

What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features

Shunyuan Zhang

Harvard Business School

szhang@hbs.edu

Dokyun Lee

Questrom School of Business, Boston University

dokyun@bu.edu

Param Vir Singh

Tepper School University, Carnegie Mellon University

psidhu@cmu.edu

Kannan Srinivasan

Tepper School University, Carnegie Mellon University

kannans@cmu.edu

ABSTRACT

We study how Airbnb property demand changed after the acquisition of *verified* images (taken by Airbnb's photographers) and explore what makes a good image for an Airbnb property. Using deep learning and difference-in-difference analyses on an Airbnb panel dataset spanning 7,423 properties over 16 months, we find that properties with verified images had 8.98% higher occupancy than properties without verified images (images taken by the host). To explore what constitutes a good image for an Airbnb property, we quantify 12 human-interpretable image attributes that pertain to three artistic aspects—composition, color, and the figure-ground relationship—and we find systematic differences between the verified and unverified images. We also predict the relationship between each of the 12 attributes and property demand, and we find that most of the correlations are significant and in the theorized direction. Our results provide actionable insights for both Airbnb photographers and amateur host photographers who wish to optimize their images. Our findings contribute to and bridge the literature on photography and marketing (e.g., staging), which often either ignores the demand side (photography) or does not systematically characterize the images (marketing).

Keywords: sharing economy, Airbnb, property demand, computer vision, deep learning, image feature extraction, content engineering

1. Introduction

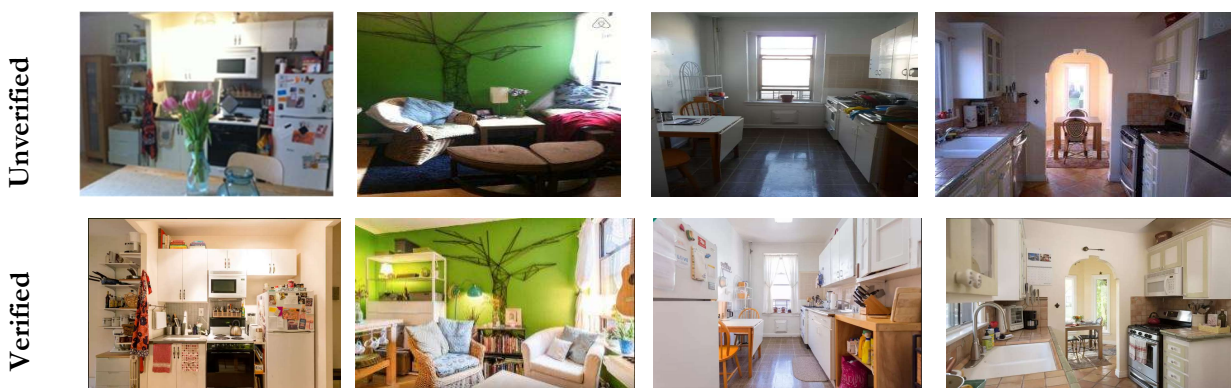
The global sharing economy market has been rapidly expanding in recent years and is projected to generate roughly \$335 billion by 2025 (PwC report 2015). On Airbnb, travelers can find lodging in home environments rather than hotels, and hosts can generate rental income from spare rooms or properties. As the world's largest home-sharing platform, Airbnb recently was valued 20% higher than Marriott and hosted 25% more guests per night than Hilton Worldwide (Winkler and Macmillan 2015).

Despite its success, Airbnb faces a significant problem in resolving the uncertainty that consumers face when evaluating property quality. The inefficiency of transferring information required by prospective guests creates transactional friction and leads to the loss of users. Reports show that quality uncertainty leads many potential consumers to choose trusted hotel brands over Airbnb (PwC report 2015; Ufford 2015).

Airbnb attempts to alleviate quality uncertainty with features such as customer reviews, host verification, property descriptions, and property images. In particular, property images provide visual information and reduce uncertainty about experiential aspects (e.g., cleanliness and mood) of units in ways that written reviews and descriptions cannot. While hotel images are taken by professional photographers, however, most Airbnb property images are taken by hosts—usually amateur photographers. Therein lies a source of inefficiency of information transfer that causes uncertainty for potential guests. Hosts often lament the poor quality of their own amateur photos and fear that their properties appear small.

To address this concern, in 2011, Airbnb launched a “photography program” that gives interested hosts free access to local professional photographers who are commissioned by the company to take photos of the host's property. A “verified” mark appears on images shot and uploaded by Airbnb's professional photographers. Figure 1 juxtaposes the original amateur photographs with Airbnb's professional photographs of the same room to illustrate the improvement.

Figure 1. Comparison of Unverified and Verified Photos



It remains unclear, however, whether the professional photography program has been beneficial for Airbnb and its hosts. In fact, the photography program has raised much controversy among Airbnb hosts and guests. It seems plausible that verified photos could attract more guests, but they also could oversell or misrepresent the property, leading to a negative impact on property demand—a fear voiced by some Airbnb hosts on the host

forum.¹ The Airbnb photography program raises a series of questions: (1) Do properties with verified photos experience higher demand? (2) If so, what systematic differences between verified and unverified photos might explain the higher demand? (3) Finally, what constitutes a good image for an Airbnb property?

To answer these questions, we collected panel data from 13,000 Airbnb listings with over 510,000 property images in seven major US cities from January 2016 through April 2017. The dataset contains rich information about each property's monthly reservations, photos, price, and other details about the property and host. The dataset captures variation in property images across both units and time, so we can observe some of the properties transition from using unverified photos to verified photos.

We use Amazon Mechanical Turk (AMT), a crowdsourcing platform for on-demand human tasks, to classify a random set of pictures into the binary categories of high- and low-quality. We use the manually classified training set to develop a scalable model that takes advantage of computer vision and deep learning. Specifically, taking pixel-level information from the images as input, we apply a convolutional neural network (CNN) to classify the aesthetic quality of each image in the training sample. The CNN model is optimized to extract a hierarchical set of features from images and learn the relationship between the set of features and the image's label (in this case, "high quality" or "low quality"). Then, we use our trained CNN image quality classifier to label the property images in the Airbnb dataset.

We employ a difference-in-difference (DID) analysis to compare the demand for properties with verified photos and with unverified photos. Hosts self-select into the Airbnb photography program, so we address endogeneity concerns through propensity score weighting (PSW). The PSW approach assumes that selection into the treatment group is captured by observed variables, and the algorithm matches treated and untreated units using estimated treatment probabilities. The PSW approach mitigates the self-selection problem only so far as the factors that affect both the treatment probability and property demand are captured in observed variables. Thus, we conduct a Rosenbaum bounds analysis to test the robustness of our results to hypothetical unobservables. The results of the Rosenbaum analysis are within the range reported in prior literature, and a series of other tests suggests similarly robust results. Nevertheless, we cannot fully rule out the possibility that selection is driven by unobserved variables in the absence of a random experiment, and we caution readers to interpret our results with this caveat in mind.

We find that the occupancy rate is 8.985% higher for properties with verified photos than for properties with unverified photos. The treatment coefficient is positive and significant even after controlling for other sources of information, such as guest reviews. The estimated treatment coefficient decreases dramatically after we incorporate photo characteristics into the demand model. Particularly, the coefficient decreases by 41.0% when we control for image quality, suggesting that a significant portion of the coefficient is explained by the higher quality of the verified images.

¹ <http://airhostsforum.com/t/professional-photography/3675/35>.

One of our key objectives is to determine what makes a good image for an Airbnb property. Our CNN model is highly accurate at predicting image quality, but the CNN-extracted features are uninterpretable. To provide better guidance for managers, we use the photography literature to identify 12 human-interpretable image attributes that are relevant to image quality in the real estate context. We theorize the relationship between each of the 12 image attributes and property demand. The 12 attributes fall under three key artistic aspects: composition, color, and the figure-ground relationship. Composition is the arrangement of visual elements in the photograph; ideally, the composition leads the viewer's eyes to the center of focus (Freeman 2007). We capture composition with four attributes: diagonal dominance, the rule of thirds, visual balance of color, and visual balance of intensity. Color can affect the viewer's emotional arousal. The marketing literature has studied the impact of color on consumer behavior particularly in the context of web design, product packaging design, and advertisement design (Gorn et al. 1997, Gorn et al. 2004; Miller and Kahn 2005). We include five aspects related to color: warm hue, saturation, brightness, contrast of brightness, and image clarity. The principle of the figure-ground relationship is one of the most basic laws of perception and is used extensively by expert photographers to plan their photographs. In visual art, the *figure* refers to the key region (i.e., foreground), and the *ground* refers to the background; photographs in which the figure is inseparable from the ground do not retain the viewer's attention. We include three attributes: the area difference, texture difference, and color difference between the figure and ground.

We use computer vision methods to score the images on the 12 image attributes, and we find systematic differences between the *verified* photos and both the unverified low-quality and unverified high-quality photos. The verified photos dominate the unverified high-quality photos on all attributes except for the rule-of-thirds and saturation. The verified photos score higher than the unverified high-quality photos across the attributes for which the theorized effect of the attribute is positive; however, the verified photos score lower than the unverified high-quality photos when the theorized effect of the attribute is negative.

We explore how the 12 image attributes might be related to property demand, though we cannot identify causal effects of these 12 image attributes on property demand given the observational nature of our data. That is, the image attributes change when a property acquires verified photos, but the choice to do so is endogenous, making any changes in the image attributes also endogenous. As a result, our analysis of the relationship between the image attributes and property demand should be viewed as exploratory. After controlling for the 12 attributes, the estimated treatment coefficient becomes statistically insignificant. The results suggest that the positive coefficient of verified photo acquisition may derive from the 12 image attributes, which together capture both vertically-differentiated quality and horizontally-differentiated taste. Airbnb professional photographers seem to be better than host photographers at capturing the attributes that matter for Airbnb property demand.

Of the 12 image attributes, the visual balance of color is most strongly related to property demand, followed by image clarity and the contrast of brightness. The visual balance of color refers to color symmetry, which can be affected by both the property itself and the position from which the image is captured. Image clarity refers to

the extent to which the image conveys visual information. The unverified low-quality images scored poorly on image clarity; the verified photos scored almost twice as high. Even without employing a professional photographer, hosts can improve image clarity through the effective use of lighting and access to a good camera. Finally, the contrast of brightness captures the difference in illumination between the brightest and dimmest points in the image; a low contrast of brightness indicates that illumination is relatively even across the image. The verified photos have a significantly lower contrast of brightness than unverified high-quality images. Interestingly, several hosts on the Airbnb community forums complained that the contrast of brightness is so low in the verified photos that they appear washed out, but we find the predicted negative relationship between the contrast of brightness and property demand. In other words, consumers seem to prefer the low contrast of brightness that appears in verified photos.

This study makes several contributions. The ability to capture and analyze unstructured data, particularly images, is a nascent field with few papers (e.g., Wang et al. 2018, Malik and Singh 2019, Netzer et al. 2019, Malik et al. 2020, Liu et al. 2020). Ours is among the first to examine how Airbnb property demand changed with the adoption of the Airbnb photography program. From a managerial perspective, our results suggest that the program is successful for hosts who choose to take advantage of it—properties with verified images achieved an 8.98% higher occupancy rate than properties without verified images. However, more than half of the Airbnb properties use low-quality images. Our findings may motivate hosts to obtain high-quality images of their properties.

Although it seems intuitive that the demand should be higher for properties with good images, the components of a “good image” are less obvious. A key practical finding of our paper is the identification of 12 interpretable image attributes that significantly correlate with property demand. Our investigation has parallels with marketing studies on color as a key image attribute that affects consumers’ perceptions and product demand. Beyond color, we consider composition and the figure-ground relationship, and we find that even the high-quality unverified images are systematically inferior to the verified images on most of the 12 attributes. Our findings may be valuable to photographers who are attempting to optimize the appeal of their images for Airbnb consumers.

Moreover, insights from our paper can guide image-content engineering efforts in short-term lodging contexts beyond Airbnb (e.g., hotels) and real estate markets. Our approach efficiently computes an extensive set of image attributes (at ~1.06 seconds per image) on a 2-Intel-Haswell (E5-2695 v3) CPU, so it can be implemented as a scalable real-time model. Our application of an existing method for image-quality classification and feature extraction can be adopted by photographers of firms and hosts to check their image quality and identify the shortcomings of their photographs. Such a practice ultimately could improve the efficiency of image-based information transfer and may increase the appeal of their properties.

Prior studies in the marketing literature have considered the effects of images and visual elements, but our paper differs from existing papers in three ways. First, the extant studies focus on the viewer’s emotional arousal and are restricted to either isolated image features (e.g., color; Gorn et al. 1997, Gorn et al. 2004) or high-level

image content or style (e.g., whether an image is “art”; see Hagtvedt and Patrick 2008). By contrast, we use the photography literature to identify 12 image attributes on which any photograph can be evaluated, and we explore their relationships with property demand. Second, the relevant literature focuses primarily on high-quality images (Bertrand et al. 2010), but e-commerce (including the sharing economy) often uses low-quality, user-generated product images. Finally, the evolving stream of marketing literature on the impact of images on consumers’ product perceptions focuses on the mere presence or absence of an image rather than specific image attributes (e.g., Mitchell and Olsen 1981; Meyers-Levy and Peracchio 1992; Peracchio and Meyers-Levy 1994, 2005; Scott 1994; Valdez and Mehrabian 1994; Gorn et al. 1997; Larsen et al. 2004; Miller and Kahn 2005).

2. Empirical Framework

2.1 Data Description

We randomly selected more than 13,000 Airbnb property listings from seven US cities (Austin, Boston, Los Angeles, New York, San Diego, San Francisco, and Seattle), and we collected data on the listings from January 2016 to April 2017. We obtained information about each property host from that host’s public profile on Airbnb.com. Each profile specifies the date on which the host joined Airbnb and whether the host had a verified Airbnb account at the time of the analysis. For each property, we obtained information about static characteristics: location (city and zip code), property type (e.g., house, apartment), property size (number of beds), amenities (e.g., pool, AC, proximity to a beach), and capacity (maximum guests). We also obtained information about dynamic characteristics: property bookings, nightly prices, guests’ reviews, property photos, and whether the photos were verified. In the appendix (Section VI), we provide detailed methods for data collection and matched sample construction. In the next section, we describe the measures of our key variables and summarize their measurement statistics.

2.2 Definitions and Measures of Key Variables

Treatment Group, Untreated Group, and Treatment Status

The panel data cover 16 one-month periods from January 2016 through April 2017. The “treatment” is the adoption of verified photos; therefore, a property is “treated” if it used verified property photos during the observation period. The sample for our main analysis consists of 7,423 unique properties that did not have verified photos in January 2016;² 212 had verified photos by the end of April 2017 (the treatment group), and the remaining 7,211 properties did not (the untreated group). We capture the treatment with three indicator variables. $TREAT_i$ equals 1 (0) if property i belongs to the treatment group (untreated group). $AFTER_{it}$ equals 1 (0) if period t is after (before) the period when property i was first observed to have verified photos. For example, if a property acquired verified property photos in March 2016, then $AFTER_{it}$ equals 0 for the periods of January

² Of the 13,000 listings in our original dataset, approximately 5,000 were already “treated” in January 2016. We exclude them from our main analyses, but as a robustness check, we repeat our main analysis with the pre-treated units, and we obtain consistent results (see Section V.7 in the appendix for details).

and February 2016 and equals 1 for the periods of March 2016 onward. Hence, the treatment status indicator $TREATIND_{it} = TREAT_i * AFTER_{it}$ equals 1 if i was treated in period t and equals 0 otherwise.

