

Belletti, Chiara; Pronkina, Elizaveta; Rossi, Michelangelo

Working Paper

Rating Systems and the End-Game Effect: When Reputation Works and when It Doesn't

CESifo Working Paper, No. 11253

Provided in Cooperation with:

Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Belletti, Chiara; Pronkina, Elizaveta; Rossi, Michelangelo (2024) : Rating Systems and the End-Game Effect: When Reputation Works and when It Doesn't, CESifo Working Paper, No. 11253, CESifo GmbH, Munich

This Version is available at:

<https://hdl.handle.net/10419/302738>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Rating Systems and the End-Game Effect: When Reputation Works and When It Doesn't

Chiara Belletti, Elizaveta Pronkina, Michelangelo Rossi

Impressum:

CESifo Working Papers

ISSN 2364-1428 (electronic version)

Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH

The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute

Poschingerstr. 5, 81679 Munich, Germany

Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de

Editor: Clemens Fuest

<https://www.cesifo.org/en/wp>

An electronic version of the paper may be downloaded

- from the SSRN website: www.SSRN.com
- from the RePEc website: www.RePEc.org
- from the CESifo website: <https://www.cesifo.org/en/wp>

Rating Systems and the End-Game Effect: When Reputation Works and when It Doesn't

Abstract

Do rating systems provide incentives to sellers when they are about to exit a market? Using data from Airbnb, this paper examines how end-of-game considerations affect hosts' effort decisions. We take advantage of a regulation on short-term rentals in the City of Los Angeles to identify hosts who anticipated their imminent exit from the platform due to non-compliance with new eligibility rules. We focus on hosts who left the platform as a result of the regulation and measure their effort with listing's ratings in effort-related categories such as check-in, communication and cleanliness. With a Difference-in-Differences and Event Study approach, we compare how listing's effort-related ratings changed, compared to ratings on location, after the regulation announcement and during its implementation. We document a statistically significant decrease in effort in the last periods of the hosts' career. Our findings provide insights for platform managers, highlighting the adverse effects of end-of-game considerations on how rating systems affect sellers' incentives for the provision of high-quality services.

JEL-Codes: D820, L140, L860.

Keywords: rating systems, online reputation, digital platforms.

Chiara Belletti
LMU Munich School of Management
Munich / Germany
chiara.belletti@lmu.de

Elizaveta Pronkina
Université Paris-Dauphine – PSL
Paris / France
elizaveta.pronkina@dauphine.psl.eu

Michelangelo Rossi
Télécom Paris & Institut Polytechnique de Paris
Paris / France
michelangelo.rossi@telecom-paris.fr

July 26, 2024

We would like to thank Marc Bourreau, Jorg Claussen, Jose De Sousa, Laszlo Goerke, Tobias Kretschmer, Ulrich Laitenberger, Mark J. Tremblay, and Nikhil Vellodi. Helpful feedback was received at the seminars at Telecom Paris, the 3rd Crowdfunding Symposium 2021, the AFREN doctoral workshop 2022, the ZEW ICT conference 2022, the CESifo area conference on Economics of Digitization 2022, the 6th Doctoral Workshop on the Economics of Digitization 2023, the 2024 Digital Economy Workshop in Rotterdam and the 15th Paris Digital Economics Conference. All errors are ours. E. Pronkina acknowledges that this work was done prior to the author joining Amazon.

1 Introduction

Online platforms use ratings to inform buyers about the quality of sellers (Dellarocas, 2003; Chen and Xie, 2008). Moreover, rating systems also motivate sellers to maintain consistently high standards. Poor performance can significantly lower their cumulative ratings, encouraging them to provide high-quality services over time (Hui, Saeedi, Shen and Sundaresan, 2016; Hui, Saeedi and Sundaresan, 2018). This extended process of reputation building helps weed out bad actors. As a result, many platforms aggregate all historical ratings into a single average, leveraging the collective wisdom of the crowd while still allowing buyers to view individual reviews for more recent transactions.¹

Maintaining high service quality is costly, so if sellers anticipate leaving the platform, they may choose to lower the quality, receive low ratings, and monetize their reputation before leaving. This behavior is particularly problematic for online platforms and marketplaces where demand is seasonal and seller turnover is high such as Airbnb (see Fradkin (2017); Farronato and Fradkin (2022)). Moreover, sellers can use this strategy more effectively when ratings are aggregated, because their average ratings will not drop rapidly. This allows them to profit from their accumulated reputation until they leave. Therefore, aggregating all reviews into a single average rating makes sense for platforms if they prioritize screening sellers at the beginning of their career – given that many sellers start with zero reputation and few will work to earn positive, consistent feedback – over avoiding sellers to “milk” their reputation before leaving the platform.

Several empirical studies have examined how sellers’ performance improves as they accumulate more and better reviews (Luca, 2011; Anderson and Magruder, 2012; Mayzlin, Dover and Chevalier, 2014; Babić Rosario, Sotgiu, De Valck and Bijmolt, 2016). Other papers have shown the power of a single review to affect consumers’ choice (Vana and Lambrecht, 2021; Varga and Albuquerque, 2023). In this paper, we shift the focus to sellers’ choices to *not* provide high-quality services and get bad reviews as a result: i.e., the process of “reputation milking”. We examine how end-game considerations, in particular the anticipation of a forthcoming exit from the platform, affect sellers’ effort decisions in their last transactions. In doing so, we answer the following question: Do rating systems provide enough incentives to sellers when they are about to exit a market?

Our analysis focuses on the behavior of hosts on Airbnb. The market for short-term rentals (STR) has been increasingly regulated worldwide and many hosts have been compelled to leave the market due to their inability to comply with stricter eligibility criteria and licensing requirements. The anticipated announcement of these policies before their effective enforcement allows hosts to anticipate their imminent

¹This design of the rating system is common in many online platforms, including Amazon, eBay, and Airbnb (Tadelis, 2016).

exit and potentially adjust their behavior accordingly. This could undermine the role of ratings in motivating hosts, as they may attempt to “milk” their accumulated reputation and reduce their effort to provide high-quality service.

To empirically evaluate how end-game considerations affect the effort choices of hosts, we collected scraped data from Airbnb in Los Angeles. We measured effort based on guest evaluations of listings in rating categories such as check-in, communication, and cleanliness. To identify hosts’ anticipated exits from the platform, we leveraged the implementation of the Home-Sharing Ordinance (HSO) in the City of Los Angeles. The HSO was approved by the Los Angeles City Council in December 2018 as a regulatory measure for the short-term rental market. Under the Ordinance, hosts are allowed to rent out their dwellings for fewer than 30 consecutive days, only if they register and pay a license fee. Eligibility for obtaining a license was conditional on specific dwelling characteristics, limiting licenses to the host’s primary residence. Airbnb enforced the regulation by removing listings that failed to provide a license number within the due period. The implementation of the HSO spanned four months, from the beginning of July to the end of October 2019, during which more than one-fifth of short-term Airbnb listings in Los Angeles left the platform. Due to the lag between policy approval (and its announcement) and the implementation, ineligible listings or those unwilling to register could anticipate their forthcoming exit from the platform.

The eligibility criteria set by the regulation were unrelated to the rating history of listings. Thus, restricting the analysis to hosts renting in the City of Los Angeles who exited due to the HSO allows us to disentangle the effect of end-game considerations from organic exits induced by the accumulation of negative reviews. Using a Difference-in-Differences (DiD) approach and an Event Study (ES) analysis, we compare the evolution of ratings in effort-related categories before and after the HSO with ratings on the listing’s location. We choose ratings on location as control since location should be independent of the host’s effort and therefore not affected by end-game considerations.

Our analysis documents that when hosts anticipate their exit, effort-related ratings in their final transactions decrease significantly. The maximum magnitude of the effect varies from about -0.07 points for check-in, -0.08 points for communication ratings and up to -0.11 points for cleanliness ratings. This corresponds to a decrease of respectively 1.5%, 1.7% to 2.4% from the pre-regulation average rating. The magnitude of the DiD estimates is remarkable given that ratings on Airbnb (on a scale of 1 to 5 stars) are particularly sticky and left-skewed. Indeed, for a listing with a rating of 5 stars, a decline of our estimated magnitude, means moving from the top percentile to around the 40th percentile of the cleanliness rating distribution and up below the 20th percentile of the rating distribution of communication and check-in ratings.²

²Computation based on the rating distribution among short-term listings in the City of Los Angeles during the time span of our

The magnitude of the end-of-game effects does not appear to be affected by pre-regulation ratings of the overall experience and location of listings. Yet, the decline in ratings for check-in and communication during the final transactions is more pronounced for “professional” hosts (hosts managing more than two listings). This suggests that the Airbnb rating system is only partially effective at screening out strategic sellers who reduce their effort levels at the expense of guests as they exit the platform. Consequently, guests may face diminished service quality from professional hosts nearing the end of their tenure, highlighting a potential vulnerability in the platform’s ability to maintain consistent service standards.

To validate our results, we exploit the geographical variation of the policy. The HSO only applied to the city of Los Angeles. The other 87 cities and unincorporated areas in Los Angeles County were not subject to the ordinance or a similar concurrent policy. We conduct a DiD analysis comparing the effort-related ratings of listings in the City of Los Angeles that left the platform during the implementation of the HSO with those of listings in other cities in the county that left the platform during the same period. Similar to our main specification, we find that listings affected by the HSO experienced a statistically significant decrease in ratings for check-in and communication during the implementation of the regulation. Ratings for location are not affected by the regulation, confirming the validity of our control group in the main specification.

Our paper contributes to several strands of the literature. The first contribution relates to how rating systems (and their design) affect sellers’ behavior. [Klein, Lambertz and Stahl \(2016\)](#), [Hui et al. \(2016\)](#), and [Hui et al. \(2018\)](#) explore how changes in rating systems impact sellers’ quality both in terms of the entry and exit of high-quality sellers (adverse selection) and sellers’ incentives to exert effort (moral hazard). Our paper contributes to this literature by highlighting how the power of reputation incentives to discipline seller behavior weakens at the end of sellers’ careers on the platform.

Consistent with our approach, [Cabral and Hortacsu \(2010\)](#) and [Fan, Ju and Xiao \(2016\)](#) examine the life cycle of sellers on online platforms, explicitly exploring the notion that sellers first build reputation and then exploit it before exiting. While both papers provide comprehensive empirical analyses supporting the processes of reputation building and milking, we exploit the HSO implementation to credibly identify sellers who anticipate their exit and subsequently reduce their effort. This allows us to measure the causal effect of exit anticipation on sellers’ behavior.

More generally, our paper contributes to the empirical literature highlighting how the signals provided by ratings are imperfect or biased ([Dellarocas and Wood, 2008](#)). Previous papers have shown that ratings can be manipulated with fake reviews ([Mayzlin et al., 2014](#); [He, Hollenbeck and Proserpio, 2022](#)) and may suffer from biases related to sellers’ pricing ([Carnehl, Schaefer, Stenzel and Tran, 2022](#)), selection ([Bondi,](#)

analysis.

2023; Bondi, Rossi and Stevens, 2023), or behavioral aspects (Fradkin, Grewal and Holtz, 2021; Proserpio, Xu and Zervas, 2018). We extend this literature by showing the limits of reputation incentives right before sellers' exit.

Finally, we also contribute to the literature on how agents decide to exert effort over their careers. Career concerns have been studied and analyzed from a theoretical point of view starting with Gibbons and Murphy (1992), Holmström (1999) and the theoretical literature on reputation (see Bar-Isaac and Tadelis (2008) for a survey). Empirical evidence on career concerns is much less common. Miklós-Thal and Ullrich (2016) study the career concerns of professional soccer players during the selections for national teams for the European Cup. Xu, Nian and Cabral (2020) explore non-pecuniary incentives for users on the Q&A platform Stack Overflow. In contrast with the previous studies, our paper focuses on the causal relationship between sellers' effort and a definitive career termination on a digital platform.

The remaining part of the paper is organized as follows: Section 2 describes the empirical setting of our work: Airbnb and the regulation of short-term listings in the City of Los Angeles. Section 3 presents our dataset and the main variables of interest. Section 4 introduces our identification strategy. Section 5 illustrates our main results followed by a series of heterogeneity analysis, placebo tests and robustness checks. We conclude and discuss our results in Section 6.

2 Background

2.1 Home-Sharing Ordinance in the City of Los Angeles

Many cities have adopted short-term rental (STR) regulations in recent years, and the City of Los Angeles is no exception (Koster, Van Ommeren and Volkhausen, 2021). In December 2018, the city council approved the Home-Sharing Ordinance, which was to be implemented and enforced in the following months. The Home-Sharing Ordinance went into effect on July 1, 2019. Starting from this date, hosts were able to register for home sharing using the city's online registration portal. Beginning November 1, 2019, the Department began overseeing enforcement of the ordinance.³ According to the HSO, in the City of Los Angeles, renting for periods shorter than 30 consecutive days is permitted only in the host's primary residence and for up to 120 days in a calendar year. Hosts are required to apply for a permit number and communicate it to the Airbnb platform. Airbnb announced that failing to provide the permit number would result in hosts being blocked from the platform.⁴

³Refer to the text of HSO at <https://clkrep.lacity.org/online/docs/2014/14-1635-S2.pdf>.

⁴Airbnb's announcement: <https://www.airbnb.com/help/article/864/los-angeles-ca.shortterm>.

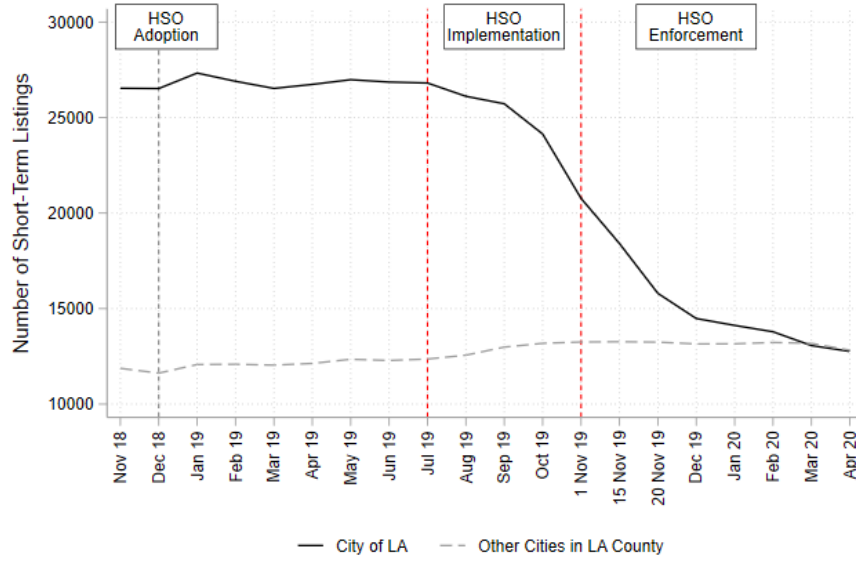


Figure 1. Number of short-term listings rented on Airbnb

Notes: The plot illustrates the evolution in the number of STR listings on Airbnb by scraping date (approximately each month). The solid line represents the number of STR listings rented on Airbnb within the City of Los Angeles. The line plots a remarkable decrease in the number of STR listings during the HSO implementation (1st of July - 1st of November 2019), and in the few first months after. The dashed line shows the more stable pattern of STR listings supply in the other 87 cities and unincorporated areas of the county of Los Angeles.

