magruder, J. (2012). Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *Economic Journal, 122* (563), 957–89.

Archak, N., Ghose, A., & Ipeirotis, P. G. (2011). Deriving the pricing power of product features by mining consumer reviews. *Management Science*, *57*(8), 1485–1509.

Argote, L., Beckman, S. L., & Epple, D. (1990). The persistence and transfer of learning in industrial settings. *Management Science*, *36*(2), 140–154.

Balasubramanian, N. (2011). New plant venture performance differences among incumbent, diversifying, and entrepreneurial firms: The impact of industry learning intensity. *Management Science*, *57*(3), 549–565.

Balasubramanian, N., & Lieberman, M. B. (2010). Industry learning environments and the heterogeneity of firm performance. *Strategic Management Journal*, *31*(4), 390–412.

Balasubramanian, N., & Lieberman, M. B. (2011). Learning-by-doing and market structure. *Journal of Industrial Economics*, *59*(2), 177–198.

Balasubramanian, N., & Sakakibara, M. (2023). Incidence and performance of spinouts and incumbent new establishments: Role of selection and redeployability within parent firms. *Strategic Management Journal, 44*(10), 2460–2488.

Barksdale, H. C., & Darden, W. R. (1972). Consumer attitudes toward marketing and consumerism. *Journal of Marketing*, *36*(4), 28–35.

Barrios, J. M., Hochberg, Y. V., & Yi, H. (2022). Launching with a parachute: The gig economy and new business formation. *Journal of Financial Economics*, *144*(1), 22-43.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing*, *84*(1), 1–25.

Bougie, R., Pieters, R., & Zeelenberg, M. (2003). Angry customers don't come back, they get back: The experience and behavioral implications of anger and dissatisfaction in services. *Journal of the Academy of Marketing Science*, *31*(4), 377–393.

Bower, J. L., & Christensen, C. M. (1996). Disruptive technologies: Catching the wave. *Harvard Business Review*, *74*(1), 43–53.

Bucher, E., Fieseler, C., Fleck, M., & Lutz, C. (2018). Authenticity and the sharing economy. *Academy of Management Discoveries*, *4*(3), 294–313.

Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. *Management Science*, *66*(2), 564-586.

Carlson, N. A., & Burbano, V. (2025). The use of LLMs to annotate data in management research: Foundational guidelines and warnings. *Strategic Management Journal*. Advance online publication. https://doi.org/10.1002/smj.70023

Carroll, C. D., Fuhrer, J. C., & Wilcox, D. W. (1994). Does consumer sentiment forecast household spending? If so, why? *American Economic Review*, *84*(5), 1397–1408.

Chakraborty, I., Kim, M., & Sudhir, K. (2022). Attribute sentiment scoring with online text reviews: Accounting for language structure and missing attributes. *Journal of Marketing Research*, *59*(3), 600–622.

Chen, Y., & Xie, J. (2005). Third-party product review and firm marketing strategy. *Marketing Science, 24*(2), 218–240.

Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science, 54*(3), 477–491.
Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research*, *43*(3), 345–354.
Cook, C., Diamond, R., Hall, J. V., List, J. A., & Oyer, P. (2021). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *Review of Economic Studies*, *88*(5), 2210–2238.
Darr, E. D., Argote, L., & Epple, D. (1995). The acquisition, transfer, and depreciation of knowledge in service organizations: Productivity in franchises. *Management Science*, *41*(11), 1750–1762.
Day, G. S. (1994). The capabilities of market-driven organizations. *Journal of Marketing*, *58*(4), 37–52.
Dellarocas, C. (2003). The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Science*, *49*(10), 1407–1424.
Dencker, J. C., Gruber, M., & Shah, S. K. (2009). Pre-entry knowledge, learning, and the survival of new firms. *Organization Science*, *20*(3), 516-537.
Egelman, C. D., Epple, D., Argote, L., & Fuchs, E. R. (2017). Learning by doing in multiproduct manufacturing: Variety, customizations, and overlapping product generations. *Management Science*, *63*(2), 405–423.
Eggers, J. P., & Song, L. (2015). Dealing with failure: Serial entrepreneurs and the costs of changing industries between ventures. *Academy of Management Journal*, *58*(6), 1785-1803.
Einav, L., Farronato, C., & Levin, J. (2016). Peer-to-Peer Markets. *Annual Review of Economics*, *8,* 615–635.
Farronato, C., & Zervas, G. (2022). *Consumer reviews and regulation: Evidence from NYC restaurants* (No. w29715). National Bureau of Economic Research.
Filippas, A., Horton, J. J., & Zeckhauser, R. J. (2020). Owning, using, and renting: Some simple economics of the "sharing economy". *Management Science*, *66*(9), 4152-4172.
Ganu, G., Elhadad, N., & Marian, A. (2009). Beyond the stars: improving rating predictions using review text content. In *WebDB* (Vol. 9, pp. 1–6).
Gao, G., Greenwood, B. N., McCullough, J., & Agarwal, R. (2012). A digital soapbox? The information value of online physician ratings. In *Academy of Management Proceedings* (Vol. 2012, No. 1, p. 16354). Briarcliff Manor, NY: Academy of Management.
Garud, R., Kumaraswamy, A., Roberts, A., & Xu, L. (2022). Liminal movement by digital platform-based sharing economy ventures: The case of Uber technologies. *Strategic Management Journal*, *43*(3), 447–475.
Gaski, J. F., & Etzel, M. J. (1986). The index of consumer sentiment toward marketing. *Journal of Marketing*, *50*(3), 71–81.
Ghose, A., & Ipeirotis, P. G. (2011). Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, *23*(10), 1498–1512.
Gillitzer, C., & Prasad, N. (2018). The effect of consumer sentiment on consumption: Cross-sectional evidence from elections. *American Economic Journal: Macroeconomics*, *10*(4), 234–269.

Gompers, P., Kovner, A., Lerner, J., & Scharfstein, D. (2010). Performance persistence in entrepreneurship. *Journal of Financial Economics*, *96*(1), 18-32.
Greenberg, J., Sands, D. B., Cattani, G., & Porac, J. (2024). Rating systems and increased heterogeneity in firm performance: Evidence from the New York City restaurant industry, 1994–2013. *Strategic Management Journal*, *45*(1), 36–65.
Haggag, K., McManus, B., & Paci, G. (2017). Learning by driving: Productivity improvements by New York City taxi drivers. *American Economic Journal: Applied Economics*, *9*(1), 70–95.
He, S., Hollenbeck, B., & Proserpio, D. (2022). The market for fake reviews. *Marketing Science*, *41*(5), 896–921.
Heydari, A., Tavakoli, M., Salim, N., & Heydari, Z. (2015). Detection of review spam: A survey. *Expert Systems with Applications*, *42*(7), 3634–3642.
Humphreys, A., & Wang, R. J. H. (2018). Automated text analysis for consumer research. *Journal of Consumer Research*, *44*(6), 1274–1306.
Jain, A. (2016). Learning by hiring and change to organizational knowledge: Countering obsolescence as organizations age. *Strategic Management Journal*, *37*(8), 1667–1687.
Jarvenpaa, S. L., & Leidner, D. E. (1999). Communication and trust in global virtual teams. *Organization Science*, *10*(6), 791-815.
Jaworski, B. J., & Kohli, A. K. (1993). Market orientation: Antecedents and consequences. *Journal of Marketing*, *57*(3), 53–70.
Jovanovic, B. (1982). Selection and the Evolution of Industry. *Econometrica, 50*(3), 649-670.
Kale, P., & Singh, H. (2007). Building firm capabilities through learning: The role of the alliance learning process in alliance capability and firm-level alliance success. *Strategic Management Journal*, *28*(10), 981–1000.
Kerr, W. R., Nanda, R., & Rhodes-Kropf, M. (2014). Entrepreneurship as experimentation. *Journal of Economic Perspectives*, *28*(3), 25-48.
Lafontaine, F., & Shaw, K. (2016). Serial entrepreneurship: Learning by doing? *Journal of Labor Economics*, *34*(S2), S217-S254.
Lieberman, M. B. (1984). The learning curve and pricing in the chemical processing industries. *RAND Journal of Economics, 15*(2), 213–228.
Lieberman, M. B. (1989). The learning curve, technology barriers to entry, and competitive survival in the chemical processing industries. *Strategic Management Journal*, *10*(5), 431–447.
Lu, S. F., & Rui, H. (2018). Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. *Management Science*, *64*(6), 2557–2573.
Luca, M. (2016). Reviews, reputation, and revenue: The case of Yelp.com. *Harvard Business School NOM Unit Working Paper* (12-016).
Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and Yelp review fraud. *Management Science*, *62*(12), 3412–3427.
Lundstrom, W. J., & Lamont, L. M. (1976). The development of a scale to measure consumer discontent. *Journal of Marketing Research*, *13*(4), 373–381.
Manso, G. (2016). Experimentation and the Returns to Entrepreneurship. *The Review of Financial Studies*, *29*(9), 2319-2340.
Mayzlin, D., Dover, Y., & Chevalier, J. (2014). Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review*, *104*(8), 2421–2455.