Property Demand

We purchased listing-level booking data from a company that specializes in collecting Airbnb property demand data. The booking data includes the number of days in a month in which the property was open (i.e., available to be booked), blocked (i.e., marked as unavailable by the host), and booked (i.e., by a guest). For property i in period t , we operationalize property demand as the occupancy rate, that is, the fraction of the open days that were booked, scaled by 100. For example, if a property in March was open for 24 days and booked for six days, then its demand for that month was $DEMAND_{it} = (6/24) * 100 = 25.00$.

Property Price

The property price for property i in period t , $NIGHTLY_RATE_{it}$, refers to the average nightly price over the days in period (Zhang et al. 2019). Property price is endogenous because it correlates with random demand shocks in the current period, which also affects property demand. To address the endogeneity concern, we use a set of instrument variables (IVs) for price. Following the extant literature, we include the characteristics of competing properties (Berry et al. 1995; Nevo 2001). The logic is that the characteristics of competing products are unlikely to be correlated with unobserved shocks in the demand for the focal property. However, the proximity of the characteristics of a property and its competitors influences the competition and, as a result, the property markup and price.³ In addition, we collect cost-related variables, that is, the factors that enter the cost/supply side but not the demand side. We use the local (zip-code level) residential utility fee obtained from OpenEI and local rental information collected from Zillow.⁴ These factors affect the host's expenses and thus provide an indirect measure of cost but are unlikely to be correlated with demand in the short-term lodging market.

Property Photos

The property photos are the set of photos posted on the property web page in the given period. Three variables characterize the property photos: photo quantity, photo quality, and the distribution of photographed room types. $IMAGE_COUNT_{it}$ is the number of photos of property available during period t . We calculate $IMAGE_QUALITY_{it}$ using machine learning techniques that are appropriate for the size of the dataset (over 510,000 images).⁵ We build a supervised image-quality classifier that classifies each image as high-quality (value

³ We compute IVs based on property type, listing type, property capacity, and number of reviews, none of which are directly controlled by the host. Competitors are defined as the properties in the same zip code.

⁴ The OpenEI dataset provides average residential, commercial, and industrial electricity rates by zip code, compiled from ABB, Velocity Suite, and the US Energy Information Administration dataset 861: <https://openei.org/doe-opendata/dataset/u-s-electric-utility-companies-and-rates-look-up-by-zipcode-feb-2011>. Zillow Research provides average home values by zip code and home size (# of bedrooms): <https://www.zillow.com/research/data/>.

⁵ The image data contains all images associated with all properties included in the dataset. That is, they include images for properties that were verified before the observation started (and hence are not included in the sample for the DiD analyses) and all images updated/added/deleted during observation periods.

of 1) or low-quality (value of 0). We calculate the average image quality, $IMAGE_QUALITY_{it}$, of all the photos associated with property i in period t . For example, if property i had 10 images in period t , and eight images are labeled as high-quality, then $IMAGE_COUNT_{it} = 10$ and $IMAGE_QUALITY_{it} = (8 * 1 + 2 * 0)/10 = 0.8$. Lastly, we include the distribution of photographed room types because professional photographers may have a more advanced understanding of which types of rooms will be most appealing to guests and, thus, present more of these aspects of properties. Specifically, we compute the proportion of the photographs that depict each of five room types: bathroom, bedroom, kitchen, living room, and outdoor area. Then, for property i in period t , the distribution is represented by a vector: $\{BATHROOM_PHOTO_RATIO_{it}, BEDROOM_PHOTO_RATIO_{it}, KITCHEN_PHOTO_RATIO_{it}, LIVINGROOM_PHOTO_RATIO_{it}, OUTDOOR_PHOTO_RATIO_{it}\}$.

Table 1 presents a summary of the statistics for the key variables at the group level. To show the overall trends in the key variables, we report statistics for the pre-treatment period (January 2016, i.e., when none of the properties in the sample were treated) and the post-treatment period (April 2017), when all the properties in the treatment group had been treated. As expected, image quality was stable over time in the control group (0.29 in January 2016 vs. 0.30 in April 2017) and improved dramatically in the treatment group (0.27 in January 2016 vs. 0.77 in April 2017), reflecting the high quality of photos shot by Airbnb professional photographers.

Table 1. Summary of the Statistics of Airbnb Properties

	Control Group		Treatment Group	
	Mean	Std. Dev.	Mean	Std. Dev.
Sample Size: # Units (N)	7,211		212	
Pre-treatment (January 2016)				
DEMAND (occupancy rate * 100)	31.07	35.93	32.57	32.30
# <i>RESERVATION DAYS</i> (demand component-days booked by guests)	5.49	8.79	6.62	8.45
# <i>AVAILABLE DAYS</i> (demand component-day not booked, not blocked)	12.91	12.34	14.87	11.86
# <i>BLOCKED DAYS</i> (demand component-days blocked by host)	12.60	13.42	9.51	12.43
IMAGE_QUALITY	0.29	0.27	0.27	0.25
<i>IMAGE_COUNT</i>	12.78	9.49	14.48	10.38
<i>BATHROOM_PHOTO_RATIO</i>	0.21	0.17	0.22	0.16
<i>BEDROOM_PHOTO_RATIO</i>	0.30	0.20	0.29	0.19
<i>KITCHEN_PHOTO_RATIO</i>	0.11	0.11	0.10	0.10
<i>LIVINGROOM_PHOTO_RATIO</i>	0.19	0.17	0.18	0.16
<i>OUTDOOR_PHOTO_RATIO</i>	0.19	0.20	0.20	0.20
<i>SECURITY_DEPOSIT</i>	181.97	347.75	202.77	333.07
<i>CLEANING_FEE</i>	48.02	53.04	54.69	58.95
<i>MAX_GUESTS</i>	3.19	2.11	3.50	2.23
<i>SUPER_HOST</i>	0.09	0.29	0.15	0.35
<i>INSTANT_BOOK</i>	0.11	0.31	0.11	0.31
<i>MINIMUM_STAY</i>	2.62	2.71	2.57	2.92
<i>RESPONSE_RATE</i>	91.00	16.29	92.25	14.31
<i>RESPONSE_TIME</i> (minutes)	261.07	365.66	225.12	338.39
<i>NIGHTLY_RATE</i> (average property price)	179.74	249.31	170.15	240.80
<i>REVIEW_COUNT</i>	16.41	26.25	20.56	26.98
<i>HAS_RATING</i>	0.68	0.47	0.78	0.41

<i>RATING_OVERALL</i> (overall score out of 100; summarizes multi-dimensional sub-ratings)	92.88	7.14	93.99	5.04
<i>RATING_COMMUNICATION</i>	9.71	0.64	9.74	0.49
<i>RATING_ACCURACY</i>	9.50	0.76	9.54	0.56
<i>RATING_CLEANLINESS</i>	9.21	0.97	9.26	0.82
<i>RATING_CHECKIN</i>	9.67	0.67	9.72	0.48
<i>RATING_LOCATION</i>	9.39	0.82	9.40	0.70
<i>RATING_VALUE</i>	9.24	0.79	9.34	0.60
<i>ZILLOW_RENTAL</i>	2,459.95	800.40	2,474.75	845.05
<i>UTILITY</i>	0.18	0.03	0.17	0.04
<i>INACTIVE PROPERTIES</i> (proportion; a property in a month is <i>inactive</i> if it's blocked full month)	0.24		0.17	
Post-treatment (April 2017)				
<i>DEMAND</i> (occupancy rate * 100)	36.79	39.46	46.58	38.37
# <i>RESERVATION DAYS</i> (demand component-days booked by guests)	8.59	10.35	10.97	10.13
# <i>AVAILABLE DAYS</i> (demand component-day not booked, not blocked)	16.23	11.56	13.41	11.49
# <i>BLOCKED DAYS</i> (demand component-days blocked by hosts)	5.17	8.79	5.62	8.94
<i>IMAGE_QUALITY</i>	0.30	0.27	0.77	0.22
<i>IMAGE_COUNT</i>	16.22	11.98	19.61	11.53
<i>BATHROOM_PHOTO_RATIO</i>	0.21	0.16	0.21	0.15
<i>BEDROOM_PHOTO_RATIO</i>	0.30	0.19	0.30	0.18
<i>KITCHEN_PHOTO_RATIO</i>	0.11	0.10	0.09	0.10
<i>LIVINGROOM_PHOTO_RATIO</i>	0.19	0.15	0.18	0.17
<i>OUTDOOR_PHOTO_RATIO</i>	0.20	0.19	0.21	0.21
<i>SECURITY_DEPOSIT</i>	149.53	325.71	179.82	280.53
<i>CLEANING_FEE</i>	41.51	55.93	46.02	51.01
<i>MAX_GUESTS</i>	3.37	2.26	3.50	2.31
<i>SUPER_HOST</i>	0.19	0.39	0.29	0.45
<i>INSTANT_BOOK</i>	0.11	0.32	0.16	0.37
<i>MINIMUM_STAY</i>	2.36	3.75	2.56	4.06
<i>RESPONSE_RATE</i>	96.63	12.17	95.82	12.50
<i>RESPONSE_TIME</i> (minutes)	150.42	266.86	167.72	287.81
<i>NIGHTLY_RATE</i> (average property price)	237.97	319.49	210.14	146.81
<i>REVIEW_COUNT</i> (number of guest reviews)	32.89	49.33	41.80	46.37
<i>HAS_RATING</i>	0.82	0.38	0.84	0.36
<i>RATING_OVERALL</i> (overall score out of 100; summarizes multi-dimensional sub-ratings)	93.80	5.34	94.16	4.45
<i>RATING_COMMUNICATION</i>	9.80	0.45	9.82	0.40
<i>RATING_ACCURACY</i>	9.60	0.60	9.66	0.51
<i>RATING_CLEANLINESS</i>	9.40	0.75	9.40	0.73
<i>RATING_CHECKIN</i>	9.79	0.45	9.78	0.45
<i>RATING_LOCATION</i>	9.47	0.67	9.50	0.65
<i>RATING_VALUE</i>	9.36	0.64	9.39	0.57
<i>ZILLOW_RENTAL</i>	2,459.80	822.06	2,426.22	823.28
<i>UTILITY</i>	0.18	0.04	0.17	0.04
<i>INACTIVE PROPERTIES</i> (proportion)	0.41		0.32	
Time-invariant				
<i>APARTMENT</i>	0.67	0.47	0.60	0.49
<i>ENTIREHOME</i>	0.59	0.49	0.61	0.49
# <i>BEDS</i>	1.70	1.22	1.81	1.21
<i>POOL</i>	0.08	0.28	0.10	0.30
<i>BEACH</i>	0.01	0.10	0.02	0.14
<i>AC</i>	0.98	0.13	0.99	0.06
<i>PARKING</i>	0.44	0.50	0.50	0.50

<i>INTERNET</i>	0.98	0.13	0.99	0.09
<i>TV</i>	0.74	0.44	0.79	0.41
<i>WASHER</i>	0.59	0.49	0.60	0.49
<i>MICROWAVE</i>	0.09	0.29	0.15	0.36
<i>ELEVATOR</i>	0.23	0.42	0.20	0.40
<i>GYM</i>	0.10	0.29	0.11	0.31
<i>FAMILY_FRIENDLY</i>	0.23	0.42	0.19	0.39
<i>SMOKE_DETECTOR</i>	0.46	0.50	0.55	0.50
<i>SHAMPOO</i>	0.35	0.48	0.45	0.50

2.3 Analysis of Property Images

2.3.1 Classifying Images as High- or Low-Quality

We build a supervised deep learning algorithm that classifies images as high- or low-quality.

Training set construction

We chose a random sample of images from our dataset to tag using AMT, a crowdsourcing platform for human tasks. A stratified (by the crude metric of quality) random sample is necessary to create a sample that is balanced as well as random. For each image, we asked the MTurkers (i.e., workers) to rate the image quality on a 1–7 Likert scale, where 1 is “very bad” and 7 is “excellent.” For guidance, we provided examples of rated photos across the quality spectrum, and we explained what we meant by “quality” (e.g., the image should be “visually pleasing” and “clearly show room/house features”). Each image was evaluated by five qualified MTurkers. In the appendix, we report details on image labeling using AMT. The image labels were further converted to a binary format (“high quality” vs. “low quality”), following practices described in the computational aesthetics literature (Datta et al. 2006). The final training sample contained 1,155 high-quality images and 1,104 low-quality images.

Training step

CNN Approach: We apply a CNN, a deep learning framework widely applied in the field of computer vision with breakthrough performances on tasks including object recognition and image classification (Krizhevsky et al. 2012). Our CNN image quality classifier (see Figure 2) represents the architecture of a classic CNN model. The CNN consists of a sequence of neural layers (also called filters); the first layer extracts features from the input image and summarizes them into an intermediate output, which becomes the input with which the next layer generates a higher-level summary, and so on. The key component in the CNN is a convolution kernel (or convolution filter), represented by an n by n weighting matrix. Given the intermediate output from the previous layer, the convolution kernel extracts features through a matrix dot product operation between the weighting matrix and the intermediate output. In other words, the sequence of layers in the CNN extracts a hierarchical set of image features from the input. A training set (in which the label, “high-quality” or “low-quality,” is already known for each image) enables the model to learn the relationships between the extracted features and the labels such that the model can extract the features that have the most discriminative power for predicting the labels. To reduce the overfitting problem in the training step, we increase the size of and random variation within the training sample by randomly applying one of three transformations to each image (e.g., data augmentation, see

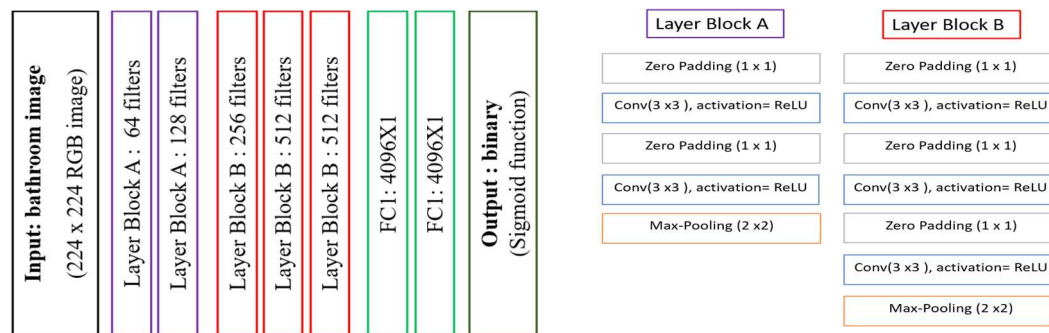
Krizhevsky et al. 2012): (1) flipping the input image horizontally, (2) rescaling the input image within a scale of 1.2, or (3) rotating the image within 20°.

Transfer Learning and Fine-Tuning the Parameters of the CNN: Because a deep learning model has many filters and parameters, it requires a large quantity of data to train the model. We overcome our limited training data by leveraging transfer learning. We start with the pre-trained parameters of the widely-applied CNN model, VGG-16 (trained on over one million images, Simonyan and Zisserman 2015) and fine-tune the parameters with 80% of our training set of Airbnb property images. We use the remaining 20% as a hold-out sample to test the performance of the trained CNN, and we achieve 90.4% accuracy). The high accuracy in predicting image quality ensures a valid interpretation of our results of image quality in the demand model. In the appendix, we provide a detailed description of the CNN architecture and technical notes on the training process.