Figure 1 depicts the direct impact of the HSO. It shows the availability of residential properties listed for short-term rent on Airbnb within the City of Los Angeles, as well as in other cities within the county.⁵ During the implementation of the HSO, the number of short-term listings in Los Angeles began to decline, and by the beginning of 2020, it had almost halved. Conversely, the other cities unaffected by similar housing policies exhibited a stable pattern, with no significant fluctuations in the supply of STR listings.

3 Data

3.1 *InsideAirbnb* Data

We collect data from *InsideAirbnb* for the county of Los Angeles. *InsideAirbnb* is a website that provides scraped data from Airbnb.⁶ It scrapes the platform on a regular basis, often once per month, and collects information about active listings. This includes fixed characteristics of the house (e.g., latitude and longitude of the listing) and time-varying characteristics (e.g., ratings, number of minimum nights for rent, and

⁵The county of Los Angeles includes the City of Los Angeles (the green area in Figure A1 in Appendix) and 87 other cities and unincorporated areas (pink areas with black borders in Figure A1 in Appendix).

⁶For more information about the dataset, see: <https://insideairbnb.com/>.

prices). We match the latitude and longitude of each listing with city borders from Los Angeles city planning official statistics to clearly identify those within the City of Los Angeles.⁷ We restrict the sample to listings that offer stays for less than 30 days. In this way, we can identify the short-term listings that were subject to the HSO rules. We additionally restrict the analysis to listings that exited the platform during the regulation implementation and entered before or at the very beginning of December 2018, before the policy was announced.⁸ To check for parallel trends in the periods before the regulation, the estimation window is extended to fifteen snapshots before the implementation of the HSO.

3.2 Measuring Hosts' Effort Through Ratings

After a stay, a feedback system allows guests and hosts to rate each other within 14 days of the end of the stay. Guests can rate the overall experience and separately rate six subcategories: the accuracy of the listing description, the check-in process, the cleanliness of the listing, the communication with the host, the location of the listing, and the value for the money. We measure hosts' effort through their listing's ratings in related categories (i.e., check-in, communication, and cleanliness ratings). Airbnb reports ratings on a five-point scale and displays the cumulative simple average over time. We observe, $\bar{R}_{i,t}^k$, the rounded cumulative average rating at time t for listing i in category k , and $n_{i,t}$, the cumulative number of new reviews for each listing i at time t . We compute the average rating per snapshot for each rating category, $r_{i,t}^k$, as:

$$r_{i,t}^k = \frac{\bar{R}_{i,t}^k n_{i,t} - \bar{R}_{i,t-1}^k n_{i,t-1}}{n_{i,t} - n_{i,t-1}} \quad (1)$$

The accumulation of negative feedback and the decision to exit the platform can influence each other. On the one hand, the benefit of exerting effort decreases when sellers are close to exiting the platform and can anticipate their departure. These sellers may choose to “milk” their reputation and “shirk” in their final transactions. On the other hand, exit may not be anticipated but actually driven by the accumulation of negative feedback in their last transactions on the platform. In fact, if sellers receive negative feedback, they may face challenges in attracting new customers and could be compelled to leave, as staying on the platform is no longer profitable.

This reverse causality poses a challenge to the causal identification of end-game considerations on sellers' effort. To overcome this issue, we exploit the implementation of the HSO as an exogenous shock in

⁷In Appendix, Figure A1 shows the location of Airbnb listings in the entire county of Los Angeles. Most of the dwellings are located within the City of Los Angeles, followed by Santa Monica, Long Beach, and West Hollywood.

⁸We look at listings that left the platform between July and the beginning of November 2019. *InsideAirbnb* scrapes platform data approximately once per month. In July 2019, scraping occurred on July 8th and 9th. At the beginning of November, the platform was scraped on November 1st and 2nd.

career concerns, identifying hosts who exited after a period of anticipation and for a policy that was, in its design, unrelated to hosts' reputation.

3.3 Host Exits Related to Regulation Implementation

We need to identify listings whose exit from the platform was independent of their hosts' accumulated reputation and could have been anticipated. To do so, we restrict the analysis to short-term listings whose last appearance in our data dates back to the period of HSO implementation in the city of Los Angeles and that appeared on the platform before the regulation was announced in December 2018.⁹ We assume that these listings exited because of the regulation, either because they were ineligible or because the host was unwilling to pay to obtain the license number. Since full enforcement of the regulation began in November 2019, listings that left the platform between July and October 2019 likely did so due to hosts anticipating profile deactivation. This allows us to confidently interpret changes in ratings as a result of strategic decisions by hosts who were well-informed about the upcoming enforcement.¹⁰

Table 1 presents summary statistics for the sub-sample of interest, which consists of 3,405 listings advertised by 2,303 hosts. The average number of total reviews, 62.1, collected over the lifetime of these listings, indicates their established presence with multiple bookings. This reflects our focus on listings that were first listed on Airbnb before December 2018, and thus have been on the platform for at least six months.

On a monthly basis, hosts accommodate approximately three guests (as estimated by the average number of reviews per month) at an average price per night of \$142 USD.¹¹ Notably, guest ratings show minimal variation and consistently maintain a high average near the maximum of 5 stars across all categories (Zervas, Proserpio and Byers, 2021). Half of the hosts in our sample manage at least two listings suggesting the presence of a significant proportion of "professional" hosts on the platform. Most likely, these are individuals who rent out properties that are not their primary residence, deviating from the principle of the sharing economy. In contrast, the other half of the sample comprises observations from single-listing hosts, who may have exited the platform during the implementation of the HSO due to unwillingness to pay for the licensing fee. Finally, we complement the platform data with administrative data from the 2018 Census, taking into account the share of ownership in households. We find that 50% of listings are located in areas with a limited share of owners (less than 20%).

⁹We consider the exit date to be the last appearance date of a listing in our sample. We exclude listings that converted from short-term to long-term rentals to remain on the platform.

¹⁰The HSO announcement was covered in the news, and the new rules were presented by Airbnb to its hosts. See: <https://www.latimes.com/local/lanow/la-me-ln-airbnb-rental-ordinance-20181211-story.html> and <https://www.cnn.com/2018/12/12/los-angeles-passes-regulation-targeting-airbnb-rental-hosts.html>.

¹¹We removed the few outliers (less than 1%) where the number of reviews per snapshot was above 31.

Table 1. Distribution of variables in the analytical sample

	Mean	Median	S.D.	Min	Max	N.
Listing's tot. reviews ($n_{i,t}$)	61.20	37.00	68.60	1.00	706.00	29,694
Listing's n. reviews per month ($n_{i,t} - n_{i,t-1}$)	3.20	3.00	2.80	0.00	31.00	29,694
Listing's price per-night (\$, USD)	142.20	115.00	119.10	10.00	1694.00	29,694
Overall rating (#stars)	4.60	4.80	0.60	1.00	5.00	29,694
Accuracy rating (#stars)	4.70	5.00	0.70	1.00	5.00	29,694
Check-in rating (#stars)	4.80	5.00	0.60	1.00	5.00	29,694
Cleanliness rating (#stars)	4.60	5.00	0.80	1.00	5.00	29,698
Communication rating (#stars)	4.80	5.00	0.60	1.00	5.00	29,694
Location rating (#stars)	4.80	5.00	0.60	1.00	5.00	29,694
Value-for-money rating (#stars)	4.60	5.00	0.80	1.00	5.00	29,694
Host's listings (#number)	5.70	2.00	8.40	1.00	87.00	29,694
Owners in neighborhood (%)	26.10	19.50	22.50	0.00	93.68	29,673

Notes: The table summarizes key statistical moments (mean, median, standard deviation, minimum, maximum value and total number of data points) for the analyzed variables in a sample of 3,405 short-term listings in the City of Los Angeles. This sample includes listings that exited after the implementation of HSO, entered before its adoption, and have a price below \$2,000 USD. For example, the table reports that the average total number of listing's reviews is 61.2. However, half of the observed listings have 37 or fewer reviews. The average value conceals significant variation, with a standard deviation of 68.6, which is larger than the mean value. The range of values spans from a minimum of 1 to a maximum of 706 reviews in a total number of 29,694 listing-snapshot observations.

4 Identification Strategy

For the selected sample of hosts exiting because of the HSO, we compare ratings in effort-related categories before and after the HSO implementation, alongside ratings on location. We select location as a control for our DiD strategy, as it is a listing attribute unaffected by the policy and independent of hosts' effort.

In Figure 2, we plot the evolution of the residual (after controlling for listing fixed effects) average ratings in different categories for each month compared to the month before the policy's implementation. The plot illustrates a significant drop in the average ratings for check-in (blue markers), cleanliness (green markers), and communication (gray markers) in the period after the policy adoption (December 2018) compared to the values in November 2018. Prior to the announcement, the average of effort-related ratings, especially for check-in and communication, remained relatively stable. Interestingly, the average rating for location (black markers) exhibits a stable pattern, showing no decline after the HSO announcement and during policy implementation. To obtain causal estimates of the end-game effect, we estimate the following Equation:

$$r_{ikt} = \beta_1 after_t + \beta_2 CategoryEffort_k \times after_t + \mu_{ik} + \delta_{it} + \phi_{kt} + \varepsilon_{ikt}, \quad (2)$$

where r_{ikt} is the rating for listing i at snapshot t in a specific category k . The indicator $CategoryEffort_k$ equals one if r_{ikt} measures effort (check-in, cleanliness, or communication categories), and zero if it relates

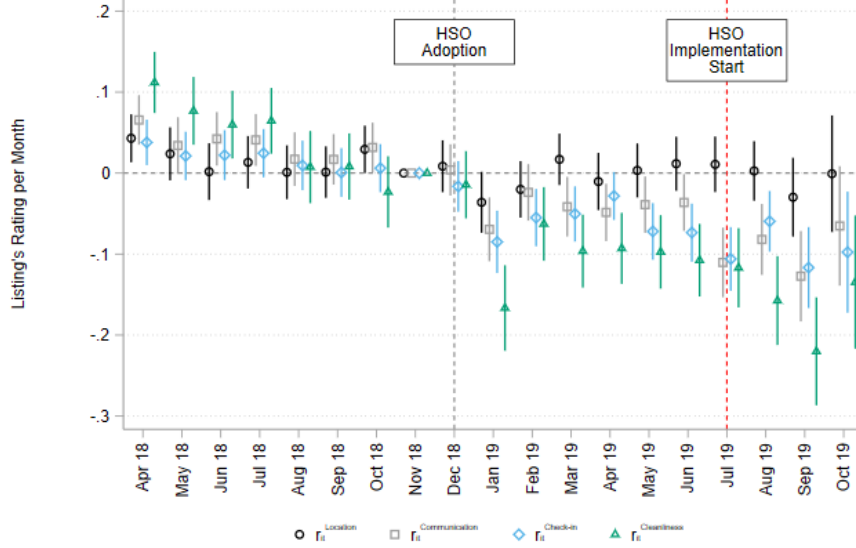


Figure 2. Evolution of ratings per listing overtime and by category

Notes: The figure plots the evolution of the residual (after controlling for listing fixed effects) average ratings in different categories for each month compared to a reference period, the month before the policy's implementation. It includes 95% confidence intervals. The sample consists of 3,405 short-term listings located in the City of Los Angeles that left the platform during the HSO implementation and entered before its adoption.

to the listing's location.¹² The $after_t$ dummy equals one for all snapshots between June 2019 and the completion of the policy implementation, and zero for snapshots from April 2018 to May 2019. From the previous figure, it appears that hosts could have started shirking before June 2019 since they could anticipate exit already at the beginning of 2019. Yet, to be conservative, we decided to use only months after the start of the implementation to ensure we isolate the effect related to their imminent exit. With this approach, we are likely underestimating the effect of the end-game on hosts' effort decisions.

We gradually include a set of fixed effects to isolate the impact of confounders. Listing-category fixed effects are denoted by μ_{ik} , which control for fixed characteristics of the listing within a specific category. Additionally, we include a listing-month fixed effect δ_{it} to control for the same guest's feedback over time. Seasonality may affect guests' evaluation of location differently from other categories. Thus, we control for the vector ϕ_{kt} , which includes category-month and category-year fixed effects. Standard errors are clustered at the listing-category level to account for correlations across snapshots for the same listing. The coefficient β_2 estimates the effect of end-game considerations on effort-related ratings.

Our identification strategy is based on the parallel trend assumption: in absence of the HSO, effort-

¹²In each regression, each listing occurs twice per snapshot. For example, in the regression where we study the evolution of ratings for check-in, a listing i at time t occurs once for the value of r_{ikt} for check-in and once for the value of r_{ikt} for location.