Miller, A., O'Mahony, S., & Cohen, S. (2024). Opening the aperture: Explaining the complementary roles of advice and testing when forming entrepreneurial strategy. *Organization Science*, *35*(1), 1–26.

Miric, M., Jia, N., & Huang, K. G. (2022). Using supervised machine learning for large-scale classification in management research: The case for identifying artificial intelligence patents. *Strategic Management Journal*, *44*(2), 491−519.

Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies*, *78*(1), 356–393.

Ocasio, W., & Joseph, J. (2018). The attention-based view of great strategies. *Strategy Science*, *3*(1), 289–294.

Proserpio, D., Xu, W., & Zervas, G. (2018). You get what you give: Theory and evidence of reciprocity in the sharing economy. *Quantitative Marketing and Economics*, *16*, 371–407.

Reimers, I., & Waldfogel, J. (2021). Digitization and pre-purchase information: The causal and welfare impacts of reviews and crowd ratings. *American Economic Review*, *111*(6), 1944–1971.

Ryu, W., McCann, B. T., & Wan, W. P. (2022). Learning by doing and corporate diversification. *Strategic Management Journal*, *43*(8), 1635–1665.

Schneider, H. (2010). Moral hazard in leasing contracts: Evidence from the New York City taxi industry. *Journal of Law and Economics*, *53*(4), 783–805.

Shane, S. (2000). Prior knowledge and the discovery of entrepreneurial opportunities. *Organization Science*, *11*(4), 448-469.

Song, J., Almeida, P., & Wu, G. (2003). Learning-by-hiring: When is mobility more likely to facilitate interfirm knowledge transfer? *Management Science*, *49*(4), 351–365.

Tadelis, S. (2016). Reputation and feedback systems in online platform markets. *Annual Review of Economics*, *8*, 321–340.

Tambe, P., Ye, X., & Cappelli, P. (2020). Paying to program? Engineering brand and high-tech wages. *Management Science*, *66*(7), 3010–3028.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal*, *18*(7), 509–533.

Vidanagama, D. U., Silva, T. P., & Karunananda, A. S. (2020). Deceptive consumer review detection: A survey. *Artificial Intelligence Review*, *53*(2), 1323–1352.

Villarroel Ordenes, F., Ludwig, S., De Ruyter, K., Grewal, D., & Wetzels, M. (2017). Unveiling what is written in the stars: Analyzing explicit, implicit, and discourse patterns of sentiment in social media. *Journal of Consumer Research*, *43*(6), 875–894.

von Hippel, E. (1986). Lead users: A source of novel product concepts. *Management Science*, *32*(7), 791–805.

Wu, Y., Ngai, E. W., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems*, *132*, 113280.

Xu, Y., Armony, M., & Ghose, A. (2021). The interplay between online reviews and physician demand: An empirical investigation. *Management Science*, *67*(12), 7344–7361.

Xu, Y., Lu, B., Ghose, A., Dai, H., & Zhou, W. (2023). The interplay of earnings, ratings, and penalties on sharing platforms: An empirical investigation. *Management Science*, *69*(10), 6128-6146.

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, *74*(2), 133–148.