Prediction step

Once the CNN classifier learns the relationship between image features and the image label, it can be applied to the unlabeled images (Figure 2). The classifier takes the unlabeled image as the input, extracts the hierarchical set of image features using the parameters of the trained classifier, and outputs the predicted label: either “1” for a high-quality image or “0” for a low-quality image.

Figure 2. Description of Architecture and Layer Description of the CNN Classifier



Filters: The number of convolution windows (i.e., number of feature maps) on each convolution layer.

Zero-padding: Pads the input with zeros on the edges to control the spatial size of the output. Zero-padding has no impact on the predicted output.

Max-pooling: Subsampling method. A 2×2 window slides through (without overlap) each feature map at that layer, and the maximum value in the window is picked as the representation of the window. This reduces computation and provides translation invariance.

2.3.2 Classifying the Image Room Type

We build another deep learning model to automatically classify the room in a property image as a bathroom, bedroom, kitchen, living room, or outdoor area. Again, we use transfer learning—we start with Places205, which was pre-trained on a large scene classification dataset (Zhou et al. 2014), and we fine-tune the classifier on our own training dataset (i.e., with known labels) that we collected from real-estate-related websites on which a vast number of indoor/outdoor images are classified into categories: 54,557 images of “bathrooms,” 59,082 images of “bedrooms,” 88,030 images of “kitchens,” 81,819 images of “living rooms,” and 5,734 images of “outdoor areas.” The trained classifier achieves 95.05% accuracy on the hold-out sample. We then apply the trained

classifier to the unlabeled property images. In the appendix, we provide a description of the training set and technical notes for the training step of the room-type classifier.

3. Methods and Results

We implement a DiD analysis (Heckman et al. 1997) with PSW to mitigate the endogeneity concern.

3.1 DiD Analysis

The DiD analysis requires the identification of a treatment group and (comparable) control group. Our treatment group consists of the 221 properties that did not have verified photos in January 2016 but acquired at least one verified photo by April 2017. The control group consists of the 7,211 properties that had only unverified photos during the observation window.

In an ideal setting, the two groups would be comparable such that the impact of treatment (i.e., the adoption of verified photos) on demand would be reflected by the demand difference in the post-treatment period. In our unrandomized (i.e., self-selecting) setting, however, the treatment is endogenous, and the two groups may not be comparable. As shown in Table 1, the treatment and control groups differ on some pre-treatment covariates. If certain differences affect both property demand and hosts' decisions about whether to join the photography program, then we cannot simply attribute any observed difference in the property demand to the treatment (Athey and Imbens 2006). To mitigate the endogeneity concern, we use the PSW method to create groups that are sufficiently comparable on the observed characteristics.

3.2 The PSW Method

The propensity score is the probability that an individual unit receives treatment, conditional on a set of observed covariates (Rosenbaum and Rubin 1983). Propensity scores can be used to “balance” samples and are widely used to make two groups comparable with respect to covariates (Rosenbaum 2002).

In practice, the true propensity scores are often unknown, so the PSW method estimates propensity scores by modeling the treatment probability as a function of the observed covariates. The propensity score of unit i , ps_i , is computed via a specified function, $ps_i = f(X_i\beta)$, where X_i is a $1 * M$ dimensional vector of pre-treatment observed covariates of unit i and β is an $M * 1$ dimensional vector of parameters for X .

The model finds a set of parameters that maximizes the treatment likelihood in the sample data (Rosenbaum and Rubin 1983). We estimate the parameters via logistic regression because the treatment assignment is a binary variable (i.e., treatment or control). The selection of X is based on a covariate balance check (see the appendix). The differences in the means of covariates between the treatment and control groups should be minimized. With parameter vector β estimated, we approximate for each unit i (with observed covariates X_i) the propensity score $\widehat{ps}_i(X_i)$, which is used to compute the sample weights.

In the X of the PSW model, we include covariates to control for the self-selection issue identified by Zhang et al. (2019)—that a rational host should adopt high-quality images if they have an appropriately high-quality

property and level of service. We control for the service quality through the *Response Rate* and *Response Time* (similar to Zhang et al. 2019) and capture the property quality with a rich set of property characteristics such as property amenities (internet, gym, beach, pool, etc.), property size and type, and host quality (super host).

Computing Sample Weights Based on Propensity Scores

We use propensity scoring for a weighting strategy—the inverse probability of treatment weighting (IPTW) method—in the DiD analysis (Austin and Stuart 2015). The IPTW method calculates a weight w for unit i by taking the inverse of unit i 's propensity score:

$$\omega_i(T, X_i) = \frac{T}{ps_i(X_i)} + \frac{1 - T}{1 - ps_i(X_i)}, \quad (1)$$

where $\widehat{ps_i(X_i)}$ is the estimated propensity score of unit i computed with its observed covariates X_i , and T is a dummy variable that equals 1 if i is in the treatment group and 0 otherwise.

The PSW results are validated to ensure that the treatment and control groups in the weighted sample are balanced on the included covariates (see the appendix for details on the PSW approach, results, and validation). As shown in Table 2, the group means are statistically indifferent, suggesting that the PSW method successfully removed systematic differences in host and property observed characteristics that might confound the treatment assignment. See Section V.5 in the appendix for the analysis details.

Table 2. PSW Validation: Covariate Balance Check

Variables	Weighted Means		T-test	
	Treated	Untreated	<i>t</i>	<i>p</i> -value
REVIEW_COUNT	20.56	19.88	0.27	0.790
IMAGE_QUALITY	0.27	0.25	1.00	0.316
IMAGE_COUNT	14.48	15.1	-0.67	0.506
NIGHTLY_RATE	170.15	191.36	-1.14	0.257
MINIMUM_STAY	2.57	2.57	-0.00	1.000
MAX_GUESTS	3.5	3.67	-0.82	0.410
RESPONSE_RATE	92.25	91.19	0.79	0.431
RESPONSE_TIME (minutes)	225.12	260.98	-1.18	0.238
SUPER_HOST	0.15	0.11	1.05	0.292
INSTANT_BOOK	0.11	0.11	-0.14	0.888
# BLOCKED DAYS	9.51	8.32	1.10	0.271
# RESERVATION DAYS	6.62	6.74	-0.16	0.877
PARKING	0.5	0.49	0.09	0.929
POOL	0.1	0.08	0.76	0.445
BEACH	0.02	0.02	0.34	0.737
INTERNET	0.99	1	-0.58	0.563
TV	0.79	0.81	-0.55	0.579
WASHER	0.6	0.57	0.81	0.419
MICROWAVE	0.15	0.13	0.64	0.523
ELEVATOR	0.2	0.21	-0.22	0.826
GYM	0.11	0.13	-0.69	0.490
FAMILY_FRIENDLY	0.19	0.2	-0.56	0.576

<i>SMOKE_DETECTOR</i>	0.55	0.52	0.62	0.534
<i>SHAMPOO</i>	0.45	0.44	0.09	0.929
<i>BATHROOM_PHOTO_RATIO</i>	0.22	0.21	1.05	0.295
<i>BEDROOM_PHOTO_RATIO</i>	0.29	0.29	-0.26	0.795
<i>KITCHEN_PHOTO_RATIO</i>	0.1	0.1	-0.15	0.880
<i>LIVINGROOM_PHOTO_RATIO</i>	0.18	0.19	-0.52	0.602
<i>OUTDOOR_PHOTO_RATIO</i>	0.2	0.21	-0.13	0.898
<i>SECURITY_DEPOSIT</i>	202.77	225.19	-0.65	0.517
<i># OPEN DAYS</i>	21.49	22.68	-1.10	0.271
<i>CANCELLATION_STRICT</i>	0.26	0.29	-0.89	0.371
<i># BEDS</i>	1.81	1.96	-1.17	0.242
<i>APARTMENT</i>	0.6	0.61	-0.27	0.786
<i>ENTIRE_HOME</i>	0.61	0.64	-0.64	0.522

3.3 Model Specification and DiD Estimator

We obtain our DiD estimator through a weighted least squares (WLS) regression, which computes sampling weights using estimated propensity scores. Let $DEMAND_{itcym}$ denote the demand for property i (in city c) in year y and month m (further, let t denote month m in year y):

$$DEMAND_{itcym} = INTERCEPT + \alpha TREATIND_{it} + \lambda CONTROLS_{it} + PROPERTY_i + SEASONALITY_{cym} + \varepsilon_{it}, \quad (2)$$

where $TREATIND_{it}$ is the treatment status indicator, which equals 1 if property i has received treatment in period t and equals 0 otherwise. The key coefficient α estimates the percentage change in property demand that is associated with the treatment (i.e., having verified photos), and ε_{it} is a random shock in period t on property i 's demand, assumed to follow an *i.i.d.* normal distribution. The vector $CONTROLS_{it}$ represents a set of control variables that may be correlated with property demand, for example, the property rules and consumer reviews.⁶ We include the property fixed effect term, $PROPERTY_i$, to capture time-invariant factors that may impact property demand, such as geographic information and property-specific characteristics. We also include the time fixed effect term $SEASONALITY_{cym}$, which captures any city-specific trends in property demand (i.e., we allow each city to have its own seasonal pattern).

Note that the key outcome variable in this study is property demand, operationalized as the monthly occupancy rate. As an empirical extension, in the appendix, we analyze whether the property price changed with treatment (i.e., we replace the DV in Eq. 2 with the property's nightly rate). We find that after controlling for

⁶ The vector $CONTROLS$ includes two metrics that measure hosts' responsiveness, $RESPONSE_RATE$ (percentage of messages/requests from guests that receive a response from the host) and $RESPONSE_TIME$ (average number of minutes to respond to a guest); MIN_STAYS , the minimum number of nights that a guest can book; MAX_GUESTS , the maximum number of guests that may stay at once; $SECURITY_DEPOSIT$, the money that the guest will be charged if the host claims that the guest damaged the property and Airbnb approves the claim; $CANCELLATION_STRICT$, whether the cancellation policy is strict (1) or not strict (0); $SUPER_HOST$, whether the host has a "super host" badge (1) or not (0), which Airbnb assigns based on consumers' reviews, the host's responsiveness, etc.; $BUSINESS_READY$, whether the property has business-related amenities (1) or not (0); HAS_RATING , whether the average guest ratings are presented on the property page (1) or not (0); and the interaction terms of HAS_RATING and the multi-dimensional ratings.

seasonality, property characteristics, and time-varying variables such as the number of reviews, the coefficient of the treatment on property price is insignificant (see Section V8 in the appendix for detailed discussion).⁷

As shown in Table 3, the estimated coefficient of the key variable *TREATIND* suggests a positive, significant treatment effect. Specifically, properties that used verified photos had 8.985% higher occupancy than the control group. On average, an untreated property was open 18.1 days per month, so the treatment coefficient corresponds to an average of 18.1 days/month * 8.985% * 12 months/year = 19.5 additional booked days in a year (or 1.6 additional days in a month). In terms of revenue, the acquisition of verified photos corresponds to an average of \$179.5/day * 19.5 days = \$3,500.30 more per year.⁸

Table 3. DiD Model: Regressing Property Demand on Verified Photos

VARIABLES	Main DiD Model (Eq. 2)	
	ESTIMATES	Robust S.E.
<i>TREATIND</i>	8.985***	1.660
<i>log REVIEW_COUNT_{t-1}</i>	9.375***	0.930
<i>NIGHTLY_RATE</i>	-0.146***	0.0320
<i>INSTANT_BOOK</i>	4.156**	1.361
<i>CLEANING_FEE</i>	0.0808***	0.0184
<i>MAX_GUESTS</i>	0.260	1.117
<i>RESPONSE_RATE</i>	0.0699	0.0430
<i>RESPONSE_TIME (minutes)</i>	-0.000477	0.00161
<i>MINIMUM_STAY</i>	0.133	0.131
<i>SECURITY_DEPOSIT</i>	0.00177	0.00201
<i>SUPER_HOST</i>	3.801*	1.494
<i>BUSINESS_READY</i>	1.806	0.985
<i>CANCELLATION_STRICT</i>	1.016	1.271
<i>HAS_RATING</i>	14.32	12.25
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	-0.212	1.420
<i>HAS_RATING</i> × <i>ACCURACY</i>	0.878	1.211
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	-1.344	1.133
<i>HAS_RATING</i> × <i>CHECKIN</i>	-2.060	1.526
<i>HAS_RATING</i> × <i>LOCATION</i>	-0.757	1.183
<i>HAS_RATING</i> × <i>VALUE</i>	2.141	1.176
Fixed Effect	Property	
Seasonality	City-Year-Month	
Num. Observations	76,901	
R-squared	0.6608	
Note: The demand model does not include the number of open nights because the D.V. is computed as a function of it (occupancy rate = #reservation days/#open days). This variable is used in the matching step (see Table A1 in the appendix). * <i>p</i> < 0.05, ** <i>p</i> < 0.01, *** <i>p</i> < 0.001.		

⁷ A possible explanation is that a change in photo quality does not necessarily reflect changes in the property or host. In a report on the smart pricing model, Airbnb did not describe property pictures as one of the many “factors at play” that the algorithm used: <https://blog.airbnb.com/smart-pricing/>.

⁸ The properties in our sample in the pre-treatment period (January 2016) were priced at \$ 179.5 per night on average.

3.4 Validating the DiD Model

We implement a set of analyses to validate our combination of the DiD model and the PSW strategy. We begin with a falsification check that examines the critical “common pre-treatment trends” assumption, followed by a Rosenbaum bounds analysis for the selection on unobservables and several additional robustness checks.

3.4.1 Falsification Check: Pre-Treatment Trends

The validity of the DiD approach (Eq. 2) relies on a critical assumption of common pre-treatment trends: that is, the two weighted groups should exhibit common trends in their demand before treatment (Angrist and Pischke 2008). We test the common trends assumption using a relative-time model of pre-treatment periods. Following the extant literature (e.g., Wang and Goldfarb 2017), we decompose the pre-treatment periods into a series of dummy variables:

$$DEMAND_{icym} = INTERCEPT + \alpha TREATIND_{it} + \sum_j \beta_j (PRE_{it}(j) \cdot TREAT_i) + \lambda CONTROLS_{it} + SEASONALITY_{cym} + PROPERTY_i + \varepsilon_{it}. \quad (3)$$

We define $PRE(1)$ as the period prior to the treatment month and set it as the reference period (i.e., we normalize its coefficient to zero). Then, $PRE(2)$ is two months prior to treatment, $PRE(3)$ is three months prior to treatment, and $PRE(4)$ represents all the periods from the beginning (i.e., January 2016) through four months prior to treatment (Autor 2003). For properties with fewer than four pre-treatment periods (e.g., a property that acquired verified photos in February 2016 would have only one pre-treatment period), the period dummies that correspond to months before January 2016 are set to zero.

The set of coefficients β_j allows us to validate the DiD model by comparing the trends in the property demand of the weighted control and treatment groups prior to treatment. The common trends assumption is validated if there is no significant pre-treatment difference between the groups, and Table 4 confirms that none of β_j the period dummy coefficients are statistically different from zero. Moreover, the set of β_j does not exhibit an increasing trend in the property demand for the treatment units compared with the control units. In other words, the demand for the treated units was not already deviating from the demand for the control units prior to treatment, which would suggest that the difference between groups was caused by an idiosyncratic shock that affected both the treatment likelihood and property demand rather than by the treatment itself.