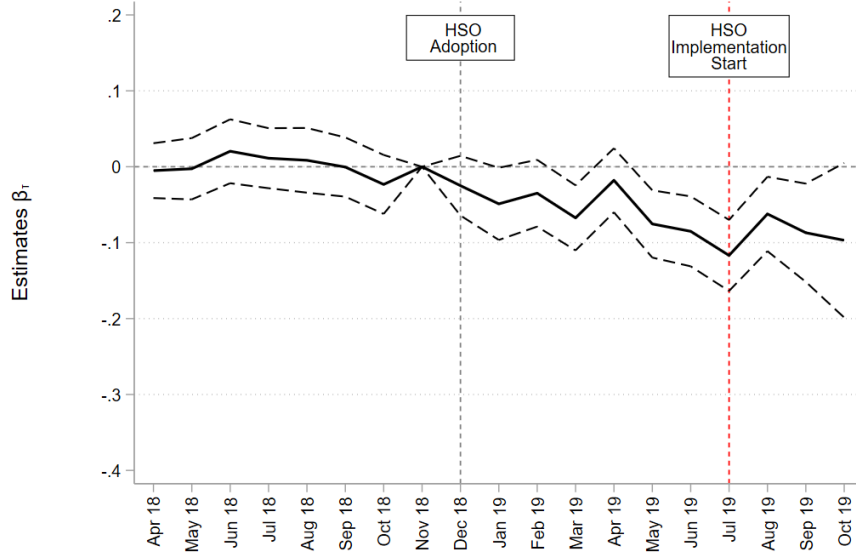


Figure 3. Event Study estimates comparing ratings on check-in with ratings on location

Notes: The figure plots the estimates for coefficients β_τ in Equation 3 along with the 95% confidence intervals for each estimate. The number of observations is 59,388 (each listing is observed twice in each period: once for the effort-related rating, once for location). The reference period, corresponding to the month before HSO approval, is normalized to zero. Standard errors are clustered at a month-listing level and get larger in the last snapshots of data due to the loss of observations when some listings left the platform after July 2019.

related ratings and location ratings would have evolved in a parallel way. In line with the analysis in Figure 2, we adopt an Event Study framework to test this assumption by verifying the absence of significantly divergent pre-trends. We estimate:

$$r_{ikt} = \sum_{\tau=Apr18}^{Oct19} \beta_\tau CategoryEffort_k \times \mathbb{1}(t = \tau) + \delta_{it} + \mu_{ik} + \varepsilon_{ikt} \quad (3)$$

The event study, illustrated by Figure 3 supports our identification strategy and shows no significant pre-trend in the difference between ratings on check-in and ratings on location before the HSO approval and its consequent announcement in the media after December 2018.¹³ The point estimates of coefficients β_τ from Equation 3 start to become significantly negative in periods after the HSO approval, and stayed below zero during its implementation.¹⁴

¹³Example of the HSO approval coverage in the media: <https://www.latimes.com/local/lanow/la-me-ln-airbnb-rental-ordinance-20181211-story.html>

¹⁴Event Study estimates for the other rating effort-related ratings categories are reported in Figures A6 and A7 in Appendix. In addition, Figures A8 and A9 report β_τ from Equation 3 also for ratings on accuracy and value-for-money.

5 Results

In this section, we report the estimates of Equation 2. Tables 2, 3 and 4 display the DiD parameters for the effort-related categories compared to location. In each column we progressively include more fixed effects.¹⁵ The results indicate a negative effect of end-game anticipation on effort. Ratings on check-in and communication decreased up to respectively -0.07 and -0.08 star points after the HSO implementation, equal to about 1.5%-1.7% of the average rating and of 12%-13% of the standard deviation. The interpretation of our estimates' magnitude should take into account the sticky and J-shaped distribution of ratings on Airbnb (Zervas et al., 2021). A back-of-the-envelope calculation allows us to quantify our results in terms of percentile change in the rating distribution. A listing with a rating of 5 stars (maximum value) in the communication rating category, a decline by 0.08 star point, means moving below the 20th percentile of the rating distribution.¹⁶ Such a decline in rating distribution can importantly impact the listing's position in the search ranking on the platform. Moreover, a negative experience also has consequences for the platform, as it may affect client retention rates.

Regarding cleanliness ratings, during the implementation of the HSO, ratings declined by up to -0.11 points, representing nearly a 2.4% decrease from the rating's pre-announcement level and 14% of the standard deviation. However, these findings are not robust to the addition of time fixed effects. One possible reason for the lack of robust results on cleanliness ratings could be the outsourcing of cleaning services to professionals by hosts. When a contract exists between hosts and cleaning professionals, it becomes less easy to adjust the level of effort, making strategic changes in cleanliness ratings less straightforward. In addition, fluctuations in cleanliness ratings could be influenced by external factors unrelated to HSO implementation, such as the cleanliness fee, which we do not include in our analysis.

We now present the results of a heterogeneity analysis aimed at examining the moderating effect of various host and listing characteristics on the impact of end-game considerations on effort decisions.

To ensure the validity of our results, we focus on characteristics related to the regulatory requirements, specifically examining the situation of those hosts who likely exited because their listings were deemed ineligible. The regulation stipulated that only primary residences could be rented for STR purposes. This implies that the target seller consisted mostly of "professional" hosts.

We assume that multi-listing hosts take more advantage of their reputation for two reasons. First, since they are the main target of the HSO, they can be sure that they will have to leave the platform, whereas non-

¹⁵Coefficients tables for the other rating categories can be found in Tables A1 and A2 in Appendix.

¹⁶See Figures A2, A4 and A3 in Appendix for a visualization of the cumulative distribution of the ratings about host's effort.

Table 2. DiD estimates for the comparison of ratings on check-in with ratings on location

	(1) r_{ikt}	(2) r_{ikt}	(3) r_{ikt}	(4) r_{ikt}
Effort category (check-in)	0.039*** (0.008)	0.000 (.)	0.000 (.)	0.000 (.)
After June 19	0.004 (0.008)	-0.000 (0.008)	0.000 (.)	0.000 (.)
Effort category (check-in) \times After June 19	-0.073*** (0.013)	-0.067*** (0.013)	-0.067*** (0.012)	-0.055*** (0.020)
Listing-Category FE		✓	✓	✓
Listing-Month FE			✓	✓
Month-Category FE				✓
Year-Category FE				✓
Standard Errors Clustering Level	listing-category	listing-category	listing-category	listing-category
R^2	0.002	0.276	0.692	0.693
Number of observations	59,388	59,388	59,388	59,388

Notes: The table reports the coefficients estimates from Equation 2 when the effort-related category under study (compared to location) is check-in. Standard errors are clustered at listing-category level and reported in parentheses. In column (2) we include listing-category fixed effect, in column (3) we add the listing-month fixed effect and in column (4) also the year-category fixed effect. In all the specifications the coefficient for the interaction of effort category and after June 2019 is negative and significant at 1% level. The magnitude of the end-game effect (coefficient of the interaction) range between -0.055 to -0.073 star points. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 3. DiD estimates for the comparison of ratings on communication with ratings on location

	(1) r_{ikt}	(2) r_{ikt}	(3) r_{ikt}	(4) r_{ikt}
Effort category (communication)	0.041*** (0.008)	0.000 (.)	0.000 (.)	0.000 (.)
After June 19	0.004 (0.008)	-0.000 (0.008)	0.000 (.)	0.000 (.)
Effort category (communication) \times After June 19	-0.078*** (0.013)	-0.075*** (0.013)	-0.075*** (0.012)	-0.052*** (0.020)
Listing-Category FE		✓	✓	✓
Listing-Month FE			✓	✓
Month-Category FE				✓
Year-Category FE				✓
Standard Errors Clustering Level	listing-category	listing-category	listing-category	listing-category
R^2	0.002	0.269	0.687	0.687
Number of observations	59,388	59,388	59,388	59,388

Notes: The table reports the coefficients estimates from Equation 2 when the effort-related category under study (compared to location) is communication. Standard errors are clustered at listing-category level and reported in parentheses. In column (2) we include listing-category fixed effect, in column (3) we add the listing-month fixed effect and in column (4) also the year-category fixed effect. In all the specifications the coefficient for the interaction of effort category and after June 2019 is negative and significant at 1% level. The magnitude of the end-game effect (coefficient of the interaction) range between -0.052 to -0.078 star points. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 4. DiD estimates for the comparison of ratings on cleanliness with ratings on location

	(1) r_{ikt}	(2) r_{ikt}	(3) r_{ikt}	(4) r_{ikt}
Effort category (cleanliness)	-0.167*** (0.009)	0.000 (.)	0.000 (.)	0.000 (.)
After June 19	0.004 (0.008)	-0.000 (0.008)	0.000 (.)	0.000 (.)
Effort category (cleanliness) \times After June 19	-0.097*** [0.014]	-0.107*** [0.014]	-0.107*** [0.013]	-0.019 [0.024]
Listing-Category FE		✓	✓	✓
Listing-Month FE			✓	✓
Month-Category FE				✓
Year-Category FE				✓
Clustering Level	listing-category	listing-category	listing-category	listing-category
R^2	0.020	0.275	0.674	0.675
Number of observations	59,388	59,388	59,388	59,388

Notes: The table reports the coefficients estimates from Equation 2 when the effort-related category under study (compared to location) is cleanliness. Standard errors are clustered at listing-category level and reported in parentheses. In column (2) we include listing-category fixed effect, in column (3) we add the listing-month fixed effect and in column (4) also the year-category fixed effect. In columns (1), (2) and (3) the coefficient for the interaction of effort category and after June 2019 is negative and significant at 1% level. Including the year fixed effect cancels the significance of the effect; however, the coefficient still remains negative. The magnitude of the end-game effect (coefficient of the interaction) range between -0.019 to -0.107 star points. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.0$.

professionals can still consider whether to stay by paying the registration fee. Second, multi-listing hosts are likely more profit-maximization driven than non-professionals, who may be find intrinsic value in participating in the sharing economy. Therefore, single-listing hosts are expected to show a smaller decline in effort even when anticipating an exit.

To approximate the likelihood of being ineligible, we identify professional hosts by the number of listings they rented on Airbnb prior to the implementation of the HSO. We assume that having more than one listing increases the likelihood of being a professional, which is defined as renting a home on a short-term basis that is not the host's primary residence.¹⁷ Then we look at the homeownership rate in the neighborhood where the listing is located. If the share is low and there is a higher prevalence of rental housing, it is more likely that the listing is not the owner's primary residence.

In line with our expectations, our findings in Figure 4 indicate that listings rented by multiple-listing hosts and those located in areas with fewer property owners exhibited a more pronounced decline in effort-related ratings during the HSO implementation. While the difference between groups in sub-figure (a) is not particularly statistically significant, in sub-figure (b), the coefficient for listings located in the lowest and third quartiles of the distribution of ownership share in the neighborhood is statistically significant.

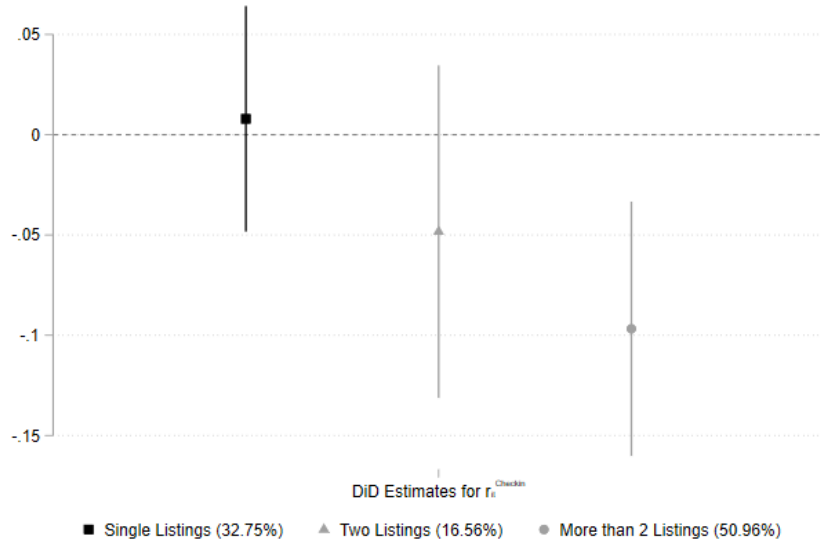
¹⁷It is possible for two listings for the same host to be in the same house or apartment. In this study, we do not distinguish between these two possibilities.

These results suggest a greater focus on profit maximization among “professional” hosts. Conversely, the intrinsic value associated with participating in the sharing economy for non-professional individuals could partially mitigate the moral hazard at the end of hosts’ careers.¹⁸

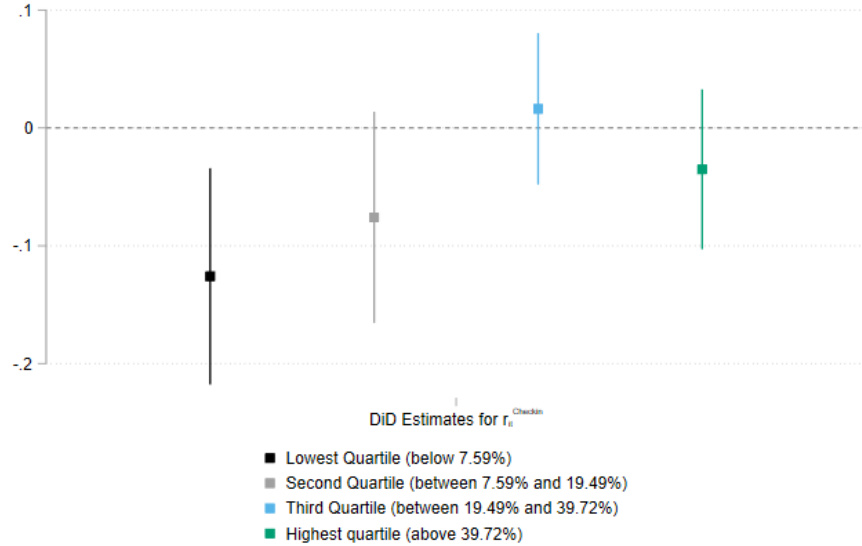
Now, we move to investigate the moderating influence of hosts’ previous reputation. We start by examining the number of reviews collected by each listing before the HSO implementation. There is a tension between two possible effects at play. On one side, the simple average of ratings in the Airbnb reputation system gives an advantage to listings with a more solid reputation, namely those with more reviews. A low rating following a decline in a host’s effort takes longer to affect the average score of these listings. On the other side, listings with more reviews could simply be the better quality ones, those with lower costs of effort and therefore those behaving well anyway. Our results in Figures A14, A15 and A16 in Appendix show mixed results. For the check-in category, the second explanation works better: hosts learn how the check-in procedure works and those more experienced (with more reviews) are more likely to keep their effort up until the end as their cost of effort is low. On the other hand, for the communication category, we observe that listings with more reviews are more likely to reduce effort compared to those with a less consolidated history of reviews in line with the first potential explanation. However, none of the heterogeneity tests show a particularly significant difference coefficient across different groups. Possibly, this is due to the selection in our sample of listings that have been on the platform at least 6 months before the regulation implementation.