Table 1: Categories and Subcategories

Categories	Subcategories	Example	% appears in data	% of negative sentiment
Overall evaluation		"My family and I really enjoyed our stay."	96.0	4.0
Location		"The apartment is centrally located."	57.5	2.0
	Proximity to attractions	Beach, Hollywood, downtown, central location	30.8	0.6
	Transportation access	Proximity to buses, airport	5.0	3.3
	Walkability	Shops, restaurants	4.4	2.5
	Neighborhood quality	Safety, comfort	15.5	7.5
Host		"The host was not friendly, and was impossible to reach."	55.3	3.8
	Friendliness	Kind, provide tips, making guests feel at home	38.2	0.6
	Communication	Responsive, resolve issues	24.6	3.9
	Cancellation of the reservation	"The host canceled my reservation the day before arrival."	1.0	97.7
Check-in		Easy, clear instructions	9.2	7.9
Cleanliness		"The house is super clean."	31.9	9.1
	Bathroom	"Some physical flaws (stains) in the private bathroom."	3.1	33.3
	Kitchen	"Rotting vegetables in fridge"	1.3	39.3
	Bedding/linen	"Fresh bedding"	1.7	29.5
	Other areas	Bedroom, living room	3.1	44.8
Property		House, apartment, room	65.2	4.0
	Comfort	Cozy, spacious, old	33.3	3.9
	Amenities (indoor)	Kitchen items, Wi-Fi, furniture	21.0	8.7
	Aesthetics	Modern décor, outdated	17.5	3.7
	Outdoor space	Pool, patio, backyard	6.3	1.5
	Privacy	"Shared with a few other guests"	4.6	12.2
	Climate control	AC, heater	2.3	39.5
Value		"worth the price", "good value", "overpriced"	6.2	10.8
Accuracy		"The listing's descriptions do not match the actual listing."	4.8	15.1
Noise		"There was some rap music playing from the work yard"	6.6	49.1
Sleep quality		"Bed was SO comfortable."	3.9	24.6
Parking		Private parking, easy street parking	9.9	22.2

Note. Multiple categories and subcategories can appear in a review. "Overall evaluation" applies when the review mentions general issues.

Table 2: Summary Statistics

Variable	N	Mean	SD	Min	P25	Median	Max
Panel A: Outcome Variables							
Bookings (nights/30 days)	438,886	17.00	8.47	0	10	18	30
Price ($)	438,886	143.39	175.15	1	70	105	10,000
Revenue ($/30 days)	438,886	2,386	2,982	0	900	1,800	300,000
Panel B: Sentiment Measure							
Lagged sentiment (3-mo MA)	438,886	0.879	0.234	−1.000	0.833	1.000	1.000
Panel C: Listing Characteristics							
Bedrooms	438,766	1.32	1.01	0	1	1	22
Amenities count	438,886	27.56	11.69	1	19	25	110
Initial price ($)	438,886	142.58	170.90	1	70	101	10,000
Room type	438,886	Entire home (65.2%), Private (31.0%), Shared (3.7%), Hotel (0.2%)					
Panel D: Host Characteristics							
Host listings count	438,535	6.41	24.81	0	1	2	2,136
First-time host	438,886	0.360	0.480	0	0	0	1
Multi-listing host	438,535	0.628	0.482	0	0	1	1

Notes: Summary statistics at the listing-month level for the 3-month moving average (MA) analysis sample (438,886 observations). Sample restricted to listing-months with a review in the current month and valid lagged 3-month moving average sentiment. *Lagged sentiment* is the one-month lag of the 3-month moving average sentiment score (−1 to +1). *First-time host* = 1 if host operates only one listing; *Multi-listing host* = 1 if host operates multiple listings. Regressions use log-transformed versions of Bookings, Price, Revenue, Bedrooms, Amenities, Initial price, and Host tenure.

Table 3: Sentiment and Host Performance

	(1) OLS	(2) Controls	(3) Listing FE
Panel A: Bookings (ln)			
$Sentiment_{t-1}$	0.336^{***}	0.289^{***}	0.100^{***}
	(0.012)	(0.011)	(0.009)
Panel B: Price (ln)			
$Sentiment_{t-1}$	0.083^{***}	0.030^{***}	0.002
	(0.016)	(0.005)	(0.002)
Panel C: Revenue (ln)			
$Sentiment_{t-1}$	0.523^{***}	0.406^{***}	0.134^{***}
	(0.025)	(0.017)	(0.015)
N	286,670	286,617	281,223
Controls	No	Yes	Yes
Time FE	No	Yes	Yes
Neighborhood FE	No	Yes	No
Listing FE	No	No	Yes

Notes: Dependent variables are in natural logarithms. *Sentiment* is the one-month lagged 3-month moving average of overall evaluation sentiment, scaled from −1 (entirely negative) to +1 (entirely positive). Column (1) presents OLS estimates without controls or fixed effects. Column (2) adds controls and time and neighborhood fixed effects. Column (3) presents the preferred specification with listing and time fixed effects, identifying effects from within-listing variation over time. Standard errors clustered by listing in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 4: Lagged Sentiment Effects