Table 4. Falsification Check: A Relative-Time Model of Pre-Treatment Trends

VARIABLES	Relative-Time Model (Eq. 3)	
	ESTIMATES	Robust S.E.
$PRE(4) * TREAT$	2.230	2.503
$PRE(3) * TREAT$	-0.894	3.164
$PRE(2) * TREAT$	1.302	3.102
$PRE(1) * TREAT$ (reference month)	--	--
$TREATIND$	9.988***	2.311
Property (Non-Photo) Characteristics		
$\log REVIEW_COUNT_{t-1}$	9.338***	0.933
$NIGHTLY_RATE$	-0.147***	0.0320
$INSTANT_BOOK$	4.166**	1.363
$CLEANING_FEE$	0.0811***	0.0184

<i>MAX_GUESTS</i>	0.199	1.108
<i>RESPONSE_RATE</i>	0.0730	0.0431
<i>RESPONSE_TIME (minutes)</i>	-0.000403	-0.00161
<i>MINIMUM_STAY</i>	0.132	0.131
<i>SECURITY_DEPOSIT</i>	0.00175	0.002
<i>SUPER_HOST</i>	3.764*	1.495
<i>BUSINESS_READY</i>	1.821	0.985
<i>CANCELLATION_STRICT</i>	1.05	1.271
<i>HAS_RATING</i>	14.57	12.22
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	-0.229	1.421
<i>HAS_RATING</i> × <i>ACCURACY</i>	0.978	1.21
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	-1.403	1.134
<i>HAS_RATING</i> × <i>CHECKIN</i>	-2.117	1.521
<i>HAS_RATING</i> × <i>LOCATION</i>	-0.716	1.182
<i>HAS_RATING</i> × <i>VALUE</i>	2.118	1.177
Fixed Effect	Property	
Seasonality	City-Year-Month	
Num. Observations	76,901	
R-squared	0.6609	
Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.		

3.4.2 Selection on Unobservables

The PSW method addresses endogeneity concerns regarding observed variables, but it does not rule out a possible hidden bias involving unobserved variables that might influence the treatment likelihood and outcome variable simultaneously. We assess the sensitivity of our estimation to a potential hidden bias with the Rosenbaum bounds test (Rosenbaum 2002), which determines how much of a change in the odds ratio of treatment would need to be due to unobservables to nullify the treatment effect identified by the PSW method. We can be more confident about the validity of the estimation results if the Rosenbaum bounds test suggests that unobservables would have to cause a large change in the odds ratio to overturn the estimated treatment effect.

Our results for the Rosenbaum bounds test with an examination of the Hodges-Lehmann estimates (Rosenbaum 1993) suggest that unobservables would need to increase the odds ratio of treatment by at least 55% (i.e., $\gamma \geq 1.55$) to overturn the positive estimated treatment effect on property demand. The results of our sensitivity analysis are similar to the results obtained in the extant literature (DiPrete et al. 2004; Sun and Zhu 2014; Manchanda et al. 2015; Li et al. 2016), which reported gamma values ranging from 1.2 to 1.6. This suggests that our study is robust to the hypothetical unobserved factors that may affect the treatment likelihood. We provide detailed methods and results in the appendix.

3.4.3 Additional Robustness Tests

In addition to the falsification test (Section 3.4.1) and the Rosenbaum analysis (Section 3.4.2), we perform a comprehensive set of seven tests that verify the robustness of our main findings. We briefly describe the tests below; detailed descriptions and results are in Section V in the appendix.

First, we run two analyses to test for potential inflation in the treatment coefficient in the long term, which could occur if Airbnb's ranking algorithm favors properties with verified images. We estimate our main DiD

specification on a set of subsamples in which we include a shorter period (four months post-treatment) for each treated unit. The estimated coefficient of *TREATIND* remains significant and positive. We also use a relative-time model to estimate the treatment effect in each month following treatment. We find that the estimated coefficient of *TREATIND* in the month following treatment is close to our main finding. The two analyses add confidence that our main finding was not driven by long-term inflation.

Second, we test whether properties with particular amenities were more likely to acquire verified photos at a time when these amenities were most attractive (e.g., a pool in the summer). We include interaction terms for the dummy *AFTER* and meaningful amenities (pool, beach, AC) in the DiD model and find consistent estimated results. The results suggest that the positive treatment coefficient is not due to higher demand for these amenities following treatment or in particular seasons.

Third, we investigate whether the positive treatment effect was driven by unobserved enhancements of the property or host. We regress the multidimensional review ratings (communication, cleanliness, etc.) on *TREATIND*. The coefficient of *TREATIND* is statistically insignificant, suggesting no substantial change in the quality of guests' experiences after the acquisition of verified photos.

Fourth, we examine whether hosts in the treatment group changed their management of the property calendar (e.g., made the property open for more nights per month) at the same time as they adopted verified photos. We regress *#_OPEN_DAYS* on *TREATIND* and other controls and find an insignificant coefficient of *TREATIND*, suggesting that the main results are not due to an increase in property availability after treatment. Additionally, we include *#_OPEN_DAYS* in demand model and have similar finding of estimated *TREATIND*.

Fifth, we replicate the DiD model with a log transformation of the numeric variables and find no change in the estimated treatment coefficient.

Sixth, Zhang et al. (2019) find that image quality affects the review probability, and the number of reviews affects property demand. We incorporate the probability that a guest writes a review for the property. In this test, we match properties based on review probability as well as other covariates. The DiD estimation results on the matched sample are consistent with our main results, and the estimated coefficient of review probability.⁹

Seventh, we use a different control group: the pre-treated units (i.e., those that adopted verified photos before January 2016 and thus were excluded from the main analyses). The pre-treated units may represent a stronger control because the pre-treated and to-be-treated properties might be more similar in the characteristics that drive the treatment likelihood (Narang and Shankar 2019). Two analyses report estimated coefficients of *TREATIND* that are consistent with our main finding.

⁹ We do not include the review writing probability in the main DiD model as: 1) review writing probability is indirectly controlled for through number of reviews and property characteristics, and 2) including it will significantly reduce the sample size. See Section V.9) for detailed discussions.

3.5 Examining the Aesthetic Qualities of Verified vs. Unverified Photos

Our main analyses in Section 3.3 established that the demand for properties with verified images was 8.98% higher than the demand for properties with unverified images. In this section, we explore the potential sources of the positive coefficient of verified photos *correlationally*, by extracting image aesthetic quality. We analyze the images and then estimate a demand equation (see Eq. 4) that captures three aspects of a property's image set: *IMAGE_COUNT* (the number of property images), *IMAGE_QUALITY* (the average of the binary labels of "high-quality" [1] and "low-quality" [0]), and the distribution of the five main types of photographed rooms. In Section 2.3 and the appendix, we define the variables and provide technical details regarding their measurements.

$$\begin{aligned} DEMAND_{itcym} = & INTERCEPT + \alpha TREATIND_{it} + \mu IMAGE_COUNT_{it} + \gamma IMAGE_QUALITY_{it} \\ & + \rho_1 BATHROOM_PHOTO_RATIO_{it} + \rho_2 BEDROOM_PHOTO_RATIO_{it} \\ & + \rho_3 KITCHEN_PHOTO_RATIO_{it} + \rho_4 LIVING_PHOTO_RATIO_{it} + \lambda CONTROLS_{it} \\ & + SEASONALITY_{cym} + PROPERTY_i + \varepsilon_{it} \end{aligned} \quad (4)$$

Table 5 reports the estimation results; the model specification does not include *IMAGE_QUALITY* in column (1) and does include it in column (2). Recall that the estimated coefficient of the key variable *TREATIND* was 8.985 in the original model; the coefficient decreases to 7.453 in column (1) with the inclusion of *IMAGE_COUNT*, which has a positive and significant coefficient (none of the room type ratios have a significant coefficient). The treatment coefficient decreases by another 41% (from 7.453 to 4.397) in column (2) with the inclusion of *IMAGE_QUALITY*, which has a positive and significant coefficient. This significant reduction suggests that the high quality of the verified images explains some but not all of the treatment effect, as there is a significant residual treatment coefficient even when we control for both covariates.

Table 5. DiD Model: Controlling for the Number and Quality of Property Images

VARIABLES	(1)		(2)	
	Without Image Quality		Including Image Quality	
	ESTIMATES	Robust S.E.	ESTIMATES	Robust S.E.
<i>TREATIND</i>	7.453***	1.777	4.397*	2.190
Property (Non-Photo) Characteristics				
<i>log REVIEW_COUNT_{t-1}</i>	9.754***	0.944	9.570***	0.942
<i>NIGHTLY_RATE</i>	-0.183***	0.0325	-0.187***	0.0325
<i>INSTANT_BOOK</i>	3.768**	1.357	3.664**	1.349
<i>CLEANING_FEE</i>	0.0931***	0.0188	0.0955***	0.0187
<i>MAX_GUESTS</i>	0.247	1.098	0.285	1.094
<i>RESPONSE_RATE</i>	0.0868*	0.0427	0.0886*	0.0427
<i>RESPONSE_TIME (minutes)</i>	-0.000232	0.00159	-0.000151	0.00160
<i>MINIMUM_STAY</i>	0.172	0.132	0.171	0.133
<i>SECURITY_DEPOSIT</i>	0.00211	0.00201	0.00210	0.00200
<i>SUPER_HOST</i>	3.999**	1.495	3.890**	1.494
<i>BUSINESS_READY</i>	1.805	0.977	1.813	0.974
<i>CANCELLATION_STRICT</i>	1.282	1.277	1.308	1.278
<i>HAS_RATING</i>	15.82	12.27	14.96	12.28
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	0.0499	1.418	0.0702	1.423

<i>HAS_RATING</i> × <i>ACCURACY</i>	0.588	1.209	0.681	1.206
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	-1.214	1.133	-1.144	1.141
<i>HAS_RATING</i> × <i>CHECKIN</i>	-2.260	1.525	-2.267	1.516
<i>HAS_RATING</i> × <i>LOCATION</i>	-0.785	1.186	-0.745	1.193
<i>HAS_RATING</i> × <i>VALUE</i>	2.124	1.176	2.023	1.182
<i>AFTER</i> × <i>POOL</i>	5.841	4.610	6.053	4.608
<i>AFTER</i> × <i>BEACH</i>	-11.29	10.91	-10.39	10.66
<i>AFTER</i> × <i>AC</i>	0.331	3.978	-0.198	3.963
Property Image Characteristics (+)				
<i>log IMAGE_COUNT</i>	6.874***	1.724	5.518**	1.777
<i>BATHROOM_PHOTO_RATIO</i>	0.777	8.457	0.779	8.439
<i>BEDROOM_PHOTO_RATIO</i>	-2.575	7.539	-2.128	7.499
<i>KITCHEN_PHOTO_RATIO</i>	18.04	11.27	17.20	11.22
<i>LIVINGROOM_PHOTO_RATIO</i>	-12.07	8.430	-12.41	8.405
<i>IMAGE_QUALITY</i>			8.984**	3.371
Fixed Effect	Property		Property	
Seasonality	City-Year-Month		City-Year-Month	
Num. Observations	76,901		76,901	
R-squared	0.6623		0.6628	
Note: + The coefficient for <i>OUTDOOR_PHOTO_RATIO</i> is not estimated; * <i>p</i> < 0.05, ** <i>p</i> < 0.01, *** <i>p</i> < 0.001.				

3.6. What Makes a Property Image Appealing?

Our CNN classifier predicted image quality with high accuracy, but the CNN-generated features are uninterpretable. That is, the features do not provide any insights for managers or photographers regarding how to improve their images based on CNN generated features. Furthermore, Airbnb property images may be appealing for more reasons than image quality. Image quality captures only the vertical differentiation among images, but images can also differ horizontally in terms of taste. In an effort to provide actionable insights, we evaluate the relationship between property demand and 12 key image attributes identified in the photography literature. We first explain these features and then theorize their impact on property demand.

3.6.1 Interpretable Image Features

Multiple image attributes may affect consumers' perceptions and choices. To investigate which image attributes correlate most strongly with Airbnb property demand, this section defines 12 key dimensions along which photographs can be compared and categorized. We use the art and photography literature to define the image attribute dimensions that are most relevant to property photos. Then, drawing from studies in this literature that pertain to the role of images in viewer perception, we theorize how each attribute could affect property demand.

The photography literature highlights 12 image attributes across three components: composition, color, and the figure-ground relationship (Datta et al. 2006; Freeman 2007; Wang et al. 2013). The features affect both the vertical quality of the photographs and horizontal qualities that might affect a property's appeal to potential guests who browse the images.

Component: Composition

Composition is the way in which visual elements are arranged, which guides the viewer's eyes to certain elements of the photograph (Freeman 2007). Expert photographers compose images that enable viewers to quickly identify the element that is the center of focus (Grill and Scanlon 1990). The most appropriate compositional attribute depends on the context; the following three compositional techniques are relevant for real estate photography.

Attribute 1: Diagonal Dominance. A photographer can guide the viewer's eyes with leading lines, and in a rectangular frame, the two diagonals of the photograph are the longest possible straight lines. In a diagonally dominant photograph, the most salient visual elements are placed close to the two diagonals (Grill and Scanlon 1990). Diagonal dominance creates a perception of spaciousness, so we predict a positive relationship between diagonal dominance and property demand. In Figure 3, it is likely that viewers will perceive that the image on the right represents a larger room than the one on the left.

Figure 3. Comparing Images on Diagonal Dominance



Attribute 2: Rule of Thirds (ROT). An image can be divided into nine equal parts with (imaginary) horizontal and vertical third lines. The ROT states that the main visual elements should be placed along the imaginary third lines or close to the four intersections (Krages 2005). These off-center focal points introduce movement into the photograph, engaging the viewer and making the image aesthetically pleasing and dynamic (Meech 2004), so we predict a positive relationship between the use of the ROT and property demand. In Figure 4, the image on the right follows the ROT better than the image on the left. Hence, when looking at the image, the viewer's attention first goes to the bed and then its counterpoint—the other vertical third line. By contrast, in the image on the left, the focus and key objects are not obvious.

Figure 4. Comparing Images on the ROT



Attributes 3 and 4: Visual Balance of Intensity and Visual Balance of Color. Visual balance is the symmetry of visual elements (in this case, intensity and color) within an image (Krages 2005). The more symmetrical the elements within the image, the greater the visual balance; the extreme case being perfect symmetry. Humans subconsciously consider visual balance to be aesthetically pleasing, and symmetry increases visual interest (Arnheim 1974; Bornstein et al. 1981). Visually balanced real estate images give viewers a feeling of order and tidiness, minimizing the cognitive demand required to process the images (Kreitler and Kreitler 1972; Machajdik and Hanbury 2010). We predict a positive relationship between visual balance and property demand. In Figure 5, the image on the right is more visually balanced than the one on the left, so the image on the right can be processed very quickly and provides a sense of order and cleanliness.

Figure 5. Comparing Images on Visual Balance

Image without Visual Balance



Image with Visual Balance



Component: Color

Color, one of the most significant elements in photography, affects the viewer's emotional arousal. Building on past research, Gorn et al. (1997) identified two dimensions of arousal: from boredom to excitement, and from tension to relaxation. Excitement is preferred to boredom, and relaxation is preferred to tension. Three dimensions of color—hue, saturation (chroma), and brightness (value)—can affect the level of arousal. Each dimension has been widely studied in the marketing literature, particularly in the contexts of web design, product packaging design, and advertisement design (Gorn et al. 2004; Miller and Kahn 2005). We add another attribute, image clarity, that is affected by the combination of the aforementioned three.