We also examine the effect of previous ratings. As mentioned before, hosts with a very high overall ratings or ratings associated with location may continue to attract guests even if their ratings start to decline. Thus, these hosts have larger margins to leverage accumulated ratings to continue their business. At the same time, hosts may have accumulated these high ratings consistently since they are of high-quality and they may not be strategic in their effort decision. To investigate this, we conducted a heterogeneity analysis based on the ratings value of the listings in the overall and location categories before the HSO announcement. However, as depicted in Figures A20 to A22 in Appendix we do not find significant difference in the estimates of β_2 (from Equation 2) given the cumulative average rating of listings before the HSO approval being at or below the maximum of 5 stars. Despite not being significant, listings with lower overall ratings and location ratings before the HSO announcement seem to decline their effort more, aligning with the idea of listings with higher ratings being of higher quality with lower costs of effort. Our results suggest a gen-

¹⁸Figures A10 and A11 in Appendix show the heterogeneity of the end-game effect by the host’s number of listings on Airbnb in Los Angeles before the HSO approval, respectively on communication and cleanliness rating. Figures A12 and A13 in Appendix illustrate show the heterogeneity of the end-game effect by the share of owners in the listing’s neighbourhood, respectively on communication and cleanliness rating.



(a) Heterogeneity analysis by host's number of listings on Airbnb before the HSO approval



(b) Heterogeneity analysis by share of owners in the listing's neighbourhood

Figure 4. Heterogeneity analysis on the rating on check-in by eligibility conditions

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on check-in. The number of observations is 59,388. The heterogeneity analysis is performed by splitting the sample based on (a) the host's number of listings (in the county of Los Angeles) on the platform before the HSO approval and (b) the share of owners in the listing's neighborhood according to the 2018 census. Sub-Figure (a) illustrates that, while not particularly statistically significant, the end-game effect is more prominent and statistically different from zero among listings hosted by individuals with more than two listings on Airbnb before HSO approval (50.96% of the sample). In Sub-Figure (b), it is observed that listings located in neighborhoods with a share of owners below 7.6% (lower quartile) exhibit a more pronounced decline in effort compared to the other three quartiles. The difference is significant, particularly in comparison with the upper half of the owners' share distribution.

erally uniform exploitation of their reputation by hosts, regardless of their previous ratings. Again, this is likely driven by the fact that we are only looking into a specific group of listings, which have been on the platform for quite some time and have little variation in the distribution of quality.

5.1 Robustness Checks and Placebos

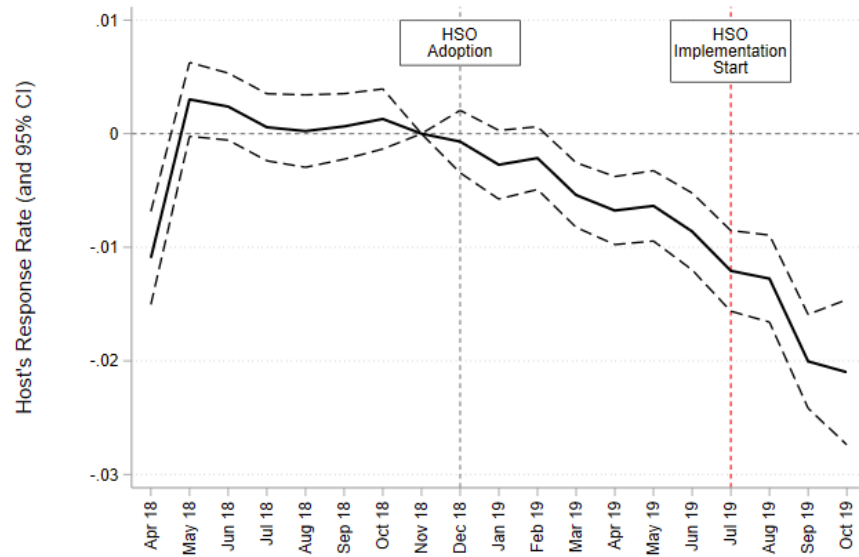
In addition to our main results, we perform a set of robustness checks and placebo analysis. First, we show that our results are robust to different clustering of the standard errors and dataset restrictions (Table A3 and Table A4 in Appendix). Additionally, we look at the evolution of other measures of effort such as the host’s response rate to guest messages and requests and the times it takes for them to answer. The descriptive results presented in Figure 5 resonates with the main findings of an effort decline. In last hosts’ transactions, the response rate decreased among sellers who left the platform in City of Los Angeles due to the regulation, while their time to answer requests and messages increased.

Finally, in Figure 6, we perform a placebo test looking at listings located in other cities of the county, thus not affected by the HSO but which showed similar characteristics and left the platform in the same period. As expected, we find no significant changes in the difference between the effort-related ratings and ratings on location.

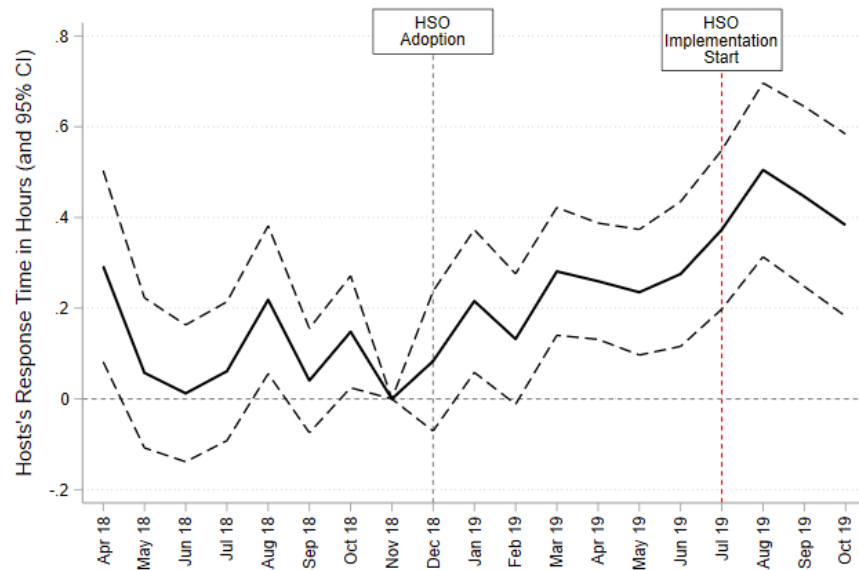
5.2 Additional Specification

We finally complement our approach with a more traditional DiD. We take advantage of the geographical variation in the HSO requirement and the richness of our data which covers listings in the whole county of Los Angeles including cities outside the administrative borders of the City of Los Angeles. The targeted sample includes listings who exited the platform during the policy implementation. The treated group corresponds with listings located in the City of Los Angeles. The control group includes all the remaining cities. This division is based on the idea that hosts who are not eligible for registration, or unwilling to register, are more likely to anticipate their unavoidable exit after regulation is enforced. Differently, hosts who are not subject to regulation can still exit the platform but are unlikely to anticipate it so much in advance. Moreover, we focus on listings leaving the platform simultaneously to rule out the risk of spillovers due to decreased competition driven by sudden exit of many listings on the platform.¹⁹

¹⁹Table A5 in Appendix summarizes the distribution of the variables used in the analysis, comparing their mean in the treated and the control group.



(a) Evolution of hosts' response rate



(b) Evolution of hosts' response time

Figure 5. Evolution of alternative measures of hosts' effort

Notes: The figure plots the evolution of residual average value of host's response rate (a) and response time (measured in hours), measure in hours (b) compared to reference period (the month before policy implementation). Vertical bars represent the the 95% confidence interval. The sample is made of 3,405 short-term listings located in the City of Los Angeles that left the platform during the HSO implementation and entered before its adoption. Sub-Figure (a) demonstrates a decline in the average hosts' response rate from the moment the policy is announced compared to November 2018. Conversely, Sub-Figure (b) reveals a statistically significant increase in the average hosts' response time to guest messages after the HSO adoption.

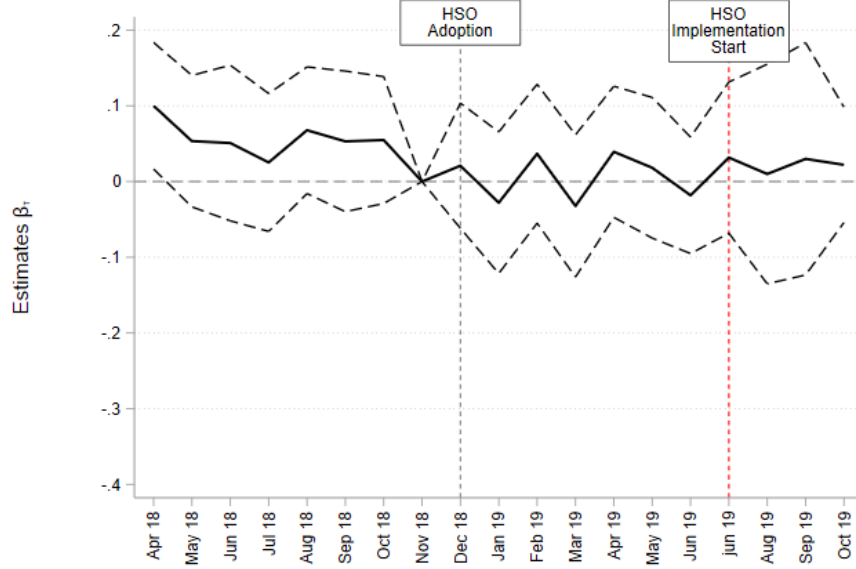


Figure 6. Placebo Event study estimates comparing ratings on check-in with ratings on location in other cities in the county of Los Angeles

Notes: The figure plots the estimates for coefficients β_t in Equation 3 for listings outside of the City of Los Angeles but belonging to the same county and having similar features (STR listings exited during the implementation of HSO, entered before its adoption, and have a price below \$2,000 USD). We show no significant changes in the DiD coefficient after the HSO is announced and during its implementation.

The main equation to estimate is:

$$r_{it}^k = \beta_1 after_t + \beta_2 LAcity_i \times after_t + \phi X_{it} + \mu_i + \delta'_t + \gamma Trend_t + \varepsilon_{it}, \quad (4)$$

where r_{it}^k is the rating for listing i at snapshot t and category k . The indicator $LAcity_i$ identified the treated group, namely listings whose exit could have been anticipated by host. It is equal to one if listing i located in the City of Los Angeles, and it is equal to zero if a listing is located in other cities of the county. As in the main specification, $after_t$ takes value one in snapshots between June 2019 and the completion of the policy implementation. In the analysis, we gradually include controls to isolate the impact of confounders. To account for seasonality, we control for the vector, δ'_t , which includes the set of month and year dummies, and for $Trend_t$, which is a daily linear trend. X_{it} includes the total number of reviews received by each listing since entry on the platform. Listing fixed effects are denoted by μ_i . Standard errors are clustered at a listing level to allow for correlation across snapshots for the same listing.

We assume that, in absence of HSO, the evolution of ratings of exiters among listings located in the City of Los Angeles and those outside the city's border would have been the same. We adopt an event study approach and show the plausibility of this assumption in our context (Figures A23, A24 and A25 in

Table 5. DiD estimates of β_2 from Equation 4

	(1) $r_{Check-in}$	(2) $r_{Cleanliness}$	(3) $r_{Communication}$	(4) $r_{Overall}$	(5) $r_{Accuracy}$	(6) $r_{Location}$	(7) r_{Value}	(8) Price
After June 19	0.028 (0.025)	0.059* (0.033)	0.070*** (0.022)	0.008 (0.027)	0.048 (0.032)	0.054** (0.025)	0.004 (0.035)	0.752 (1.413)
City of LA \times After June 19	-0.041* (0.022)	-0.013 (0.029)	-0.052*** (0.019)	-0.006 (0.023)	-0.030 (0.028)	0.005 (0.021)	0.001 (0.030)	-1.868 (1.423)
Listing-Month Controls (X'_{it})	✓	✓	✓	✓	✓	✓	✓	✓
Time Linear Trend	✓	✓	✓	✓	✓	✓	✓	✓
Time FEs	✓	✓	✓	✓	✓	✓	✓	✓
Listing FE	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.284	0.270	0.279	0.282	0.262	0.289	0.222	0.976
Number of observations	35,441	35,441	35,441	35,441	35,441	35,441	35,441	35,441

Notes: The table reports the coefficients β_2 from Equation 4. Standard errors are clustered at the listing level and are reported in parentheses. Columns (1) to (6) report the results using the ratings for the categories of check-in, cleanliness, communication, overall experience, accuracy, location, and value for money. In Column (8), we use the same specification with the outcome variable being the listings' prices. The coefficient for the interaction of the city of LA and after June 2019 is negative only for two of the three effort-related rating categories (check-in and communication). All other categories, including location, are not affected, as well as prices. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In Table 5, we document a statistically significant negative coefficient for both check-in and communication ratings. Including all the controls, anticipation decreases a rating by around -0.04 points (at a 10% significance level) for check-in and -0.05 (at a 1% significance level) for communication. As for the main specification results, coefficient magnitude should be interpreted given the limited variance of ratings on the platform. Again, the absence of a significant effect on cleanliness ratings may be interpreted as being somehow related to the possibility that these services are often outsourced to external cleaning companies. Meanwhile, the lack of an effect on location ratings confirms the validity of our choice for the control group in the main specification. Finally, while the anticipation of a forthcoming career termination correlates with lower effort exertion, we do not find evidences of moral hazard passing through changes in pricing strategy.

6 Conclusion

In this work, we study how end-game considerations impact sellers' behavior in online marketplaces. Using a panel of listings on Airbnb in the City of Los Angeles, we studied hosts' effort decisions in response to a regulatory shock. The policy generated anticipation of exit among a group of hosts advertising short-term rentals on the platform by introducing additional costs and eligibility criteria for hosts to rent for short-term periods. This led to a significant drop in the number of listings on Airbnb during the policy implementation, from July to November 2019. When the policy was announced, hosts could anticipate if they were going to abandon the platform in the coming months. Accordingly, we focused on listings affected by the regulation

that left the platform during its implementation and compared the evolution of ratings reflecting hosts' effort with ratings on location. We also applied an additional Difference-in-Differences strategy by comparing listings located in the City of Los Angeles with listings located in other cities of the county, where the regulation did not apply.

With both identification strategies, we document a negative and significant impact of end-game considerations on ratings reflecting hosts' effort. When hosts expect few remaining periods on the platform, they tend to shirk, resulting in a decrease in effort-related ratings. Our results suggest that reputation incentives vary across a seller's life on the platform and, specifically, that sellers have less incentive to exert effort when they approach the end of their careers. The specific context of Airbnb reveals that reputation systems relying solely on the simple averaging of ratings over the entire seller's lifespan are inadequate in mitigating moral hazard during sellers' final transactions on digital platforms.