	Bookings		Price		Revenue	
	(1)	(2)	(3)	(4)	(5)	(6)
$Sentiment_{t-1}$	0.100^{***}	0.077^{***}	0.002	−0.002	0.134^{***}	0.096^{***}
	(0.009)	(0.011)	(0.002)	(0.002)	(0.015)	(0.019)
$Sentiment_{t-2}$		0.050^{***}		0.007^{**}		0.082^{***}
		(0.011)		(0.003)		(0.019)
N	281,223	238,083	281,223	238,083	281,223	238,083
Listing FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variables are in natural logarithms. *Sentiment* is the 3-month moving average of overall evaluation sentiment (−1 to +1). Odd columns include only one-month lagged sentiment; even columns include both one-month and two-month lags. All specifications include listing and time fixed effects and control for log nearby listings. Standard errors clustered by listing in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 5: Experience Heterogeneity—First-time vs. Repeat Hosts

	Bookings		Price		Revenue	
	First	Repeat	First	Repeat	First	Repeat
	(1)	(2)	(3)	(4)	(5)	(6)
$Sentiment_{t-1}$	0.078***	0.103***	0.001	0.003	0.115***	0.135***
	(0.017)	(0.010)	(0.003)	(0.003)	(0.028)	(0.018)
N	98,955	181,088	98,955	181,088	98,955	181,088
Listing FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variables are in natural logarithms. *Sentiment* is the one-month lagged 3-month moving average of overall evaluation sentiment (−1 to +1). First-time hosts are those operating their first Airbnb listing; repeat hosts have operated previous listings on the platform. All specifications include listing and time fixed effects and control for log nearby listings. Standard errors clustered by listing in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 6: Experience Heterogeneity—Single vs. Multi-listing Hosts

	Bookings		Price		Revenue	
	Single	Multi	Single	Multi	Single	Multi
	(1)	(2)	(3)	(4)	(5)	(6)
$Sentiment_{t-1}$	0.077***	0.102***	0.001	0.003	0.115***	0.132***
	(0.017)	(0.010)	(0.003)	(0.003)	(0.027)	(0.018)
N	100,280	179,775	100,280	179,775	100,280	179,775
Listing FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variables are in natural logarithms. *Sentiment* is the one-month lagged 3-month moving average of overall evaluation sentiment (−1 to +1). Single-listing hosts operate one listing; multi-listing hosts operate two or more listings simultaneously. All specifications include listing and time fixed effects and control for log nearby listings. Standard errors clustered by listing in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 7: Own-Category Learning

Category	Coefficient	SE	*N*
Overall evaluation	0.060***	(0.001)	300,597
Location	0.038***	(0.001)	296,797
Host	0.038***	(0.001)	296,077
Cleanliness	0.029***	(0.001)	276,798
Property	0.041***	(0.001)	299,232
Check-in	0.031***	(0.002)	173,406
Accuracy	0.032***	(0.003)	109,539
Value	0.036***	(0.003)	130,562
Noise	0.036***	(0.002)	137,098
Sleep quality	0.038***	(0.003)	97,456
Parking	0.035***	(0.002)	178,112

Notes: Dependent variable is the maximum sentiment improvement in the same category over the subsequent 1–3 months. The independent variable is an indicator equal to 1 if sentiment in that category declined by more than 0.1 points in the current month. All specifications include listing and time fixed effects. Standard errors clustered by listing in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table 8: Cross-Category Spillover Matrix