Attribute 5: Warm Hue. Hues (e.g., red, green) are believed to be a major driver of emotion. The warmth in an image is affected by the actual colors of the subject, but the photographer can also manipulate hue by warming up or cooling down a picture during post-processing. Warm hues (such as red and yellow) elicit higher levels of excitement (Gorn et al. 2004; Valdez and Mehrabian 1994), while cool hues (such as blue and green) elicit higher levels of relaxation. Hence, we predict a positive relationship between warm hues and property demand. In Figure 6, we present a cool photo of a living room on the left and a warm photo of a living room on the right.

Figure 6. Comparing Images on Warm Hue

Image with Cool Colors

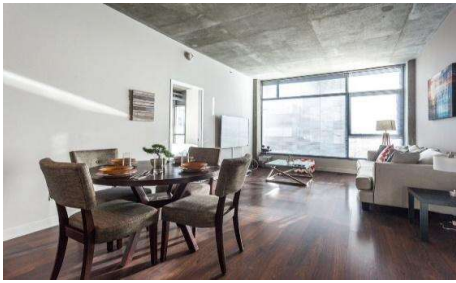


Image with Warm Colors



Attribute 6: Saturation. Saturation refers to the richness of color. Highly saturated images are colorful, while weakly saturated images contain low levels of pigmentation. Saturation is positively associated with happiness and purity; less-saturated colors are associated with sadness and distress (Valdez and Mehrabian 1994; Gorn et al. 2004). Thus, we predict that real estate images with saturated colors can induce positive emotions in viewers and that there is a positive relationship between saturation and property demand. To illustrate the difference in emotional arousal, Figure 7 displays two images of the same room with different levels of saturation.

Figure 7. Comparing Images on Saturation

Image with Low Saturation



Image with High Saturation



Attributes 7 and 8: Brightness and the Contrast of Brightness. The photography literature identifies two image attributes regarding image illumination: brightness and its contrast. Brightness is the level of overall illumination; viewers prefer bright images as they induce a sense of relaxation but do not affect the level of excitement (Valdez and Mehrabian 1994; Gorn et al. 1997). Furthermore, sufficient illumination makes the content of an image clear to viewers because images convey information through pixel brightness, so we predict a positive relationship between brightness and property demand. Meanwhile, the contrast of brightness is the variance in the illumination and describes whether the illumination is evenly distributed over the image with a smooth flow. In other words, a low contrast of brightness indicates an even distribution of illumination across the photograph. An uneven distribution (i.e., high contrast) of brightness may induce a feeling of harshness, so we predict a negative relationship between the contrast of brightness and property demand. For example, in

Figure 8, the image on the right has a higher level of brightness and more uniform illumination (i.e., lower contrast of brightness) than the image on the left.

Figure 8. Comparing Images on Illumination (Brightness and its Contrast)

Image with Low Brightness and High Contrast



Image with High Brightness and Low Contrast



Attribute 9: Image Clarity. Clarity reflects the intensity of hues in the HSV (i.e., hue, saturation, value) space (Levkowitz and Herman 1993). An image is “dull” if it is dominated by desaturated colors or has near-zero hue intensities in some color channels (He et al. 2011). Amateur photographers often shoot dull photos, inducing a so-called haze effect in which some parts of the image look unclear and ill-focused. By contrast, images with “clear” color reduce the friction in information transfer, so we predict a positive relationship between image clarity and property demand. In Figure 9, the photo on the right has higher clarity than the one on the left.

Figure 9. Comparing Images on Clarity

Image with Dull Color



Image with Clear Color



Component: The Figure-Ground (FG) Relationship

Attributes 10, 11, and 12: Area Difference, Color Difference, and Texture Difference. The FG relationship within an image is evaluated in relation to three aspects—area, color, and texture. The principle of the FG relationship is one of the most basic laws of perception and is used extensively by expert photographers to plan their photographs. In visual art, the *figure* refers to the key region (i.e., foreground), and the *ground* refers to the background. The FG relationship describes the separation between the figure and ground. Gestalt theory states that objects that share visual characteristics, such as size, color, and texture, are seen as belonging together (Arnheim 1974). Hence, the figure is more salient if it differs from the ground in size, color, and texture. In advertising research, consumers pay more attention to images with clear FG relationships (Larsen et al. 2004;

Schloss and Palmer 2011), and we predict a positive relationship between FG separation and property demand. Figure 10 presents one set of images in which the figure is clearly separable from the ground and another set in which the separation is not obvious.

Figure 10. Comparing Images on the Figure-Ground Relationship

10a. Clear Separation of Figure from Ground



10b. Unclear Separation of Figure from Ground



3.6.2 Measurement of Image Attributes and Related Statistics

Table 6 summarizes and briefly describes the 12 attributes.

Table 6. The 12 Image Attributes and Their Descriptions

COMPONENT	ATTRIBUTE	DESCRIPTION
Composition	1 Diagonal Dominance	Alignment of the key objects with the diagonals
	2 Visual Balance-Intensity	Symmetry of key objects around the imaginary vertical central line
	3 Visual Balance-Color	Symmetry of colors around the imaginary vertical central line
	4 Rule of Thirds	Alignment of the key objects with the intersections of two imaginary horizontal and two vertical lines
Color	5 Warm Hue	Dominance of warm colors (e.g., red and yellow) over cold colors (e.g., blue and purple)
	6 Saturation	The richness/vividness of image colors
	7 Brightness	Overall level of illumination
	8 Contrast of Brightness	The variance in the illumination flow across the image.
Figure-Ground Relationship	9 Image Clarity	Whether image colors have sufficient intensity
	10 Size Difference	Difference in area between the figure and ground
	11 Color Difference	Difference in color between the figure and ground
	12 Texture Difference	Difference in texture between the figure and ground

Statistics of Image Attributes

We measure the 12 attributes using computer vision algorithms; detailed methods appear in the appendix. In general, an image-processing task involves the segmentation of an image into patches, detection of salient regions, and computation involving those regions. We compare the attributes across three groups of property images:

Group LQ: All low-quality images (all unverified; $N = 368,626$)

Group HQ_UN: All high-quality unverified images ($N = 69,380$)

Group HQ_V: All verified images (all high-quality; $N = 72,608$)

The first three columns of Table 7 present the means (standard deviations) for the image attributes by group. The fourth (rightmost) column presents the difference between the means of the HQ_UN and HQ_V groups along with the two-sample t -statistics in parentheses; statistically significant differences (at a 5% significance level) are bolded.

The low-quality images rate lower than the high-quality images on all image attributes except for the contrast of brightness, which is theorized to have a negative relationship with the property demand. More interestingly, the unverified high-quality images also perform significantly worse than the verified high-quality images on most attributes. We conclude that there is a systematic difference in the high-quality images taken by Airbnb photographers versus by other photographers.

Table 7. Summary Statistics: Mean (Standard Deviation) of Image Attributes of Verified vs. Unverified High-Quality Images

COMPONENT	IMAGE ATTRIBUTE	LQ	HQ_UN	HQ_V	HQ_V
		Low Quality 368,626 Obs.	High-Quality Unverified 69,380 Obs.	High-Quality Verified 72,608 Obs.	V.S. HQ_UN
		Mean (Std. Dev.)	Mean (Std. Dev.)	Mean (Std. Dev.)	Difference (t -statistic)
Composition ¹⁰	Diagonal Dominance	-0.342 (0.160)	-0.281 (0.109)	-0.236 (0.081)	0.045*** (88.56)
	Visual Balance of Intensity	-0.865 (0.110)	-0.774 (0.103)	-0.757 (0.105)	0.017*** (30.78)
	Visual Balance of Color	-59.281 (19.460)	-53.093 (15.509)	-50.096 (15.070)	2.997*** (36.93)
	Rule of Thirds	-0.147 (0.082)	-0.089 (0.045)	-0.089 (0.047)	0.0003 (1.23)
Color	Warm Hue	0.738 (0.230)	0.751 (0.208)	0.789 (0.181)	0.038*** (36.77)
	Saturation	59.023 (37.528)	73.942 (31.300)	73.683 (26.929)	-0.259 (0.87)
	Brightness	136.029 (32.488)	154.212 (27.558)	175.802 (22.593)	21.590*** (161.75)

¹⁰ Note that the composition measurements are negative because they reflect distances, and we subtract all distances from zero to preserve the absolute magnitude when the direction is reversed. A higher value (i.e., less negative) suggests better performance on that composition attribute. For example, a higher value of diagonal dominance suggests that the image is more diagonally dominant.

Figure-Ground Relationship	Contrast of Brightness	60.601 (13.628)	58.029 (13.056)	53.996 (12.990)	-4.033*** (58.33)
	Image Clarity	0.324 (0.232)	0.413 (0.217)	0.595 (0.195)	0.182*** (166.38)
	Size Difference	-0.405 (0.181)	-0.181 (0.188)	-0.140 (0.153)	0.041*** (45.16)
	Color Difference	23.090 (20.056)	33.054 (17.552)	39.063 (15.580)	6.009*** (68.29)
	Texture Difference	0.043 (0.033)	0.057 (0.026)	0.059 (0.018)	0.002*** (16.92)

Note: SD in parentheses for the first three columns; t-statistics in parentheses for the rightmost column; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

3.6.3 Relating the Interpretable Image Attributes to Property Demand

The 12 human-interpretable attributes have well-studied relationships with image quality, but the tastes of Airbnb consumers do not necessarily align perfectly with image quality. The lack of exogenous variation in the 12 attributes prevents us from drawing causal conclusions. Nevertheless, the DID analysis may offer valuable exploratory insights. Table 8 presents the results from the estimation of Eq. 5:¹¹

$$\begin{aligned}
DEMAND_{itcym} = & INTERCEPT + \alpha TREATIND_{it} + \mu IMAGE_COUNT_{it} \\
& + \rho_1 BATHROOM_PHOTO_RATIO_{it} + \rho_2 BEDROOM_PHOTO_RATIO_{it} \\
& + \rho_3 KITCHEN_PHOTO_RATIO_{it} + \rho_4 LIVING_PHOTO_RATIO_{it} \\
& + \eta IMAGE_ATTRIBUTES_{it} + \lambda CONTROLS_{it} + SEASONALITY_{cym} + PROPERTY_i + \varepsilon_{it}
\end{aligned} \tag{5}$$

Recall that we theorized that the contrast of brightness has a negative effect on property demand, and the remaining 11 image attributes have positive effects. The coefficients in Table 8 are of the expected sign, that is, significantly negative for the contrast of brightness and significantly positive for the other attributes. We can conclude that, controlling for all other variables, the 12 image attributes are significantly correlated with property demand.

Among the four composition attributes, the largest coefficients belong to the visual balance of color and visual balance of intensity. Both features describe the symmetry of the image, which is determined by both the physical arrangement of elements in the photograph and the photographer's position from which the image is taken.

Among the five color attributes, the largest magnitudes of coefficients belong to image clarity and the contrast of brightness. Image clarity is obviously desirable; images that do not convey information clearly cannot make elements of the property appear attractive to the consumer. One would expect most skilled photographers to prioritize image clarity, but Table 7 shows a significant difference in image clarity between the unverified and verified high-quality images; unsurprisingly, the verified photos score almost twice as high as the unverified low-quality images. Meanwhile, a low value is usually preferred for the contrast of brightness, though several hosts

¹¹ For ease of understanding, we use standardized image attributes (i.e., variables are normalized to zero-mean and unit-variance).

have complained (on the Airbnb host forums) that the contrast of brightness is so low in the verified photos that they appear washed out, which might drive guests away.¹² Our results do not find evidence to support this concern. The lower contrast of brightness in the verified photos (Table 7) may actually make property images more appealing to viewers.

Among the three FG relationship features, the most prominent one is the size difference, as suggested by its largest coefficient. When the figure has a much larger size than the background, it is better able to capture the viewer's attention. A photographer's ability to manipulate the size difference may be physically constrained; for instance, if there is little size difference between the two, it will be hard to separate the figure from the ground.

With the inclusion of the 12 image attributes, the coefficient of the key variable *TREATIND* falls to 1.721 and is statistically insignificant (Table 7), suggesting that the treatment effect is associated with the ability of Airbnb professional photographers to achieve more optimal values of the 12 interpretable attributes.

Table 8. DiD Model: Controlling for Interpretable Image Attributes

COMPONENT	VARIABLES	ESTIMATES	Robust S.E.
	<i>TREATIND</i>	1.721	6.575
	Property (Non-Photo) Characteristics		
	<i>log REVIEW_COUNT</i>	9.279***	0.927
	<i>NIGHTLY_RATE</i>	-0.194***	0.0325
	<i>INSTANT_BOOK</i>	3.245*	1.351
	<i>CLEANING_FEE</i>	0.0955***	0.0185
	<i>MAX_GUESTS</i>	0.159	1.093
	<i>RESPONSE_RATE</i>	0.0946*	0.0424
	<i>RESPONSE_TIME (minutes)</i>	-0.000203	0.00158
	<i>MINIMUM_STAY</i>	0.159	0.130
	<i>SECURITY_DEPOSIT</i>	0.00189	0.00208
	<i>SUPER_HOST</i>	3.614*	1.518
	<i>BUSINESS_READY</i>	2.182*	0.971
	<i>CANCELLATION_STRICT</i>	1.670	1.256
	<i>HAS_RATING</i>	14.02	11.76
	<i>HAS_RATING</i> × <i>COMMUNICATION</i>	0.0507	1.393
	<i>HAS_RATING</i> × <i>ACCURACY</i>	0.156	1.200
	<i>HAS_RATING</i> × <i>CLEANLINESS</i>	-0.879	1.112
	<i>HAS_RATING</i> × <i>CHECKIN</i>	-2.198	1.472
	<i>HAS_RATING</i> × <i>LOCATION</i>	-0.473	1.156
	<i>HAS_RATING</i> × <i>VALUE</i>	2.059	1.165
	<i>AFTER</i> × <i>POOL</i>	6.789	4.455
	<i>AFTER</i> × <i>BEACH</i>	-9.047	10.31
	<i>AFTER</i> × <i>AC</i>	1.474	2.410
	Property Image Characteristics		
	<i>log IMAGE_COUNT</i>	4.530*	1.824
	<i>BATHROOM_PHOTO_RATIO</i>	3.431	8.526

¹² See discussion at <https://airhostsforum.com/t/airbnb-plus-program-anyone-else-who-thinks-the-new-photos-suck/30120>.

		BEDROOM_PHOTO_RATIO	0.282	7.642
		KITCHEN_PHOTO_RATIO	15.57	11.28
		LIVINGROOM_PHOTO_RATIO	-10.39	8.484
12 Human-Interpretable Image Attributes				
Composition	1	DIAGONAL_DOMINANCE	2.516**	0.945
	2	VISUAL_BALANCE_INTENSITY	4.618***	1.350
	3	VISUAL_BALANCE_COLOR	8.869***	2.143
	4	RULE_OF_THIRDS	3.537**	1.106
Color	5	WARM_HUE	4.715*	2.363
	6	SATURATION	3.920*	1.927
	7	BRIGHTNESS	3.434*	1.679
	8	CONTRAST_OF_BRIGHTNESS	-4.897*	2.411
	9	IMAGE_CLARITY	6.212**	2.175
Figure-Ground Relationship	10	SIZE_DIFFERENCE	3.807*	1.541
	11	COLOR_DIFFERENCE	2.728*	1.372
	12	TEXTURE_DIFFERENCE	2.313*	1.090
Fixed Effect			Property	
Seasonality			City-Year-Month	
Num. Observations			76,901	
R-squared			0.6670	
Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.				