These findings have important implications for the design of reputation systems. Platforms could mitigate the risk of moral hazard in the last transactions of sellers by introducing additional tools, such as performance-pay contracts or giving higher weight to more recent feedback. Measures like censoring ratings after a certain number of reviews may be detrimental, potentially stimulating hidden actions by sellers due to the decreasing marginal benefits of positive ratings after a certain point. Future research could further explore these managerial strategies to improve the effectiveness of reputation systems and ensure consistent service quality throughout a seller's tenure on digital platforms.

References

- Anderson, Michael, and Jeremy Magruder (2012) 'Learning from the crowd: Regression discontinuity estimates of the effects of an online review database.' *The Economic Journal* 122(563), 957–989
- Babić Rosario, Ana, Francesca Sotgiu, Kristine De Valck, and Tammo HA Bijmolt (2016) 'The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors.' *Journal of marketing research* 53(3), 297–318
- Bar-Isaac, Heski, and Steven Tadelis (2008) 'Seller reputation.' *Foundations and Trends® in Microeconomics* 4(4), 273–351
- Bondi, Tommaso (2023) 'Alone, together: A model of social (mis) learning from consumer reviews.' *Available at SSRN 4453685*

- Bondi, Tommaso, Michelangelo Rossi, and Ryan Stevens (2023) ‘The good, the bad and the picky: Reference dependence and the reversal of product ratings.’ *Available at SSRN 4453681*
- Cabral, Luis, and Ali Hortacsu (2010) ‘The dynamics of seller reputation: Evidence from eBay.’ *The Journal of Industrial Economics* 58(1), 54–78
- Carnehl, Christoph, Maximilian Schaefer, André Stenzel, and Kevin Ducbao Tran (2022) ‘Value for money and selection: How pricing affects airbnb ratings.’ *Innocenzo Gasparini Institute for Economic Research Working Paper Series*
- Chen, Yubo, and Jinhong Xie (2008) ‘Online consumer review: Word-of-mouth as a new element of marketing communication mix.’ *Management science* 54(3), 477–491
- Dellarocas, Chrysanthos (2003) ‘The digitization of word of mouth: Promise and challenges of online feedback mechanisms.’ *Management science* 49(10), 1407–1424
- Dellarocas, Chrysanthos, and Charles A. Wood (2008) ‘The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias.’ *Management Science* 54(3), 460–476
- Fan, Ying, Jiandong Ju, and Mo Xiao (2016) ‘Reputation premium and reputation management: Evidence from the largest e-commerce platform in China.’ *International Journal of Industrial Organization* 46, 63–76
- Farronato, Chiara, and Andrey Fradkin (2022) ‘The welfare effects of peer entry: the case of airbnb and the accommodation industry.’ *American Economic Review* 112(6), 1782–1817
- Fradkin, Andrey (2017) ‘Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb.’ *Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb (March 21, 2017)*
- Fradkin, Andrey, Elena Grewal, and David Holtz (2021) ‘Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on airbnb.’ *Marketing Science* 40(6), 1013–1029
- Gibbons, Robert, and Kevin J Murphy (1992) ‘Optimal incentive contracts in the presence of career concerns: Theory and evidence.’ *Journal of Political Economy* 100(3), 468–505
- He, Sherry, Brett Hollenbeck, and Davide Proserpio (2022) ‘The market for fake reviews.’ *Marketing Science* 41(5), 896–921

- Holmström, Bengt (1999) ‘Managerial incentive problems: A dynamic perspective.’ *The review of Economic studies* 66(1), 169–182
- Hui, Xiang, Maryam Saeedi, and Neel Sundaresan (2018) ‘Adverse selection or moral hazard, an empirical study.’ *The Journal of Industrial Economics* 66(3), 610–649
- Hui, Xiang, Maryam Saeedi, Zeqian Shen, and Neel Sundaresan (2016) ‘Reputation and regulations: Evidence from ebay.’ *Management Science* 62(12), 3604–3616
- Klein, Tobias J, Christian Lambertz, and Konrad O Stahl (2016) ‘Market transparency, adverse selection, and moral hazard.’ *Journal of political economy* 124(6), 1677–1713
- Koster, Hans RA, Jos Van Ommeren, and Nicolas Volkhausen (2021) ‘Short-term rentals and the housing market: Quasi-experimental evidence from airbnb in los angeles.’ *Journal of Urban Economics* 124, 103356
- Luca, Michael (2011) ‘Reviews, reputation, and revenue: The case of yelp.com.’ *Harvard Business School NOM Unit Working Paper No. 12-016*
- Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014) ‘Promotional reviews: An empirical investigation of online review manipulation.’ *American Economic Review* 104(8), 2421–55
- Miklós-Thal, Jeanine, and Hannes Ullrich (2016) ‘Career prospects and effort incentives: Evidence from professional soccer.’ *Management Science* 62(6), 1645–1667
- Proserpio, Davide, Wendy Xu, and Georgios Zervas (2018) ‘You get what you give: theory and evidence of reciprocity in the sharing economy.’ *Quantitative marketing and economics* 16(4), 371–407
- Tadelis, Steven (2016) ‘Reputation and feedback systems in online platform markets.’ *Annual review of economics* 8(1), 321–340
- Vana, Prasad, and Anja Lambrecht (2021) ‘The effect of individual online reviews on purchase likelihood.’ *Marketing Science* 40(4), 708–730
- Varga, Marton, and Paulo Albuquerque (2023) ‘The impact of negative reviews on online search and purchase decisions.’ *Journal of Marketing Research* p. 00222437231190874
- Xu, Lei, Tingting Nian, and Luis Cabral (2020) ‘What makes geeks tick? A study of stack overflow careers.’ *Management Science* 66(2), 587–604

Zervas, Georgios, Davide Proserpio, and John W Byers (2021) 'A first look at online reputation on airbnb, where every stay is above average.' *Marketing Letters* 32, 1–16

APPENDIX - Rating Systems and the End-Game Effect: When Reputation Works and When it Doesn't

A Additional Results

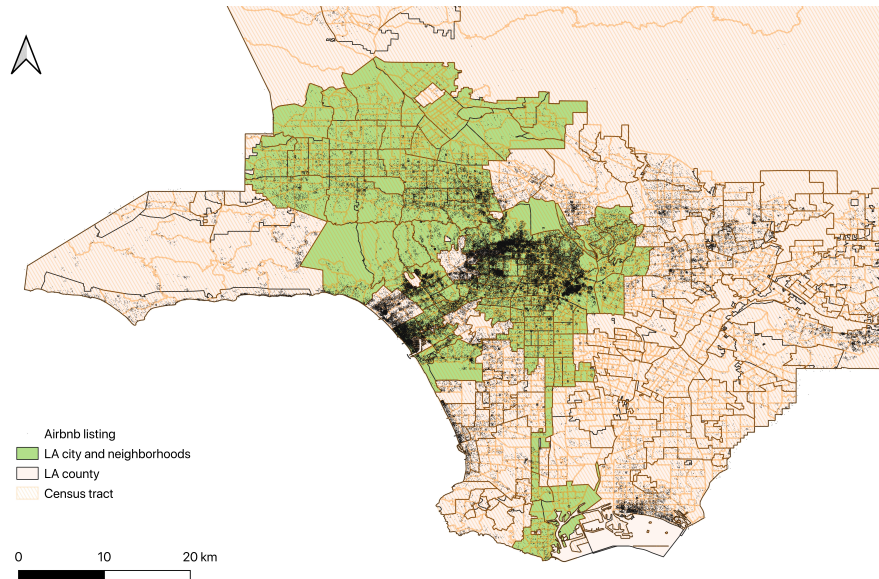


Figure A1. Location of Airbnb listings in the county of Los Angeles

Notes: Authors' own computations based on *InsideAirbnb* data. The figure displays the locations of all listings present in the *InsideAirbnb* dataset. The green area represents the city of Los Angeles, while the pink area between borders encompasses other regions within the county of Los Angeles. The majority of dwellings are situated within the City of Los Angeles, followed by Santa Monica, Long Beach, and West Hollywood.

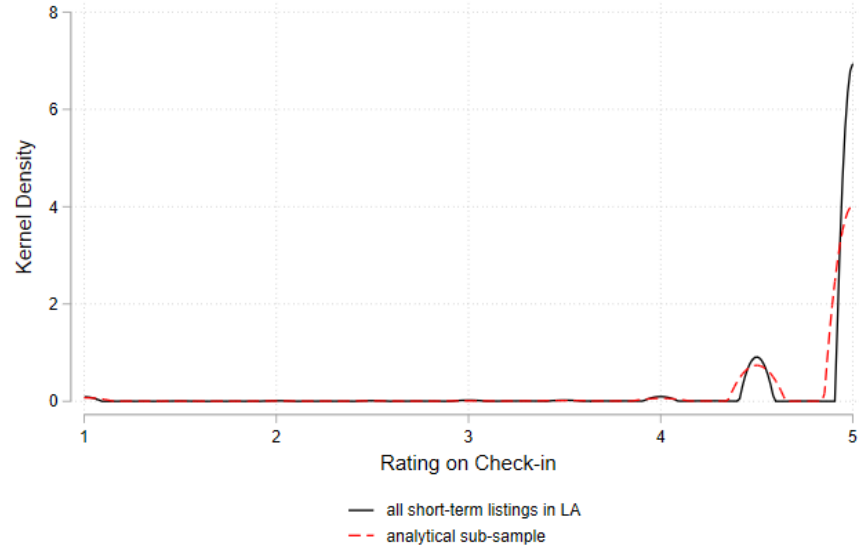


Figure A2. Distribution of the ratings on check-in in the City of Los Angeles

Notes: The figure illustrates the distribution of the ratings on check-in in the City of Los Angeles during the period of our analysis. It confront the overall distribution, with the distribution of ratings only in our analytical sub-sample

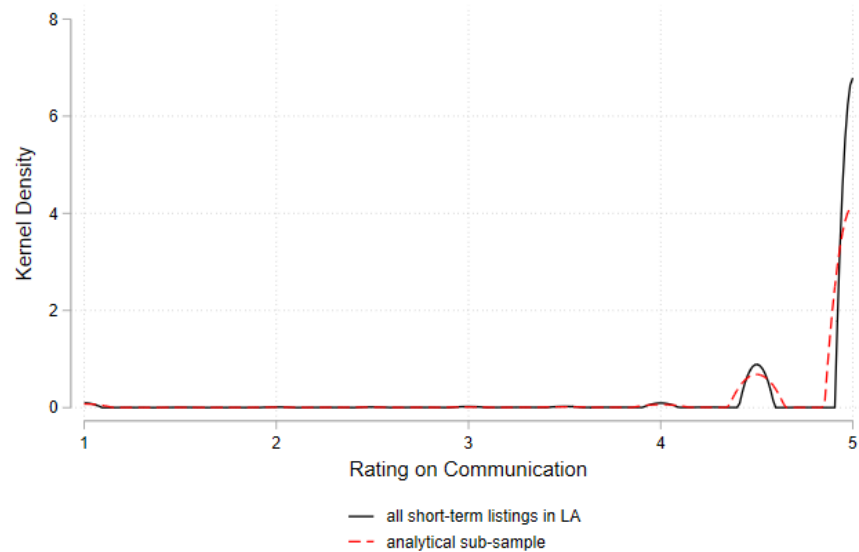


Figure A3. Distribution of the ratings on communication in the City of Los Angeles

Notes: The figure illustrates the distribution of the ratings on communication in the City of Los Angeles during the period of our analysis. It confront the overall distribution, with the distribution of ratings only in our analytical sub-sample

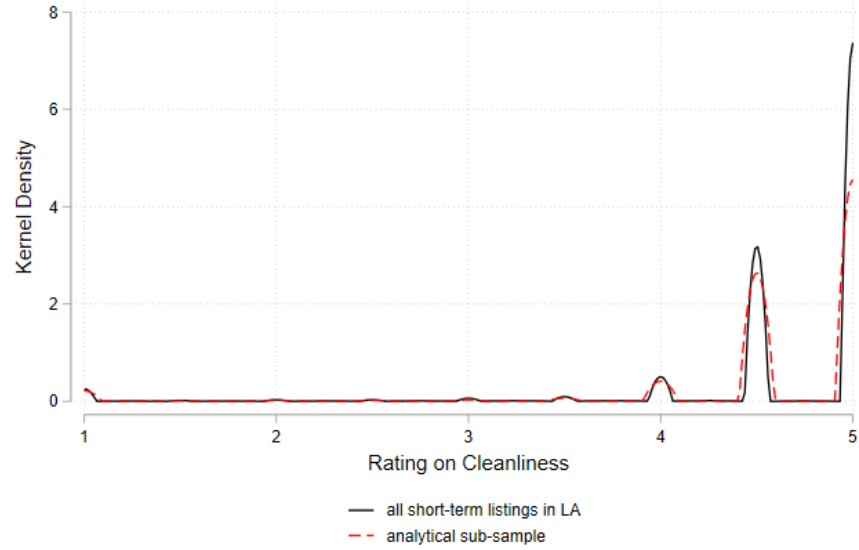


Figure A4. Distribution of the ratings on cleanliness in the City of Los Angeles

Notes: The figure illustrates the distribution of the ratings on cleanliness in the City of Los Angeles during the period of our analysis. It confront the overall distribution, with the distribution of ratings only in our analytical sub-sample

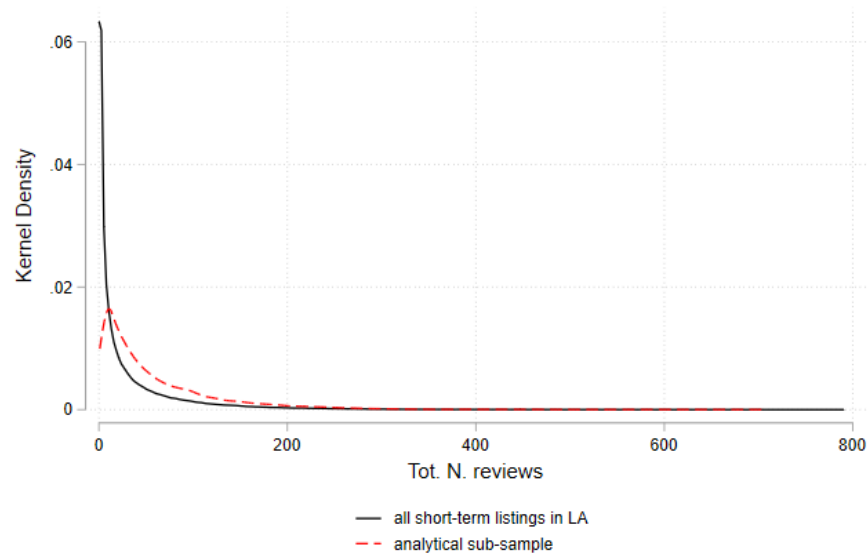


Figure A5. Distribution of listing's number of reviews in the City of Los Angeles