Shock	Ovrl	Loc	Host	Cln	Prop	Chk	Acc	Val	Nois	Slp	Park
Overall	**0.060**	0.009	0.020	0.014	0.018	0.003	0.002	0.003	0.005	0.002	0.002
Location	0.009	**0.038**	0.005	0.005	0.008	0.002	0.002	0.002	0.002	0.002	0.003
Host	0.015	0.005	**0.038**	0.005	0.006	0.004	0.002	0.002	0.001	0.001	0.001
Cleanliness	0.016	0.004	0.007	**0.029**	0.011	0.003	0.002	0.001	0.002	0.001	0.001
Property	0.014	0.007	0.006	0.009	**0.041**	0.001	0.003	0.002	0.002	0.002	0.002
Check-in	0.021	0.004	0.010	0.008	0.015	**0.031**	0.005	0.002	0.004	0.001	0.005
Accuracy	0.018	0.005	0.007	0.000	0.016	0.003	**0.032**	0.004	0.003	0.000	0.006
Value	0.027	0.007	0.010	0.010	0.023	0.003	0.007	**0.036**	0.004	0.004	0.007
Noise	0.015	0.007	−0.003	0.003	0.005	0.003	0.003	0.001	**0.036**	0.005	0.008
Sleep	0.023	0.004	0.006	0.011	0.008	0.005	0.004	0.002	0.016	**0.038**	0.007
Parking	0.008	0.004	0.006	0.002	0.000	0.003	0.000	0.003	0.007	0.000	**0.035**

Notes: Each cell shows the coefficient from regressing the maximum 3-month improvement in the column category on a negative shock for the row category. Diagonal elements (bold) represent own-category learning; off-diagonal elements represent cross-category spillovers. Column abbreviations: Ovrl = Overall evaluation, Loc = Location, Cln = Cleanliness, Prop = Property, Chk = Check-in, Acc = Accuracy, Val = Value, Nois = Noise, Slp = Sleep quality, Park = Parking. All specifications include listing and time fixed effects.

Table 9: Category Improvement Effects on Revenue

Category	Coefficient	SE	*N*
Overall evaluation	0.254***	(0.043)	339,483
Location	0.189***	(0.033)	334,250
Property	0.180***	(0.031)	337,585
Cleanliness	0.152***	(0.034)	310,173
Host	0.146***	(0.033)	333,593
Parking	0.049	(0.055)	196,586
Noise	0.027	(0.077)	151,387
Sleep quality	−0.015	(0.117)	106,885
Value	−0.023	(0.094)	145,024
Check-in	−0.144**	(0.069)	191,364
Accuracy	−0.211**	(0.099)	119,994

Notes: Dependent variable is ln(revenue). *Improvement* is defined as current minus initial sentiment in each category, capturing cumulative improvement since listing entry. Categories are ordered by coefficient magnitude. All specifications include listing and time fixed effects and control for log nearby listings. Standard errors clustered by listing in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

APPENDIX

Table A1: First-Difference Specification

	(1) Δ Bookings	(2) Δ Price	(3) Δ Revenue
Δ *Sentiment*$_{t-1}$	0.066***	−0.000	0.101***
	(0.011)	(0.001)	(0.019)
N	272,093	272,093	272,093
Time FE	Yes	Yes	Yes

Notes: First-difference removes all time-invariant listing-level unobserved heterogeneity. Δ denotes the change from month $t-1$ to month t. Standard errors clustered by listing in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Table A2: Alternative Specifications

	(1) Bookings	(2) Price	(3) Revenue
Panel A: Two-Month Lagged Sentiment			
Sentiment$_{t-2}$	0.097***	0.006**	0.141***
	(0.010)	(0.003)	(0.016)
N	238,083	238,083	238,083
Panel B: Low Initial Sentiment Subsample			
Sentiment$_{t-1}$	0.151***	−0.001	0.203***
	(0.018)	(0.004)	(0.029)
N	49,797	49,797	49,797
Panel C: Cumulative Improvement Measure			
Improvement	0.165***	0.016	0.254***
	(0.024)	(0.012)	(0.043)
N	339,539	339,539	339,539
Panel D: Inverse Probability Weighting			
Sentiment$_{t-1}$	0.116***	−0.000	0.156***
	(0.012)	(0.003)	(0.020)
N	281,174	281,174	281,174
Listing FE	Yes	Yes	Yes
Time FE	Yes	Yes	Yes

Notes: Panel A uses two-month lagged sentiment instead of one-month lag. Panel B restricts the sample to listings in the bottom quartile of initial sentiment scores. Panel C uses a cumulative improvement measure defined as current minus initial sentiment. Panel D employs inverse probability weighting (IPW) to address potential selection into sentiment improvement; first-stage propensity scores are estimated using listing characteristics and lagged performance, and observations with propensity scores below 0.05 or above 0.95 are trimmed. All specifications include listing and time fixed effects and control for log nearby listings. Standard errors clustered by listing in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.