4. Discussion and Conclusions

This study employs a DiD analysis combined with the PSW method to investigate Airbnb's photography program. We find that the properties that acquired verified photos had an 8.98% higher occupancy rate, translating into an average increase of \$3500.30 in annual revenue, than the properties without verified photos. The treatment coefficient decreased by 41% when we controlled for image quality, suggesting that a sizeable portion (but not all) of the coefficient of verified photos on property demand is attributable to the high quality of the professional images.

Furthermore, we shed light on what makes a good image for an Airbnb property. We show that the verified and unverified photos differ significantly on 12 interpretable image attributes identified by the photography literature. We theorize the impact of each of the 12 attributes on property demand, and we find supportive evidence for all 12 hypothesized relationships. We cannot draw causal conclusions from our observational study, but our findings nevertheless may inform content-engineering strategies among Airbnb photographers, hosts who take their own photographs, and real estate marketers more broadly.

We use the PSW method to mitigate the effect of self-selection on our DID analysis. There are, however, possible reasons that may cause risk to the PSW and DiD analyses: in a related study, motivated by our findings and leveraging the foundations of our model, Zhang et al. (2019) investigate the reason as to why a large proportion of Airbnb hosts do not adopt high quality images even when Airbnb offers them for free. They find that high-quality images lead to higher expectations for the property and level of service. If guests hold inappropriately high expectations, they are more likely to be dissatisfied and less likely to write a review, if such expectations are not met. In the long run, a property may face lower demand from its relatively low number of

accumulated reviews. That is, the service quality and property quality, affect hosts' verified photo adoptions, and hence beside other factors, create a self-selection issue for our study. As discussed in Section 3.2, our PSW model helps to mitigate the issue, through controlling for *Response Rate* and *Response Time*, and a rich set of property characteristics. Additionally, note Zhang et al. (2019) finds that the direct effect of image quality on demand is always positive, in the short- and long- term for all properties (consistent with our long-term robustness test, see V.7 of the appendix). The negative effect of images' quality on property demand comes *indirectly* through a lower rate of review accumulation for properties using high-quality images. Similar to Zhang et al. (2019), we control for the number of reviews in the demand model.

The PSW approach accounts for and matches properties on observed characteristics, but unobserved aspects of property quality may also affect both the property demand and treatment likelihood. We assess the robustness of our results to unobserved factors with two follow-up analyses: a sensitivity analysis (Rosenbaum Bounds Test, Section 3.4.2) and a DiD estimation with the pre-treated units (i.e., those that acquired verified photos before January 2016) as the control group, such that any unobservables that drive self-selection should affect the treatment and control groups equally (Section 3.4.3; Section V of the appendix). The results increase our confidence that the PSW method adequately mitigated the endogeneity concern.

There are a few limitations to this research. First, our trained CNN classifier has a relatively high accuracy of 90.4%, but it is not perfectly accurate;¹³ future studies may further improve the classifier's performance with more training data. Second, we ignore an important element of the user search process on Airbnb. Typically, a potential guest browses several properties that appear on an Airbnb search page based on user-specified criteria (e.g., location, dates,). A single image appears for each property on the search result page and may influence which properties the guest chooses to evaluate further (at which point the guest could see all images available for the property). Without access to consumer search processes, we cannot explicitly incorporate relevant information into our analysis, but future research may pursue such an analysis as more data (e.g., on the search process and transaction) become available to researchers. Lastly, as noted previously, our observational data do not enable us to make causal claims about the relationship between image features and property demand. Future research can investigate the causal impact of image attributes on property revenue if the data and context allow (e.g., in a randomized controlled trial).

This paper relates property demand to both high-level and low-level dimensions of image features. Certain industries could benefit from the documented insights regarding the 12 image attributes. For example, home rental markets, such as Airbnb and VRBO, could reduce the issue of quality uncertainty by incentivizing their hosts to display high-quality property images. In the related industry of real estate (Zillow.com, Redfin.com,

¹³ In our analyses, the *IMAGE_QUALITY* variable in the econometric model is the mean of the quality labels for the property's set of images. Though some images likely were misclassified during the machine learning step, some misclassifications should cancel out in the computed mean quality, leading to a relatively accurate mean value.

RE/MAX, etc.), a platform such as Zillow.com could use our results to improve listing images. Finally, our study is among the first to investigate the difference in demand generated by properties with verified and unverified photos and to examine the correlations between identifiable image attributes and property demand.

References

- Angrist, J., and Pischke, J. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.
- Arnheim, R. 1974. *Art and Visual Perception: A Psychology of the Creative Eye*. Berkeley: University of California Press.
- Athey, S., and Imbens, G.W. 2006. Identification and Inference in Nonlinear Difference-in-Differences Models. *Econometrica* **74**(2): 431–497.
- Austin, P.C., and Stuart, E.A. 2015. Moving towards Best Practice When Using Inverse Probability of Treatment Weighting (IPTW) Using the Propensity Score to Estimate Causal Treatment Effects in Observational Studies. *Statistics in Medicine* **34**(28): 3661–3679.
- Autor, D. 2003. Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing. *Journal of Labor Economics* **21**(1): 1–42.
- Berry, S., Levinsohn, J., and Pakes, A. 1995. Automobile Prices in Market Equilibrium. *Econometrica* **63**(4): 841–890.
- Bertrand, M., Karlan, D., Mullainathan, S., Shafir, E., and Zinman, J. 2010. What's Advertising Content Worth? Evidence from a Consumer Credit Marketing Field Experiment. *The Quarterly Journal of Economics*, **125**(1): 263–306.
- Bornstein, M.H., Ferdinandsen, K., and Charles, G.G. 1981. Perception of Symmetry in Infancy. *Developmental Psychology* **17**(1): 82–86.
- Datta, R., Joshi, D., Li, J., and Wang, J.Z. 2006. Studying Aesthetics in Photographic Images Using a Computational Approach. *ECCV* **3953**: 288–301.
- DiPrete, T.A., and Gangl, M. 2004. Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments. *Sociological Methodology*, **34**: 271–310.
- Freeman, M. 2007. *The Photographer's Eye: Composition and Design for Better Digital Photos*. (1st ed.). Focal Press.
- Gorn, G.J., Chattopadhyay, A., Sengupta, and Tripathi, J.S. 2004. Waiting for the Web: How Screen Color Affects Time Perception. *Journal of Marketing Research* **41**(2): 215–225.
- Gorn, G.J., Chattopadhyay, A., Yi, T., and Dahl, D.W. 1997. Effects of Color as an Executional Cue in Advertising: They're in the Shade. *Management Science* **43**(10): 1387–1400.
- Grill, T., and Scanlon, M. 1990. *Photographic Composition*. Watson-Guptill.
- Hagtvedt, H., and Patrick, V.M. 2008. Art Infusion: The Influence of Visual Art on the Perception and Evaluation of Consumer Products. *Journal of Marketing Research* **45**(3): 379–389.
- He, K., Sun, J., and Tang, X. 2011. Single Image Haze Removal Using Dark Channel Prior. *Pattern Analysis and Machine Intelligence. IEEE Transactions* **33**(12): 2341–2353.
- Heckman, J., Ichimura, H., and Todd, P. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *The Review of Economic Studies* **64**(4): 605–654.
- Krages, B. 2005. *Photography: The Art of Composition*. Allworth Press, USA.
- Kreitler, H., and Kreitler, S. 1972. *Psychology of the Arts*. Duke University Press.
- Krizhevsky, A., Sutskever I., and Hinton, G.E. 2012. Imagenet Classification with Deep Convolutional Neural Networks. In *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, Vol. 1, (pp. 1097–1105).

- Larsen, V., Luna, D., and Peracchio, L.A. 2004. Points of View and Pieces of Time: A Taxonomy of Image Attributes. *Journal of Consumer Research* **31**(1): 102–111.
- Levkowitz, H., and Herman, G.T. 1993. GLHS: A Generalized Lightness, Hue and Saturation Color Model. *CVGIP: Graphical Models and Image Processing* **55**(4): 271–285. doi:10.1006/cgip.1993.1019.
- Li, J., Moreno, A., and Zhang, D. 2016. Pros vs Joes: Agent Pricing Behavior in the Sharing Economy. Available at SSRN: <https://ssrn.com/abstract=2708279>.
- Liu, L., Dzyabura, D., and Mizik, N. 2020. Visual Listening In: Extracting Brand Image Portrayed on Social Media. *Marketing Science* **39**(4): 669–686. <https://doi.org/10.1287/mksc.2020.1226>
- Machajdik, J., and Hanbury, A. 2010. Affective Image Classification Using Features Inspired by Psychology and Art Theory. In *Proceedings of the International Conference on Multimedia* (pp. 83–92) ACM.
- Malik, N., Singh, P. V., and Srinivasan, Kannan, 2019. A Dynamic Analysis of Beauty Premium (February 26, 2019). Available at SSRN: <https://ssrn.com/abstract=3208162> or <http://dx.doi.org/10.2139/ssrn.3208162>
- Malik, N., and Singh, P.V. 2019. Deep Learning in Computer Vision: Methods, Interpretation, Causation and Fairness, *Tutorials in Operations Research* (pp. 73–100).
- Manchanda, P., Packard, G., and Patabhiramaiah, A. 2015. Social Dollars: The Economic Impact of Customer Participation in a Firm-Sponsored Online Customer Community. *Marketing Science* **34**(3): 367–387.
- Meech, S. 2004. *Contemporary Quilts: Design, Surface and Stitch*. Batsford: London, UK.
- Meyers-Levy, J., and Peracchio, L. A. 1992. Getting an Angle in Advertising: The Effect of Camera Angle on Product Evaluations. *Journal of Marketing Research* **29**: 454–461.
- Miller, E.G., and Kahn, B.E. 2005. Shades of Meaning: The Effect of Color and Flavor Names on Consumer Choice. *Journal of Consumer Research* **32**(1): 86–92.
- Mitchell, A., and Olsen, J.O. 1981. Are Product Attribute Beliefs the Only Mediator of Advertising Effects on Brand Attitude? *Journal of Marketing Research* **18**: 318–332.
- Narang, Unnati & Shankar, Venkatesh. (2019). Mobile App Introduction and Online and Offline Purchases and Product Returns. *Marketing Science*. 38. 10.1287/mksc.2019.1169.
- Nevo, A. 2000. Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry. *The RAND Journal of Economics* **31**(3): 395–421.
- Netzer, O., Lemaire, A., & Herzenstein, M. (2019). When Words Sweat: Identifying Signals for Loan Default in the Text of Loan Applications. *Journal of Marketing Research*, 56(6), 960–980.
- Peracchio, L.A., and Meyers-Levy, J. 1994. How Ambiguous Cropped Objects in Ad Photos Can Affect Product Evaluations. *Journal of Consumer Research* **21**: 190–204.
- PwC Report. 2015. *Consumer Intelligence Series: The Sharing Economy*.
- Rosenbaum, P.R. 1993. Hodges-Lehmann Point Estimates of Treatment Effect in Observational Studies. *Journal of the American Statistical Association* **83**: 1250–1253.
- Rosenbaum, P. R. 2002. *Observational Studies*. (2nd ed.). New York: Springer.
- Rosenbaum, P.R., and Rubin, D.B. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **70**(1): 41–55.
- Schloss, K. B., and Palmer, S. E. 2011. Aesthetic Response to Color Combinations: Preference, Harmony, and Similarity. *Attention, Perception, & Psychophysics* **73**(2) 551–571. Available at <http://doi.org/10.3758/s13414-010-0027-0>.

- Scott, L.M. 1994. Images in Advertising: The Need for a Theory of Visual Rhetoric. *Journal of Consumer Research* **21**: 252–273.
- Simonyan, K., and Zisserman, A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. Presented at *International Conference on Learning Representations 2015 (ICLR 2015)*.
- Sun, M., and Zhu, F. 2013. Ad Revenue and Content Commercialization: Evidence from Blogs. *Management Science* **59**(10): 2314–2331.
- Ufford, S. 2015. The Future of the Sharing Economy Depends on Trust. Available at <http://www.forbes.com/sites/theyec/2015/02/10/the-future-of-the-sharing-economy-depends-on-trust/#21a3370b58ff>.
- Valdez, P., and Mehrabian, A. 1994. Effects of Color on Emotions. *Journal of Experimental Psychology: General* **123**(4): 394–409.
- Wang, Q., Li, B., and Singh, P.V. 2018. Copycats versus Original Mobile Apps: A Machine Learning Detection Method and Empirical Analysis. *Information Systems Research*, 29(2): 273–291.
- Wang, X., Jia, J., Yin, J., and Cai, L. 2013. Interpretable Aesthetic Features for Affective Image Classification. In *IEEE International Conference on Image Processing* (pp. 3230–3234). Melbourne, Australia.
- Wang, K., and Goldfarb, A. 2017. Can Offline Stores Drive Online Sales?. *Journal of Marketing Research* **54**(5): 706–719.
- Winkler, R., and Macmillan, D. 2015. The Secret Math of Airbnb's \$24 Billion Valuation. Available at <https://www.wsj.com/articles/the-secret-math-of-airbnbs-24-billion-valuation-1434568517>.
- Zhou, B., Lapedriza, A., Xiao, J., Torralba, A., and Oliva, A. 2014. Learning Deep Features for Scene Recognition Using Places Database. In *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Vol. 1. December 2014.
- Zhang, Shunyuan, Nitin Mehta, Param Vir Singh, and Kannan Srinivasan. Can Lower-quality Images Lead to Greater Demand on AirBnB?. Technical report, Working Paper, Carnegie Mellon University, 2019.

**An Appendix to
What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features**

Table of Contents

An Appendix to	34
I. Classifying Image Quality Using a Deep Learning-based Classification Model	36
1. Training Set Construction	36
2. Image Quality Classifier Training.....	38
II. Algorithm and Concepts for Image Attribute Computation	42
1. Visual (Image) Saliency	42
<i>Definition</i>	42
<i>Calculation</i>	42
2. Salient Region	42
<i>Definition</i>	42
<i>Detection</i>	42
3. FG	43
<i>Definition</i>	43
<i>Detection</i>	43
III. Measurement of Interpretable Image Attributes	45
a. Composition.....	45
b. Color	46
c. FG Relationship	46
IV. Room Type Classification.....	48
Training Set of Indoor/Outdoor Photos	48
Training a Room Type Classifier on a Collected Training Set.....	49
V. Robustness Checks, Empirical Extension, and Exclusion of Alternative Explanations	51
1) Validating the Propensity Score Method	51
2) Sensitivity Analysis of the Propensity Score Method (Rosenbaum Bounds Test).....	52
3) Addressing the Possibility of an Inflated Long-Term Effect	54

4) Adding Interaction Terms with Meaningful Amenities.....	59
5) Testing Changes in the Property or the Host's Unobserved Quality (via Multidimensional Ratings) ..	61
6) Alternative Specifications: Logging All Numeric Variables.....	63
7) Including Pre-treated Units in the Control Group	64
8) Additional Analyses.....	67
9) Testing Host Behavior – Changes in the Number of Open Days	70
VI. Data Description and Sample Construction.....	73
VII. Example of Property Images with 1 SD of Improvement in Key Image Features.....	75

I. Classifying Image Quality Using a Deep Learning-based Classification Model

1. Training Set Construction

Image Quality Assessment Survey on Amazon Mechanical Turk

We describe the steps for creating the dataset to train our classifier using Amazon Mechanical Turk (AMT). AMT is a platform of Amazon Web Services that enables users to outsource small tasks to a large group of workers at a relatively low cost. It has been widely used for human intelligence tasks, such as data collection and data cleaning. As a crowdsourcing method, AMT has been found to be quite efficient and accurate (Casalboni 2015; Laws et al. 2011).