Notes: The figure illustrates the distribution of the listing's number of reviews in the City of Los Angeles during the period of our analysis. It confront the overall distribution, with the distribution of reviews count only in our analytical sub-sample

Table A1. DiD estimates for the comparison of ratings on accuracy with ratings on location

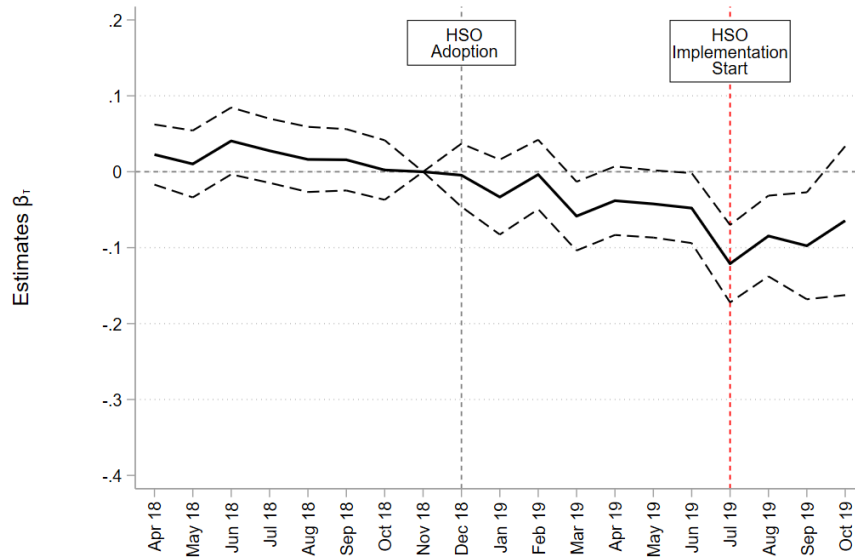
	(1)	(2)	(3)	(4)
	r_{ikt}	r_{ikt}	r_{ikt}	r_{ikt}
Effort category (accuracy)	-0.030*** (0.009)	0.000 (.)	0.000 (.)	0.000 (.)
After June 19	0.004 (0.008)	-0.000 (0.008)	0.000 (.)	0.000 (.)
Effort category (accuracy) \times After June 19	-0.100*** (0.014)	-0.102*** (0.014)	-0.102*** (0.013)	-0.056** (0.022)
Listing-Category FE		✓	✓	✓
Listing-Month FE			✓	✓
Month-Category FE				✓
Year-Category FE				✓
Standard Errors Clustering Level	listing-category	listing-category	listing-category	listing-category
R^2	0.004	0.262	0.683	0.683
Number of observations	59,388	59,388	59,388	59,388

Notes: The table reports the coefficients estimates from Equation 2 when the effort-related category under study (compared to location) is accuracy. Standard errors are clustered at listing-category level and reported in parentheses. In column (2) we include listing-category fixed effect, in column (3) we add the listing-month fixed effect and in column (4) also the year-category fixed effect. In all the specifications the coefficient for the interaction of effort category and after June 2019 is negative and significant at 1% level. The magnitude of the end-game effect (coefficient of the interaction) range between -0.060 to -0.104 star points. Although less directly associated with effort, accuracy ratings also decline after the HSO approval, while there were no significant differences before it. This suggests that accuracy ratings can be interpreted as a measure of host effort. For instance, hosts anticipating an exit may not be spending time updating descriptions of the listing that may include outdated information about appliance and dwelling status or other details.

Table A2. DiD estimates for the comparison of ratings on value-for-money with ratings on location

	(1)	(2)	(3)	(4)
	r_{ikt}	r_{ikt}	r_{ikt}	r_{ikt}
Effort category (value-for-money)	-0.174*** (0.009)	0.000 (.)	0.000 (.)	0.000 (.)
After June 19	0.004 (0.008)	-0.000 (0.008)	0.000 (.)	0.000 (.)
Effort category (value-for-money) \times After June 19	-0.123*** (0.015)	-0.124*** (0.015)	-0.124*** (0.013)	-0.046* (0.024)
Listing-Category FE		✓	✓	✓
Listing-Month FE			✓	✓
Month-Category FE				✓
Year-Category FE				✓
Standard Errors Clustering Level	listing-category	listing-category	listing-category	listing-category
R^2	0.024	0.246	0.673	0.675
Number of observations	59,388	59,388	59,388	59,388

Notes: The table reports the coefficients estimates from Equation 2 when the effort-related category under study (compared to location) is value-for-money. Standard errors are clustered at listing-category level and reported in parentheses. In column (2) we include listing-category fixed effect, in column (3) we add the listing-month fixed effect and in column (4) also the year-category fixed effect. In the first three specifications the coefficient for the interaction of effort category and after June 2019 is negative and significant at 1% level, significance decreases to 10% in column (4). The magnitude of the end-game effect (coefficient of the interaction) range between -0.044 to -0.115 star points. Since Table 3 shows that end-game concerns do not affect price, the decline in value-for-money can be interpreted as a negative evaluation over sellers' effort.

**Figure A6.** Event Study estimates comparing ratings on communication with ratings on location

Notes: The figure plots the estimates for coefficients β_{τ} in Equation 3 along with the 95% confidence intervals for each estimate. Number of observations is 60,880 (each listing is observed twice in each period: once for the effort-related rating, once for location). The reference period, corresponding to the month before HSO approval, is normalized to zero. Standard errors are clustered at a month-listing level. It is important to note that standard errors increased in the last snapshots of data due to the loss of observations when some listings left the platform after July 2019.



Figure A7. Event Study estimates comparing ratings on cleanliness with ratings on location

Notes: The figure plots the estimates for coefficients β_τ in Equation 3 along with the 95% confidence intervals for each estimate. Number of observations is 60,880 (each listing is observed twice in each period: once for the effort-related rating, once for location). The reference period, corresponding to the month before HSO approval, is normalized to zero. Standard errors are clustered at a month-listing level. It is important to note that standard errors increased in the last snapshots of data due to the loss of observations when some listings left the platform after July 2019.

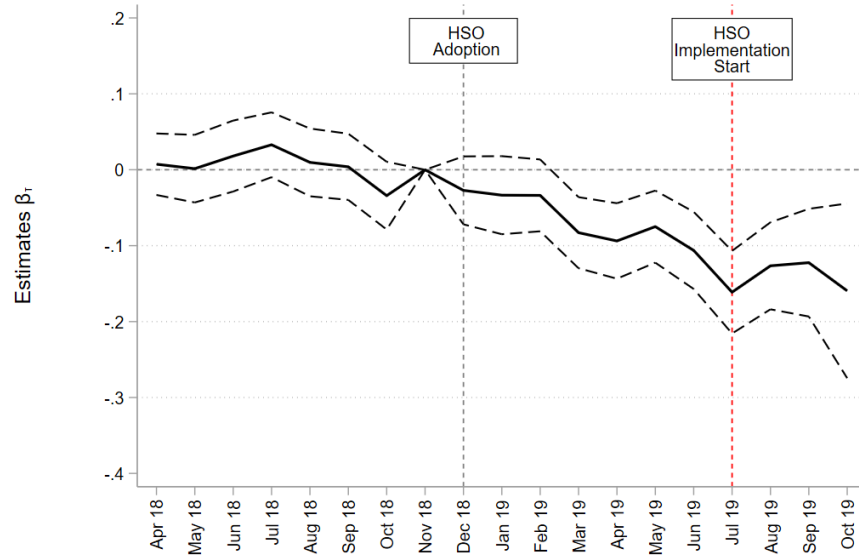


Figure A8. Event Study estimates comparing ratings on accuracy with ratings on location

Notes: The figure plots the estimates for coefficients β_τ in Equation 3 along with the 95% confidence intervals for each estimate. Number of observations is 60,880 (each listing is observed twice in each period: once for the effort-related rating, once for location). The reference period, corresponding to the month before HSO approval, is normalized to zero. Standard errors are clustered at a month-listing level. It is important to note that standard errors increased in the last snapshots of data due to the loss of observations when some listings left the platform after July 2019.

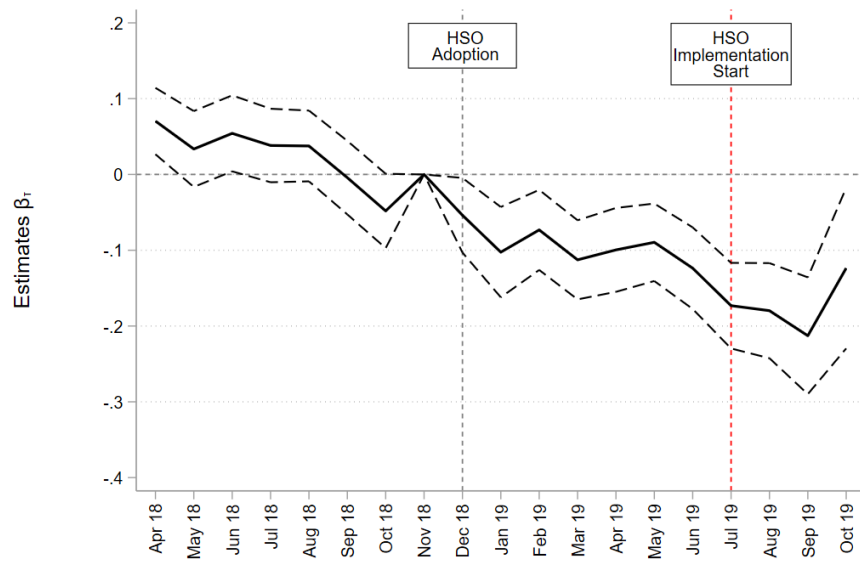


Figure A9. Event Study estimates comparing ratings on value-for-money with ratings on location

Notes: The figure plots the estimates for coefficients β_τ in Equation 3 along with the 95% confidence intervals for each estimate. Number of observations is 60,880 (each listing is observed twice in each period: once for the effort-related rating, once for location). The reference period, corresponding to the month before HSO approval, is normalized to zero. Standard errors are clustered at a month-listing level. It is important to note that standard errors increased in the last snapshots of data due to the loss of observations when some listings left the platform after July 2019.

Additional Robustness and Heterogeneity Analysis

Table A3. DiD estimates (on check-in ratings) robustness to different specifications and sampling

	(1)	(2)	(3)	(4)	(5)	(6)
	r_{ikt}	r_{ikt}	r_{ikt}	r_{ikt}	r_{ikt}	r_{ikt}
Effort category (check-in)	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
After June 19	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Effort category (check-in) \times After June 19	-0.055***	-0.055***	-0.055***	-0.056***	-0.055***	-0.051**
	(0.020)	(0.018)	(0.021)	(0.021)	(0.020)	(0.022)
Listing-Category FE	✓	✓	✓	✓	✓	✓
Listing-Month FE	✓	✓	✓	✓	✓	✓
Month-Category FE	✓	✓	✓	✓	✓	✓
Year-Category FE	✓	✓	✓	✓	✓	✓
Sample				price<500	price<1000	distance \geq 0.003
Standard Errors Clustering Level	listing-category	zip-code	ct10	listing-category	listing-category	listing-category
R^2	0.693	0.693	0.693	0.691	0.692	0.692
Number of observations	59,388	58,860	59,388	58,458	59,252	51,164

* p<0.1, ** p<0.05, *** p<0.01

Notes: The table tests the robustness of the coefficients on check-in ratings estimated from Equation 2 (column (1)) to different clustering levels of the standard errors (column (2) and column (3) and to different samples (column (4), column (5) and column (6)). Standard errors are reported in parentheses.

Table A4. DiD estimates (on communication ratings) robustness to different specifications and sampling

	(1)	(2)	(3)	(4)	(5)	(6)
	r_{ikt}	r_{ikt}	r_{ikt}	r_{ikt}	r_{ikt}	r_{ikt}
Effort category (communication)	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
After June 19	0.000	0.000	0.000	0.000	0.000	0.000
	(.)	(.)	(.)	(.)	(.)	(.)
Effort category (communication) \times After June 19	-0.052***	-0.054**	-0.052***	-0.051**	-0.051**	-0.057***
	(0.020)	(0.021)	(0.018)	(0.020)	(0.020)	(0.022)
Listing-Category FE	✓	✓	✓	✓	✓	✓
Listing-Month FE	✓	✓	✓	✓	✓	✓
Month-Category FE	✓	✓	✓	✓	✓	✓
Year-Category FE	✓	✓	✓	✓	✓	✓
Sample				price<500	price<1000	distance \geq 0.003
Standard Errors Clustering Level	listing-category	zip-code	ct10	listing-category	listing-category	listing-category
R^2	0.687	0.687	0.687	0.687	0.687	0.688
Number of observations	59,388	58,860	59,388	58,458	59,252	51,164

* p<0.1, ** p<0.05, *** p<0.01

Notes: The table tests the robustness of the coefficients on communication ratings estimated from Equation 2 (column (1)) to different clustering levels of the standard errors (column (2) and column (3) and to different samples (column (4), column (5) and column (6)). Standard errors are reported in parentheses.

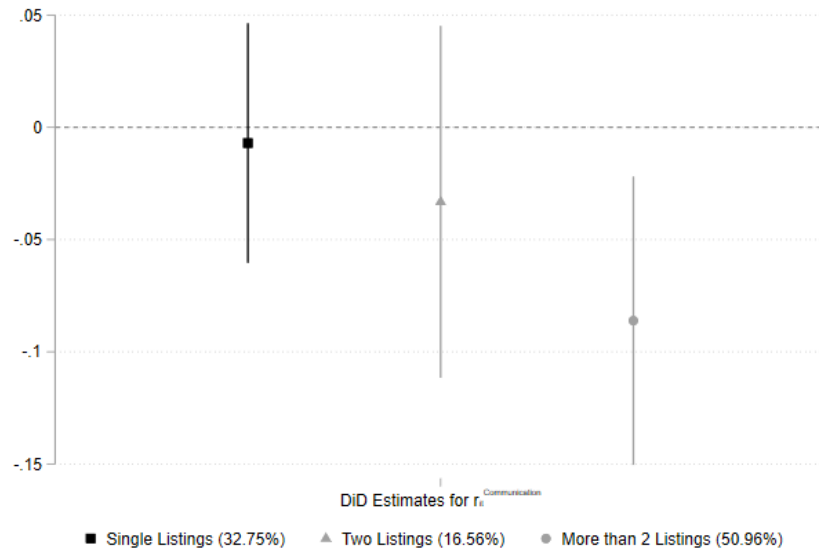


Figure A10. Heterogeneity analysis on ratings on communication by host's number of listings on Airbnb before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on communication and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the host's number of listings on the platform before the HSO approval.