To construct a labeled training set for supervised learning, we selected 3,000 Airbnb property images from our dataset and used AMT to tag each image based on its quality. In the selection of images for AMT tagging, full random sampling was not optimal, as we did not have information on the distribution of image quality beforehand. We used stratified random sampling to ensure that a sufficient number of images were evaluated and labeled for different categories of image quality. A random sample stratified by a crude metric of quality was necessary, as it ensures that the sample is balanced and random. We randomly selected 500 images from the pool of verified images, as these were guaranteed to be taken by professional photographers and are most likely to be of high quality. Then, from all the unverified images, a human judge chose 8,000 images that *looked bad*, and 500 images from this group were randomly sampled. From the unverified images, we chose 5,000 images that we judged to be in between *excellent* and *very bad*. Of these, 500 images were randomly sampled. Lastly, we randomly sampled 1,500 images from the entire sample. Constructing the AMT data in this way ensured that we would have a sufficient random sample of images from each stratum (i.e., subgroup of images with a certain quality). We also manually reviewed the selected images to ensure that no image was repeatedly sampled.

For the AMT tagging task, we created a survey instrument that asked the Turker to assign a score to a displayed image based on its aesthetic quality. To provide accurate guidelines for image evaluation, we borrowed instructions from professional photography forums, as well as Airbnb's guidelines for shooting good property photos. We also provided example photos. The quality measurement was based on a Likert scale from 1 to 7, in which 1 is *very bad* and 7 is *excellent*. Figure A1 shows an example question on the survey given to the Turkers. To ensure high-quality and consistent responses from the Turkers, we required them to have an approval rate of higher than 95% and to have completed at least 50 approved tasks.

Figure A1. Example of the AMT Aesthetic Quality Assessment Task

Instructions

We are interested in assessing the aesthetic quality of images. Give an aesthetic score to an image on a scale of 1-7 on how beautiful/pleasing/visually appealing you think the image is, where 1 is "very bad" and 7 is "excellent". Here is a short non-exhaustive guideline for judging images.


Reasons for a rating closer to 7:



- looks good
- visually pleasing
- well lit
- clearly shows room/house features
- conveys well what the room/house may look like in person




Note 1: Most of our images are room/house images, however there are a few pictures of outdoors (e.g., pictures of plants/gardens). The score should be given on the aesthetic quality of an image, regardless of its topic (i.e., whether it's a picture of indoor or a picture of outdoor).

Note 2: this is a subjective measure, but you can calibrate your standard by looking at our examples, to rate each image.

Examples of images with aesthetic scores of 1, 4, and 7:

example score = 1: 

example score = 4:  

example score = 7:   

Give a score to an image on a scale of 1-7 on its aesthetic quality where 1 is "very bad" and 7 is "excellent".

1	2	3	4	5	6	7
0	0	0	0	0	0	0
very bad	bad	below average	average	good	very good	excellent

Constructing the Training Set from Tagged Images

After the AMT survey, we obtained 3,000 tagged images, each of which was evaluated by five Turkers, to be pre-processed for the construction of our training set. Following previous studies on aesthetic quality (Datta et al. 2006; Datta et al. 2008; Marchesotti et al. 2011), we computed for each image, i , the mean aesthetic score, score_i , given by five Turkers. We then set two thresholds: $\theta_1 = \overline{\text{score}} + \text{gap}/2$ and $\theta_2 = \overline{\text{score}} - \text{gap}/2$. Next, we annotated image i with the term *high quality* if the average $\text{score}_i \geq \theta_1$ and with the term *low quality* if the average $\text{score}_i \leq \theta_2$. The mean score for all 3,000 images was $\overline{\text{score}} = 4.5$. Images with an average score of θ_1 and θ_2 were excluded from the training set, leaving an artificial gap between high- and low-quality images. As argued by Datta et al. (2006), the reasons for creating a gap between high- and low-quality images is that close aesthetic scores (e.g., 4.4 and 4.5) are unlikely to reflect that the images differ in aesthetic quality; rather, they indicate noise in the quality measurement process. Increasing the value of the gap makes it easier to distinguish between high- and low-quality images, but it leads to a smaller training set, as images lying within the gap are dropped. To choose an optimal value, we varied the value of the gap and selected the one that resulted in the best performance of the trained classifier on a hold-out set (Datta et al. 2006; Marchesotti et al. 2011). Using this method, we chose a gap of 0.8, which left us with a training set of 2,259 images. As 50% of the training set was selected at random, we repeated the analysis with the selected 50% and found similar results for the image labeling tasks.

2. Image Quality Classifier Training

Architecture of the Convolutional Neural Network (CNN) Framework

After a training set was constructed, the next task was to build an image quality classifier using labeled data. We applied CNN, an emerging deep learning framework that is widely applied in the field of computer vision and has been shown to perform very well for tasks, such as object recognition and image classification (Krizhevsky et al. 2012; Simonyan and Zisserman 2015).

As shown in Figure A2, a CNN model consists of a sequence of layers, each having multiple *neurons*. The number of neurons can vary from one to thousands. These neuron layers perform matrix multiplication based on an input, generating an output to serve as the input for the next layer. Both the input and output take the form of multi-dimensional matrices. The sequence of layers makes the neural network *deep*.

Images serve as the first input for the deep learning framework. In our training task, we resized all the images to 224×224 pixels, determined the pixel intensity of each image, and represented the image with a 3D array (matrix) that contains pixel information for the three channels (RGB). This was done to alleviate the computational burden and ensure that the image size aligned with the pre-trained VGG16 model (described below).

The last output layer predicts the binary label for its input, which, after passing through the whole network, is an N -dimensional vector extracted from the image. For an image in our training set (represented by IMG^k), the output layer applies a sigmoid function and predicts the label regarding the image's quality:

$$\widehat{\text{Label}}(\text{Image Quality}|IMG^k) = \begin{cases} 1 \text{ (high quality)} & \text{if } \frac{1}{1+\exp(-(X^T W_1 + W_0))} \geq 0.5 \\ 0 \text{ (low quality)} & \text{if } \frac{1}{1+\exp(-(X^T W_1 + W_0))} < 0.5 \end{cases}$$

where X^T represents the output from the layer preceding the output layer (in our model, this was the FC2 layer, which produces a 4096×1 vector), W_1 represents the weight parameters, W_0 represents the bias (a constant) connecting the preceding layer to the output layer, and $\frac{1}{1+\exp(-(X^T W_1 + W_0))}$ is the probability that the image is of high quality, given the X^T and W_1 and W_0 values for the output layer.

Throughout the CNN model, a sequence of weights on each layer define the intermediate extracted vectors from each layer, including X^T . These weights are adjusted during the training process to optimize the model's predictive performance.

Operation of Key Layers in the CNN

A few key layers determine the performance of a CNN: the convolution layer, the zero-padding layer, and the max-pooling layer. We describe these below.

Convolution Layer

The convolution layer is the most important and unique layer in the CNN. It consists of a stack of convolution filters or convolution kernels. For example, the two convolution layers in Layer Block A (shown in Figure A2) consist of 64 and 128 convolution filters, respectively. A convolution filter is a matrix in which each element represents a numeric value. For example, in Layer Block A, the convolution layers have a size of 3×3 and hence consist of nine numeric values.¹⁴ This matrix, treating an image or an intermediate input as a matrix, operates dot production by *sliding* through the input. For an input of a relatively large size (e.g., 224×224), a 3×3 convolution filter operates dot production for every 3×3 section. The convolution operation is beneficial because it reduces the dimensionality of parameters and well explores and reserves the (local) spatial relationships of the input. Regarding the second benefit, if a convolution kernel extracts an oriented edge of an object, operating this kernel on every small square (e.g., 3×3) of an image would enable the extraction of all edges with that direction from the image. Many kernels that extract edges do so in all directions, potentially constructing the contour of an object. As can be seen in Figure A2, each block consists of a varying number of convolutional filters (e.g., 64, 128, 256, and 512 filters). These kernels extract features from the input data, which represent the extracted features from the preceding layers. Toward the output layer (i.e., layers closer to the output layer) in the CNN, the filters combined together extract higher- and lower-level features. That is, the CNN can extract a hierarchical structure of related features to predict the output labels.

Zero-Padding Layer

The zero-padding layer adds numerical arrays consisting of all 0 values to the edge of an intermediate output from a layer. The size of the zero-padding layer is a hyperparameter in the CNN model. Typically, as was done in our model, the intermediate output is padded with 1*M 0 vectors on each side, causing the width and height to both increase by 1 after the zero-padding value. As zeros do not contribute to the matrix multiplication procedure, the zero-padding layer does not affect the features extracted by the layers. In addition, the zero-padding layer allows us to control the spatial size of the intermediate outputs, and it can prevent the outputs from decreasing too quickly after layers of convolution operations.

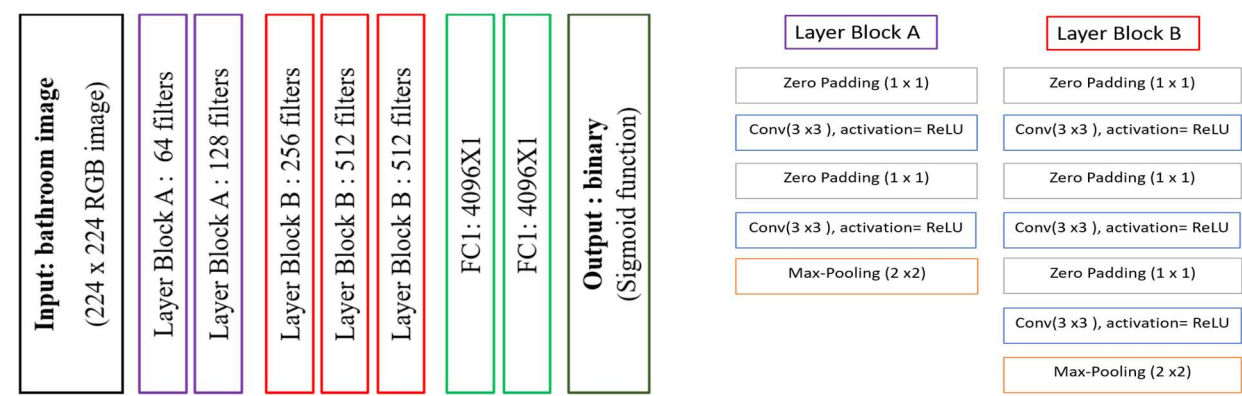
Max-Pooling Layer

Inserting a max-pooling layer between successive convolution layers is common in CNNs. A max-pooling layer is a small square filter (in our model, a 2×2 matrix). Similar to the convolution filter, a max-pooling layer applies to every 2×2 square patch in input data. Its function is to select and preserve only the maximum value in that 2

¹⁴ The size of a convolution layer is a choice of the model architecture. A 3×3 or 5×5 configuration is common.

$\times 2$ square. Adding max-pooling layers can reduce the spatial size of the intermediate features and the dimension of the trained parameters in the model, and it can help to efficiently prevent overfitting.

Figure A2. Description of the Architecture and Description of the Layers of the CNN Classifier



Filters: Indicate the number of convolution windows (i.e., number of feature maps) on each convolution layer.

Zero-padding layer: Pads the input with zeros on the edges to control the spatial size of the output. It has no impact on the predicted output.

Max-pooling: Subsampling method. A 2×2 window slides through each feature map (without overlap) at that layer, and then the maximum value in the window is selected as a representation of the window. This reduces computation and ensures translation invariance.

Training the CNN

We randomly split the dataset into training and validation sets, with 80% of the examples forming the training set and the remainder being used as the validation set. To reduce the overfitting problem in the training step, we use data augmentation and implement a real-time (i.e., during training) image transformation for each image in the training sample by randomly (1) flipping the input image horizontally, (2) rescaling the input image within a scale of 1.2, and (3) rotating the image within 20° (images being rotated by a random value between 0° and 20°). This method introduces random variation in the training sample, increases the training set size, and reduces overfitting (Krizhevsky et al. 2012).

To effectively learn features from a relatively small sample size, we apply the idea of *transfer learning*, which involved building our model on top of an existing well-trained CNN model and then fine-tuning it. As the features extracted from images are generic to some extent (e.g., almost all CNNs extract edge information at the first layer), transfer learning is quite common in deep learning and is suggested as an effective approach for dealing with problems caused by limited data (Girshick et al. 2014; Lin et al. 2015; Zhang et al. 2015). In this study, we used VGG16 (Simonyan and Zisserman 2015), as it is a conceptually simple and popular pre-trained

model. We removed the last fully connected layers from the original VGG16 model, as they contain more data-specific features, and then we added the output layer as the last layer. Figure 2A presents the architecture of our image classification model. The parameters were initialized using the pre-trained weights, except for the output layers, for which the parameters were initialized following LeCun's uniform scaled initiation method (LeCun et al. 1998). Then, we fixed the parameters on the first 25 layers and fine-tuned the model. The model was trained on the training set using a NVIDIA K80 GPU, and then performance was tested on the hold-out set at the end of the training. Optimization was performed with an adaptive method of gradient descent (Adadelata optimization; Zeiler 2012) for each mini-batch of 16 examples.

II. Algorithm and Concepts for Image Attribute Computation

In this section, we define the key concepts used in the process of image attribute computation. These key concepts include image saliency, the key/salient region, and the figure-ground (FG). We first discuss each concept's definition and then present the image algorithm used to detect, extract, or compute the concept.

1. Visual (Image) Saliency

The basic unit determining image saliency is visual saliency at the pixel level. The overall saliency score for a local patch of an image can be computed based on the pixel saliency within the local region.

Definition

Saliency describes a concept that originates from visual unpredictability. In images, it is often captured by variations in, for example, boundaries and colors. Studies on cognitive psychology and computer vision have investigated how humans process and pay attention to visual information and have found that we allocate our attention to parts of given information (e.g., the regions of an image) while cognitively ignoring other parts. Visual uniqueness is salient in the sense that it easily attracts the attention of viewers.

Calculation

In general, the models proposed for calculating visual saliency are based on local contrast to surroundings. The contrasts are determined using features, such as color, intensity, edge density, and orientation. A simple example is the *gradient* of pixel intensity. A pixel with great contrast is assigned a high saliency value.

2. Salient Region

Definition

Following the definition of visual saliency, salient regions are defined as the regions of an image that have a high overall saliency score.

Detection

The detection of a salient region involves four steps: (1) segmenting an image into local *patches*, (2) assigning a saliency score to each path, (3) merging similar patches into a *region*, and (4) finding the most salient region. These steps are discussed in greater depth below.

- 1) *Segmenting an image*: This process generally involves grouping the pixels of an image into multiple parts containing pixels that are similar to one another. Segmentation can be based on edges (detected edges are assumed to define the boundaries of objects), colors, or other factors. A segmentation algorithm is

intended to compare two intensity differences: the difference across the boundary of patches and the difference between two neighboring pixels within the same patch.