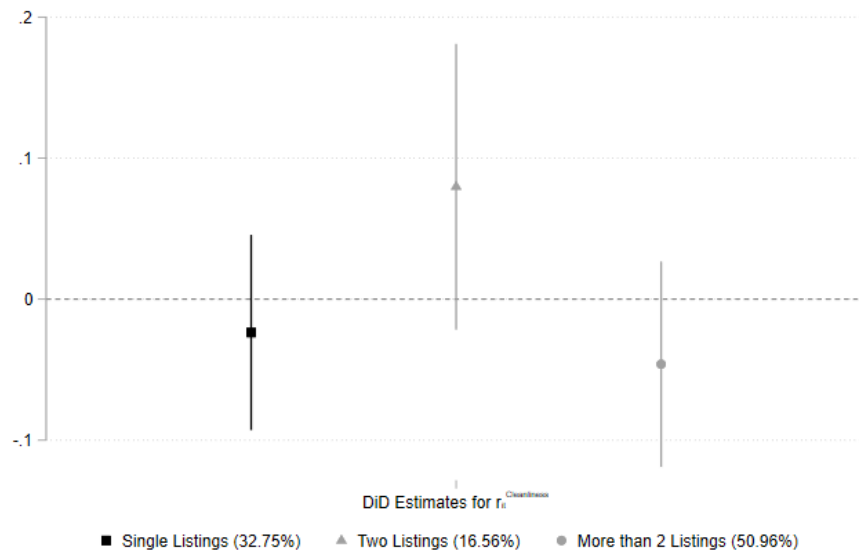


Figure A11. Heterogeneity analysis on ratings on cleanliness by host's number of listings on Airbnb before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on cleanliness and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the host's number of listings on the platform before the HSO approval.

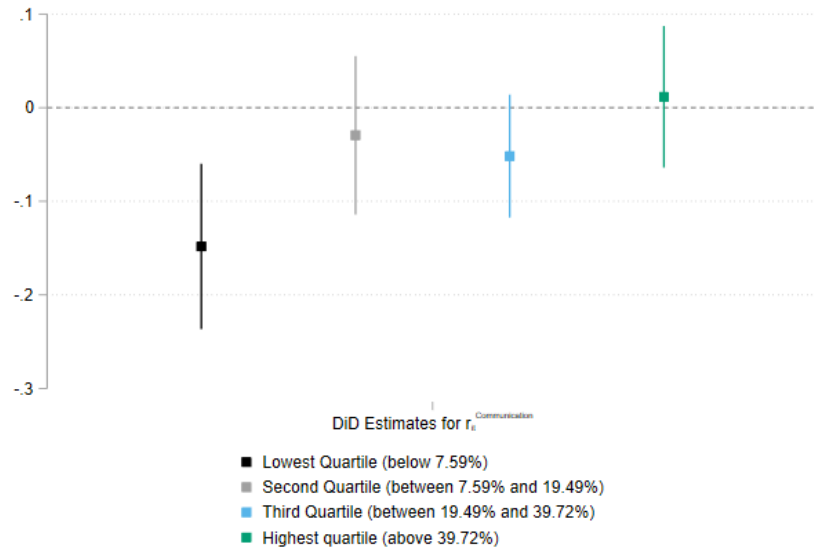


Figure A12. Heterogeneity analysis on ratings on communication by share of owners in the listing's neighbourhood

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on communication and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the share of owners in listing's neighbourhood according to the 2018 census.

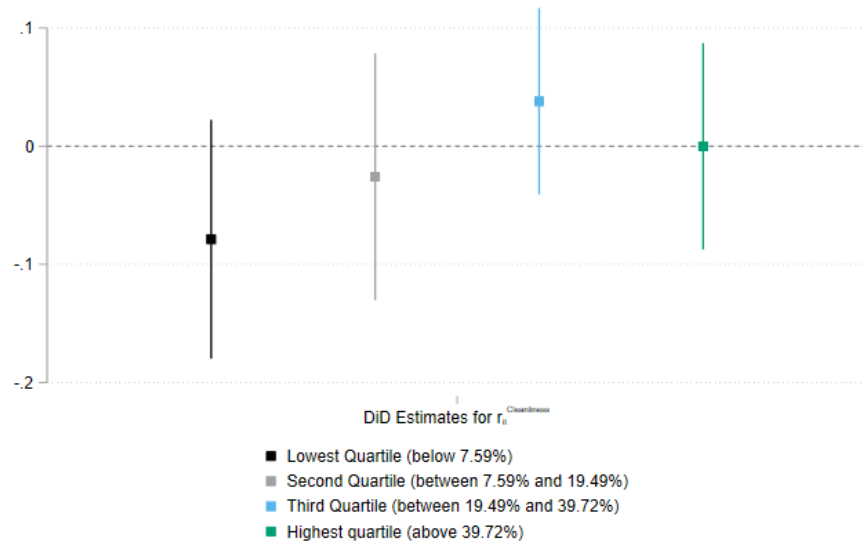


Figure A13. Heterogeneity analysis on ratings on cleanliness by share of owners in the listing's neighbourhood

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on cleanliness and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the share of owners in listing's neighbourhood according to the 2018 census data.

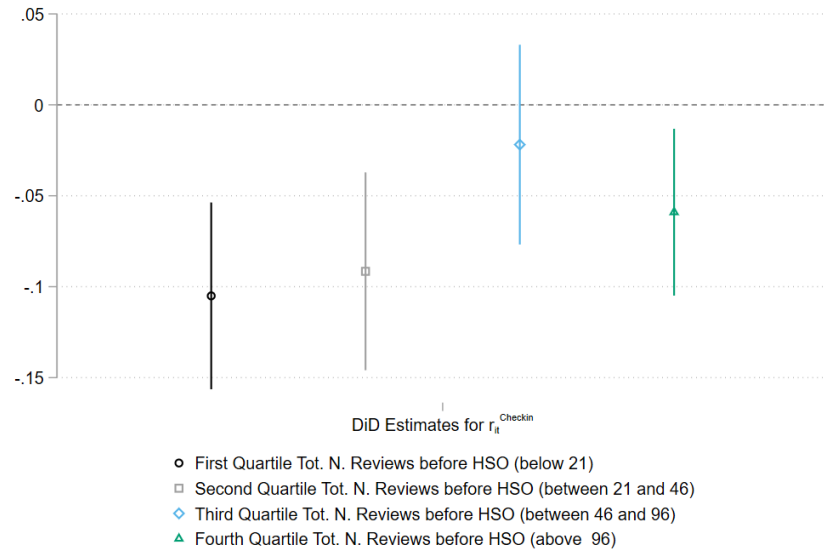


Figure A14. Heterogeneity analysis on ratings on check-in by total number reviews before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on check-in and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the quarter of the distribution of the total number of reviews before the HSO approval.

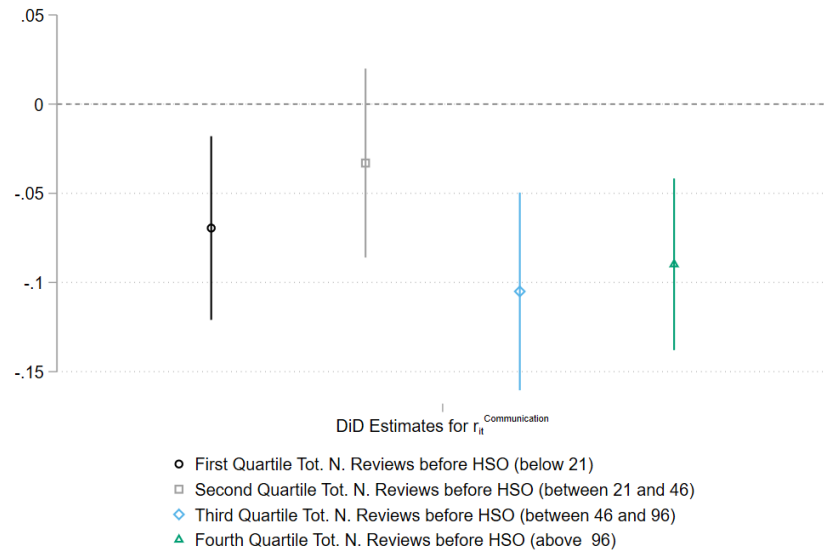


Figure A15. Heterogeneity analysis on ratings on communication by total number reviews before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on communication and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the quarter of the distribution of the total number of reviews before the HSO approval.

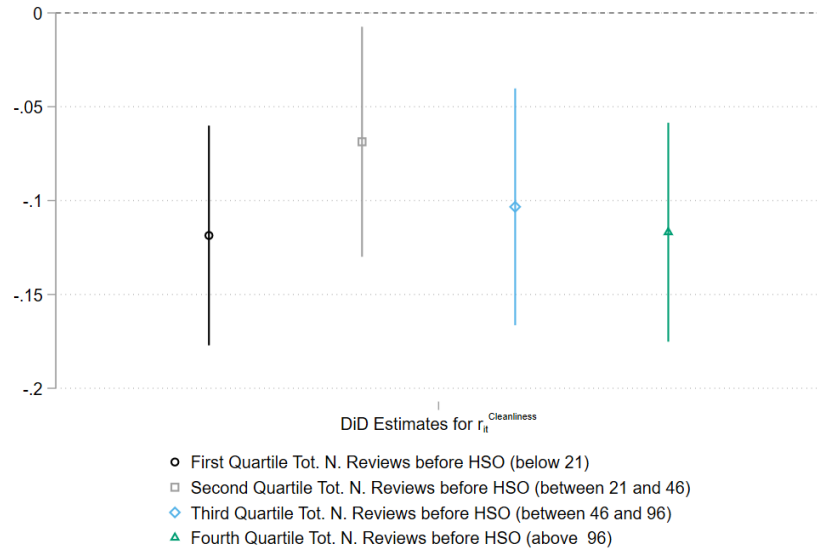


Figure A16. Heterogeneity analysis on ratings on cleanliness by total number reviews before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on cleanliness and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the quarter of the distribution of the total number of reviews before the HSO approval.

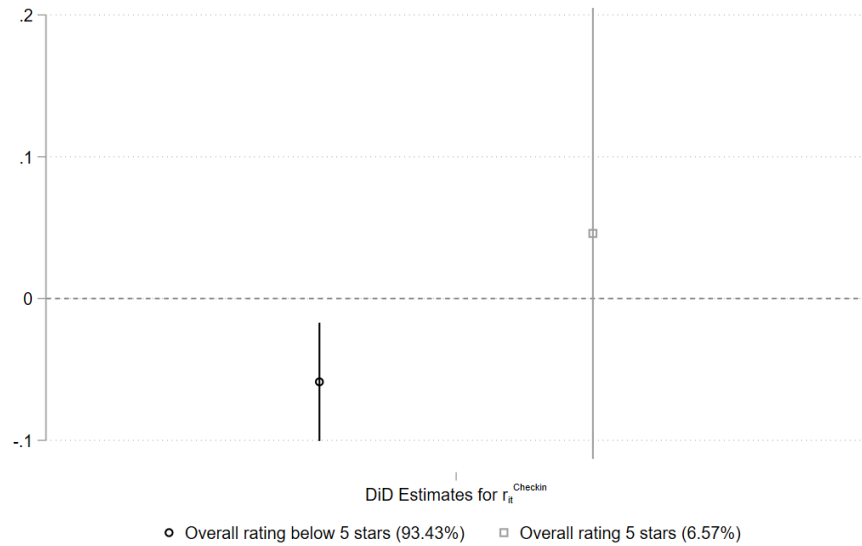


Figure A17. Heterogeneity analysis on ratings on check-in by listing's overall rating before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on check-in and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the value of $\bar{R}_{it}^{overall}$ before December 2018.

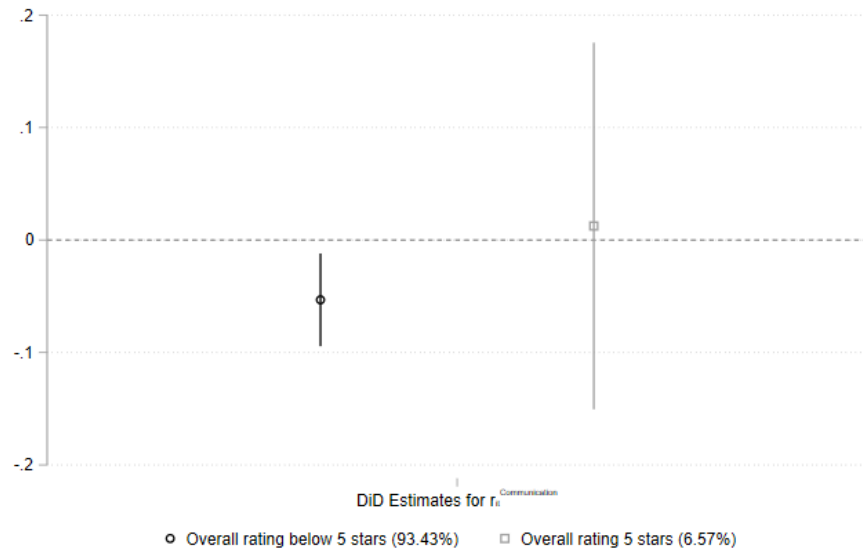


Figure A18. Heterogeneity analysis on ratings on communication by listing's overall rating before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on communication and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the value of $\bar{R}_{it}^{overall}$ before December 2018.

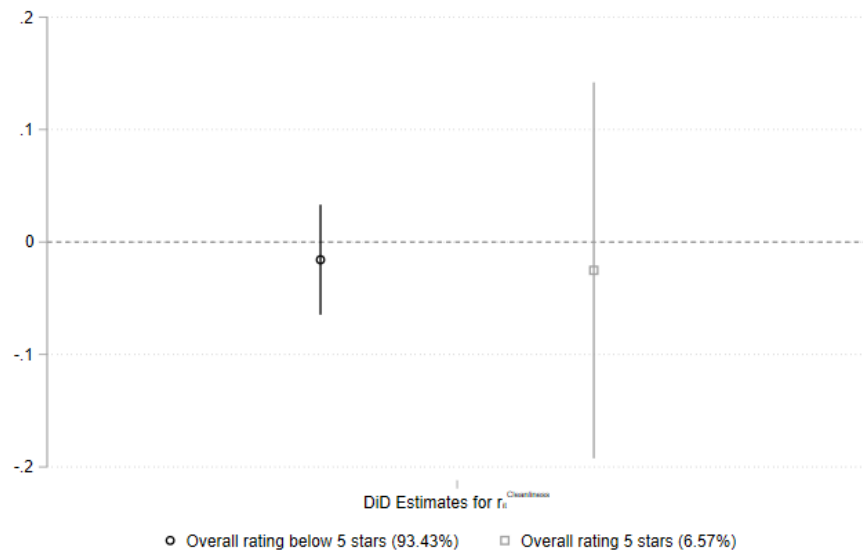


Figure A19. Heterogeneity analysis on ratings on cleanliness by listing's overall rating before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on cleanliness and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the value of $\bar{R}_{it}^{overall}$ before December 2018.

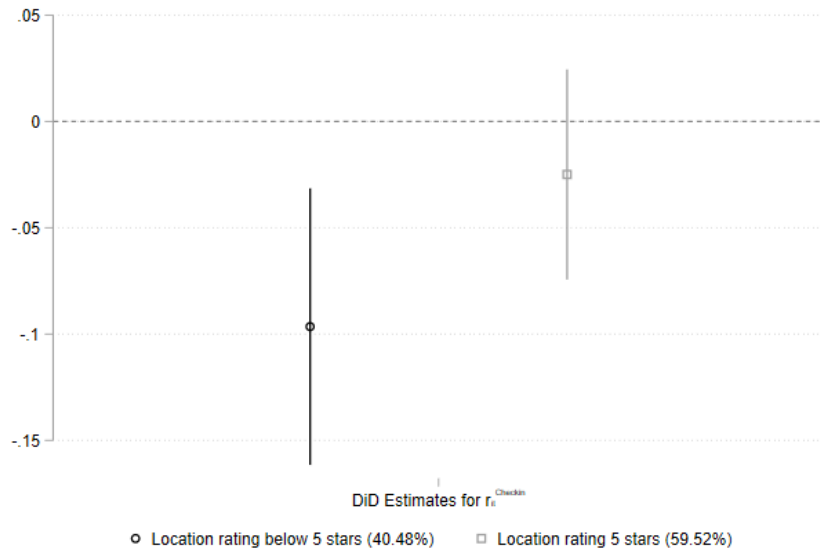


Figure A20. Heterogeneity analysis on ratings on check-in by listing's rating on location before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on check-in and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the value of $\bar{R}_i^{location}$ before December 2018.

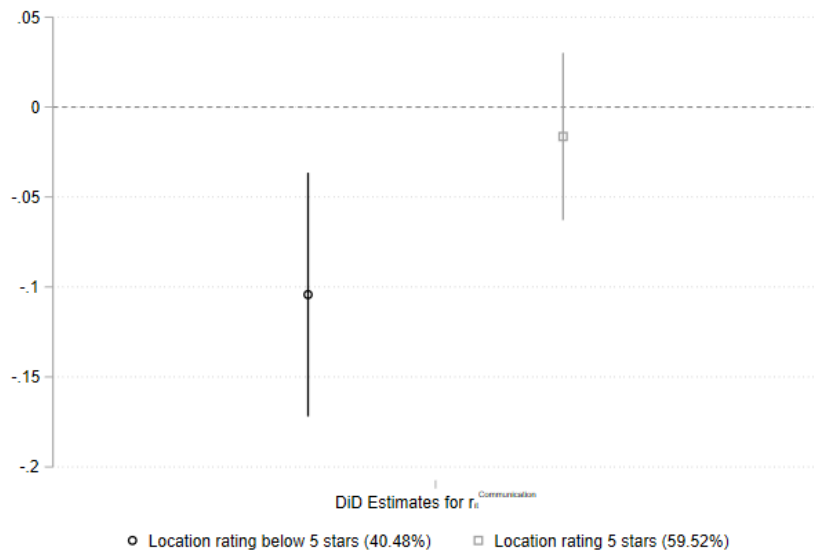


Figure A21. Heterogeneity analysis on ratings on communication by listing's location rating before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on communication and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the value of $\bar{R}_i^{location}$ before December 2018.

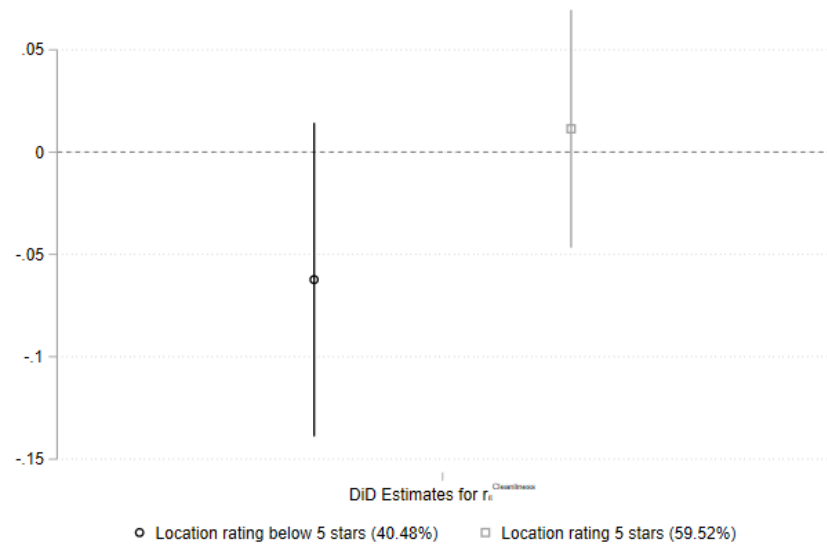


Figure A22. Heterogeneity analysis on ratings on cleanliness by listing's location rating before the HSO approval

Notes: The figures plot the estimates of β_2 in Equation 2 for ratings on cleanliness and the estimates' 95% confidence intervals. Number of observations is 60,880. The heterogeneity is performed by splitting the sample given the value of $\bar{R}_{it}^{location}$ before December 2018.

Additional Results for the Alternative Specification

Table A5. Summary statistics for treated and control listings in alternative specification

	(1) Treated		(2) Control		(3) Difference	
	Mean	S.D.	Mean	S.D.	Δ	p-Value
Listing's tot. reviews ($n_{i,t}$)	27.54	47.32	20.30	39.36	7.24	(0.00)
Listing's n. reviews per month ($n_{i,t} - n_{i,t-1}$)	2.70	2.78	2.48	2.88	0.22	(0.05)
Listing's price per night (\$, USD)	144.19	140.72	134.41	182.67	9.78	(0.11)
Overall Rating (#stars)	4.70	0.53	4.70	0.60	0.00	(0.85)
Accuracy Rating (#stars)	4.81	0.50	4.80	0.54	0.01	(0.76)
Check-in Rating (#stars)	4.85	0.47	4.86	0.48	-0.01	(0.47)
Cleanliness Rating (#stars)	4.71	0.58	4.71	0.63	0.00	(0.90)
Communication Rating (#stars)	4.85	0.47	4.83	0.57	0.02	(0.30)
Location Rating (#stars)	4.81	0.45	4.77	0.58	0.04	(0.03)
Value-for-money Rating (#stars)	4.72	0.56	4.72	0.62	0.00	(0.98)
Host's listings (#number)	5.60	9.18	5.62	16.84	-0.02	(0.97)
Owners in neighborhood (%)	26.87	23.12	45.37	25.78	-18.50	(0.00)
Number of listings	3,072		755		3,827	

Notes: The table compares the characteristics of treated (column(1)) and control group (column (2)) before the HSO implementation. Column (3) report the difference between the two groups and test its statistical significance with t-test (of which we report the p-Value).

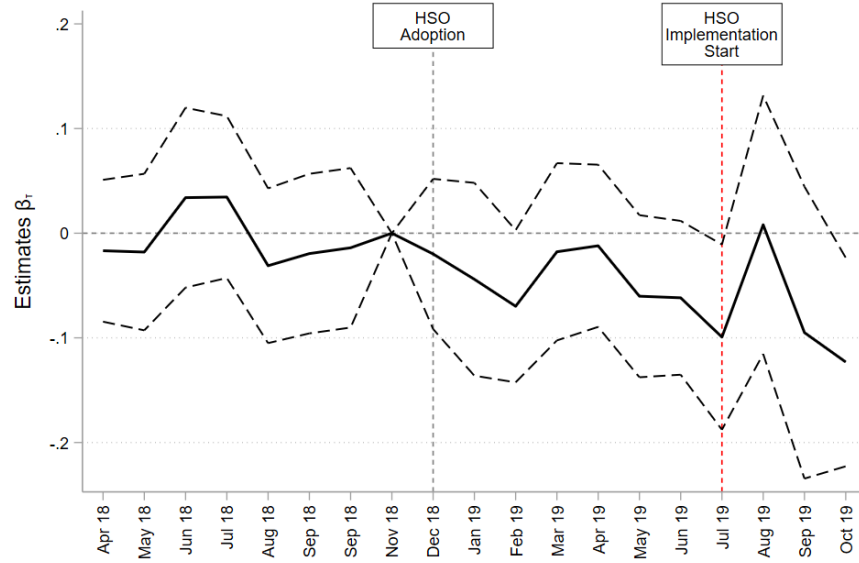


Figure A23. Event study estimates comparing ratings on check-in in the City of Los Angeles with other cities in the county.

Notes: The plot reports the estimates of β_τ from $r_{it}^{check-in} = \sum_{\tau=Apr18}^{Oct19} \beta_\tau LAcity_i \times \mathbb{1}(t = \tau) + \phi X_{it} + \gamma LinearTrend_t + \mu_i + \varepsilon_{it}$ and the confidence interval at 95%. Coefficients after the HSO approval become significantly different from the coefficient at the reference period - November 2018 (normalized at zero). We do not find significant pre-trend and the end-game effect is particularly strong after the policy announcement.

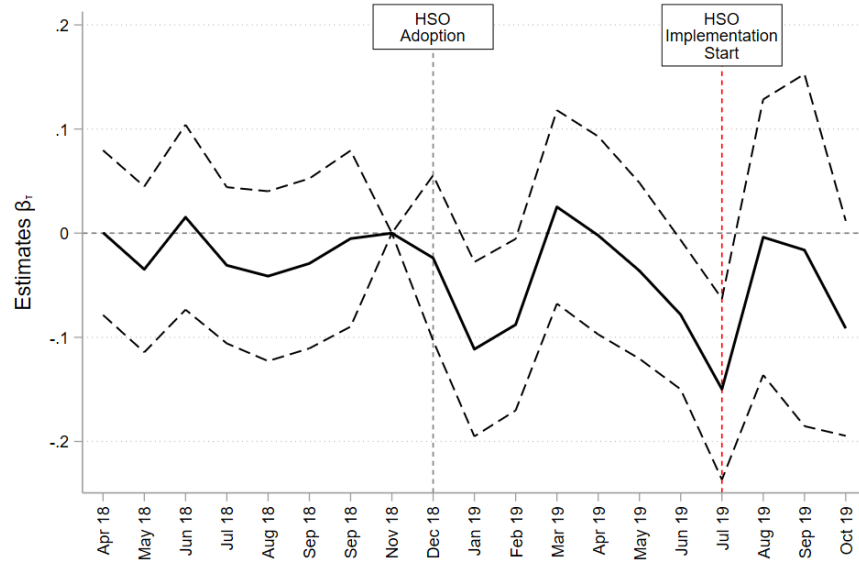


Figure A24. Event study estimates comparing ratings on communication in the City of Los Angeles with other cities in the county.

Notes: The plot reports the estimates of β_τ from $r_{it}^{communication} = \sum_{\tau=Apr18}^{Oct19} \beta_\tau LAcity_i \times \mathbb{1}(t = \tau) + \phi X_{it} + \mu_i + \gamma LinearTrend_t + \varepsilon_{it}$ and the confidence interval at 95%. Coefficients after the HSO approval become significantly different from the coefficient at the reference period - November 2018 (normalized at zero). We do not find significant pre-trend and the end-game effect is particularly strong after the policy announcement.

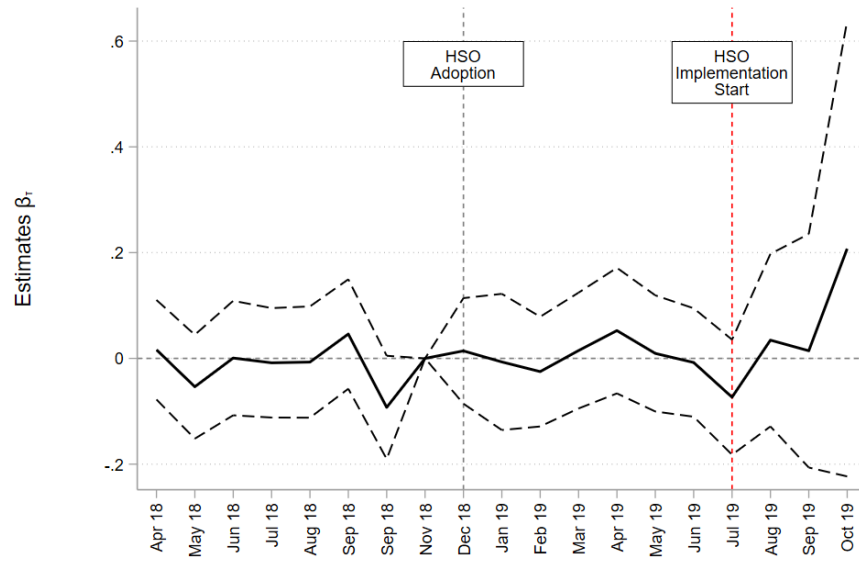


Figure A25. Event study estimates comparing ratings on cleanliness in the City of Los Angeles with other cities in the county.

Notes: The plot reports the estimates of β_τ from $r_{it}^{cleanliness} = \sum_{\tau=Apr18}^{Oct19} \beta_\tau LAcity_i \times \mathbb{1}(t = \tau) + \phi X_{it} + \mu_i \gamma + LinearTrend_t + \varepsilon_{it}$ and the confidence interval at 95%. Coefficients after the HSO approval are not significantly different from the coefficient at the reference period - November 2018 (normalized at zero). We find no end-game effect on the listings' ratings on cleanliness.