- 2) *Assigning a saliency score to a patch*: Each pixel is assigned a saliency value. The saliency score of a patch is calculated by averaging the scores of all pixels within the patch.
- 3) *Merging patches into a region*: If neighboring patches have similar colors, then they are merged into a larger region.
- 4) *Finding the most salient region*: The salient region is found by selecting the region with the highest average salient score.

3. FG

Definition

The *figure* is the foreground of an image, and the *ground* is the background. Only one figure and one ground can be detected for each image. This is different from the detection of salient regions, for which multiple regions can be detected.

Detection

The figure is detected and extracted from an image, and then the ground is defined as the rest of the image. Detection of the foreground is an extension of image segmentation, as a pixel is assigned a value of either 1 (foreground) or 0 (background).

Detailed Algorithm

We used GrabCut, a state-of-the-art model for foreground extraction (Rother et al. 2004). In this model, an image is treated as a graph, with each pixel serving as a node and pixel similarity being defined as an edge. GrabCut implements the expectation–maximization (EM) algorithm and min-cut algorithm to iteratively assign a foreground/background label to pixels and to cut the graph into two subgraphs: one representing the foreground and the other representing the background.

- 1) Initially, an arbitrary rectangle separates the image into two parts. The pixels in the rectangle are labeled “1” (foreground), and those outside it are labeled “0” (background). The initial position of the rectangle can be arbitrary. Alternatively, one can specify the rectangle’s location or hard label some pixels with good prior knowledge of where the foreground might be.
- 2) A Gaussian mixture model (GMM) is trained with the EM algorithm based on the distribution of pixel color statistics. From the GMM, we can determine the probability that each pixel belongs to a particular mode (or cluster). That is, the GMM labels a pixel as a *probable foreground* or a *probable background*.

- 3) A graph is built in which each node represents a pixel, and the edge weight between two pixels represents pixel similarity. Similar pixels will be assigned a low edge weight and vice versa. Pixel similarity can be computed based on intensity, color, or texture.
- 4) Two additional nodes are created on the graph: the source node and the sink node. All pixel nodes labeled as a probable foreground (probable background) are connected to the source node (sink node), with the edge weight between each pixel node and the source node (sink node) representing the probability that the pixel belongs to the foreground (background).
- 5) Next, the graph is cut into two parts, and the min-cut algorithm is implemented by minimizing the cost function, defined as the sum of the edge weights across all edges that are cut. The min-cut algorithm penalizes a cut if it will cause two similar pixels to be separated into two subgraphs. Intuitively, if two pixels both have a high probability of being in the foreground (background), then it is desirable for them to be labeled “1” (“0”) at the end of this iteration.
- 6) Steps 2–5 are repeated until convergence in the pixel labeling is achieved.

III. Measurement of Interpretable Image Attributes

Using the computer vision algorithm described above, we perform image processing tasks to segment images into patches and detect key/salient regions. After salient regions are detected, subsequent computation is performed to measure image attributes. This section discusses the steps for computing image attribute measurements after image processing.

a. Composition

Four image attributes are categorized during the composition step. How well an image performs concerning a particular attribute is evaluated by distance, such as the distance between two pixels. A smaller distance indicates better performance. For all four composition attributes, we compute the distance metrics and then subtract the metrics from zero.

Diagonal Dominance (Attribute 1): Diagonal dominance captures how close an image's key region is positioned to the two diagonals of the image. For an image, we first identify the key region and then measure the weighted Manhattan distance from the key region to each diagonal (Liu et al. 2010; Wang et al. 2013).¹⁵ The measurement of diagonal dominance is computed by subtracting the minimum weighted distance from zero. A greater diagonal dominance value suggests that the image is more diagonally dominant.

Rule of Thirds (ROT) (Attribute 2): We first divide an image into nine equal parts with two (imaginary) equally spaced vertical lines and two (imaginary) equally spaced horizontal lines. Then, we calculate the Euclidean distance from the centroid of the key region to each of the intersections (Wang et al. 2013). If the minimum distance is small, then the image follows the ROT, with its key region close to at least one intersection. The ROT is measured by subtracting the minimum distance from zero. If an image follows the ROT more closely, the ROT value is greater.

Visual Balance Intensity (Attribute 3): In this measurement, we split the image along its vertical central line. On each half of the image, we identify a key region and compute the distance from its centroid to the vertical central line (Liu et al. 2010). A relative distance measure is calculated by subtracting the shorter distance from the longer one and dividing the difference by the longer distance. Then, visual balance intensity is computed by subtracting the relative distance from zero. A greater value for this measure suggests that the image is more (vertically) visually balanced in terms of pixel intensity.

Visual Balance Color (Attribute 4): The color measurement of visual balance compares the left half of an image with the right half based on color. We first calculate the Euclidean distance in terms of color intensity (i.e., RGB

¹⁵ The Manhattan distance between two points on an image is measured as the number of pixels between them, with only horizontal and vertical paths allowed.

channel) between each pixel and its symmetrical pixel (across the vertical line). Then, visual balance color is computed by subtracting the average difference from zero. A greater value suggests that the image is more visually balanced in terms of color around its vertical central line.

b. Color

Five image attributes related to color are computed. The measurements are based on pixel intensity or related values (e.g., hue and saturation).

Warm Hue (Attribute 5): The warm hue measurement captures the warmth of an image, which is defined by the relative proportion of warm hues (e.g., yellow) to cool hues (e.g., green). The measurement is computed in the HSV (hue, saturation, and volume) space. Specifically, we calculate the pixel hues that fall outside the cool range (30–110) on the hue spectrum (Wang et al. 2013). If an image contains more warm hues, such as yellow and orange, it will have a greater warm hue value.

Saturation (Attribute 6): We compute pixel saturation in the HSV space by averaging the saturation values of all pixels in the image. A greater value indicates higher saturation (for example, the image contains more saturated colors).

Brightness (Attribute 7): The brightness of an image is defined as the overall level of illumination. We calculate the intensity of each pixel and then average the intensity values across all pixels in the image. A brighter image has a greater brightness value.

Contrast of Brightness (Attribute 8): Contrast is calculated as the *SD* of pixel intensity in the whole image. A lower contrast of brightness value suggests that the brightness is more evenly distributed across the image.

Image Clarity (Attribute 9): Image clarity captures the portion of hues with sufficient intensity. We measure pixel brightness on a 0–1 scale and then compute the portion of pixels whose brightness is 0.7–1 (Wang et al. 2013). A clear image has a higher image clarity value.

c. FG Relationship

The FG relationship is described as the difference between the figure and its ground in terms of three metrics: size, color, and texture. An image with a good FG relationship has a clearly separable figure and ground (i.e., greater differences).

Size Difference (Attribute 10): The size difference attribute compares the size of the figure with that of the ground. We detect the figure and ground in an image and calculate the size of each in relation to the whole image (Cheng et al. 2011). Then, the size difference is computed by subtracting the ground's size ratio from the figure's

size ratio. A greater size difference value indicates that the image has a figure occupying a relatively larger area, standing out from the ground.

Color Difference (Attribute 11): The color difference attribute captures the difference in colors between the figure and the ground. We compute the Euclidean distance between the mean color of the figure and that of the ground. A high color difference value suggests that the figure and ground contain distinct colors. In such cases, the figure can be easily distinguished from the ground.

Texture Difference (Attribute 12): Texture difference measures the difference between the figure and the ground in terms of texture, which is captured by the edge density within a local region. For the figure and ground, we use the Canny edge detector to detect the edge and then compute edge density. Then, we measure the absolute difference between the two densities. A high texture difference value suggests that the figure and ground have a clear separation based on texture.

IV. Room Type Classification

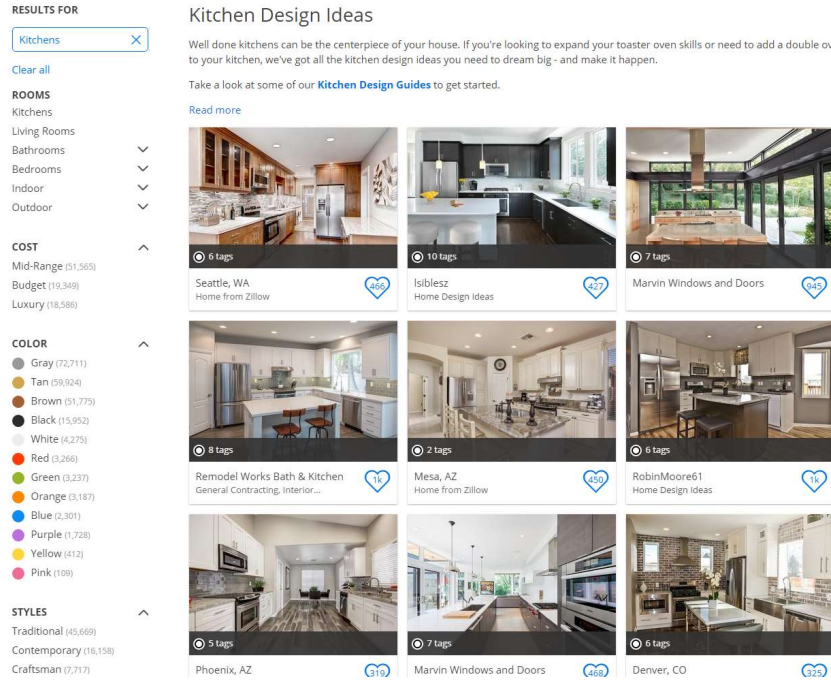
We build a deep learning model to automatically categorize the type of scene photographed. The goal is to compute the distribution of different types of rooms depicted in property images. Controlling the distribution in our demand model helps us address the concern that professional photographers know which types of places appeal more to consumers and thus present these aspects of properties.

We build a deep learning model to automatically classify the room type (bathroom, bedroom, kitchen, living room, or outdoor area) depicted in a given property image. Using transfer learning with a deep learning model that was pre-trained on a large scene classification dataset, Places205 (Zhou et al. 2014), we optimize the classifier for a dataset we collected, which consists of 54,557 images of bathrooms, 59,082 images of bedrooms, 88,030 images of kitchens, 81,819 images of living rooms, and 5,734 images of outdoor areas. The average classification accuracy is 95.05% on a hold-out set across the five categories.

Training Set of Indoor/Outdoor Photos

To train a room type classifier, we need a large number of room images labeled with a room type. We would have preferred to use original images from Airbnb.com, but the images on property web pages are not labeled, and it would incur a high labor cost to have someone (e.g., AMT) manually label images for us. Therefore, we crawled data from real estate–related websites on which a vast number of indoor/outdoor images are classified into categories. Figure A3 shows a portion of a web page displaying 23 images of kitchens at multiple properties.

Figure A3 A Real Estate Web Page Showing Kitchens



From the website, we collected images aligning with the five room types: bathroom, bedroom, living room, kitchen, and outdoor area. We then split the dataset, 80% of which was used as a training set and the remaining 20% used as a hold-out test set.

Training a Room Type Classifier on a Collected Training Set

We used the VGG16 ConvNet model, which was pre-trained on the Places205 dataset (Zhou et al. 2014) and then fine-tuned on our training set.¹⁶ The model was used for a task to classify 205 categories of places. To transfer the pre-trained model to our study, we removed the output layer in the pre-trained model and then added an output layer designed for our specific task. The added output layer was a 5*1 vector in which each element indicated the predicted probability of assigning the corresponding label. To ensure that all the probabilities add up to 1 (as we assume that a room belongs to only one category), the Softmax function was used to calculate the predicted probability (the function ensures that all the predicted probabilities add up to 1). For example, the probability that an image is assigned room type k is computed as follows:

$$\text{Prob}(\text{Room Type} = k | X^T) = \frac{x^T w_1^k + w_0^k}{\sum_{j=0}^4 x^T w_1^j + w_0^j}, \quad (1)$$

¹⁶ The pre-trained VGG16 model (including the architecture and parameters) can be accessed at <http://places.csail.mit.edu/downloadCNN.html>.

where $j = 0, \dots, 4$ represents the room type (bathroom, bedroom, kitchen, living room, or outdoor area, respectively), X^T represents the output from FC2 (the second fully connected layer), W_1^j represents the weight connecting FC2 to the j^{th} node on the output layer, and W_0^j represents the bias connecting FC2 to the j^{th} node on the output layer.

V. Robustness Checks, Empirical Extension, and Exclusion of Alternative Explanations

This section reports a series of analyses performed to test the robustness of our main results and/or exclude alternative explanations. We begin by presenting the validation of the propensity score method used in our empirical model, which included a propensity score weighting (PSW) strategy and a sensitivity assessment of unobservables.

1) Validating the Propensity Score Method

To ensure that the propensity score (PS) approach effectively eliminates potential systematic imbalances between the treatment and control groups, one needs to show that the PSs have balanced the covariates for matched or weighted samples.

We implemented a balance check, which compares, in terms of the covariates, the weighted means of the treatment group, $\bar{X}_{treatment} = \frac{\sum_{i \in treatment} \omega_i X_i}{\sum_{i \in treatment} \omega_i}$, and the control group, $\bar{X}_{control} = \frac{\sum_{i \in control} \omega_i X_i}{\sum_{i \in control} \omega_i}$. Here, X_i is a 1*M dimensional vector of the pre-treatment covariates (i.e., the covariates observed before the treatment) of unit i , and ω_i is the sample weight for unit i , computed based on the estimated PSs.

Table 2 presents the weighted group means and tests for the differences in the means for each variable X^m ($m = 1, 2, \dots, M$). As shown by the t -statistics in the table, the weighted samples are not statistically different at the 95% significance level. That is, the systematic differences in the weighed samples are negligible after performing the PSW method. Therefore, we validated that our PSW method effectively eliminated imbalances in the sample and that our weighted treatment and control groups are comparable in terms of the observed covariates that may affect the treatment selection process.

Table A9. PSW Validation: Covariate Balance Check

Variables	Weighted Means in		T-test	
	Treated	Untreated	t	p -value
<i>REVIEW_COUNT</i>	20.56	19.88	0.27	0.790
<i>IMAGE_QUALITY</i>	0.27	0.25	1.00	0.316
<i>IMAGE_COUNT</i>	14.48	15.1	-0.67	0.506
<i>NIGHTLY_RATE</i>	170.15	191.36	-1.14	0.257
<i>MINIMUM_STAY</i>	2.57	2.57	-0.00	1.000
<i>MAX_GUESTS</i>	3.5	3.67	-0.82	0.410
<i>RESPONSE_RATE</i>	92.25	91.19	0.79	0.431
<i>RESPONSE_TIME (minutes)</i>	225.12	260.98	-1.18	0.238

<i>SUPER_HOST</i>	0.15	0.11	1.05	0.292
<i>INSTANT_BOOK</i>	0.11	0.11	-0.14	0.888
<i># BLOCKED DAYS</i>	9.51	8.32	1.10	0.271
<i># RESERVATION DAYS</i>	6.62	6.74	-0.16	0.877
<i>PARKING</i>	0.5	0.49	0.09	0.929
<i>POOL</i>	0.1	0.08	0.76	0.445
<i>BEACH</i>	0.02	0.02	0.34	0.737
<i>INTERNET</i>	0.99	1	-0.58	0.563
<i>TV</i>	0.79	0.81	-0.55	0.579
<i>WASHER</i>	0.6	0.57	0.81	0.419
<i>MICROWAVE</i>	0.15	0.13	0.64	0.523
<i>ELEVATOR</i>	0.2	0.21	-0.22	0.826
<i>GYM</i>	0.11	0.13	-0.69	0.490
<i>FAMILY_FRIENDLY</i>	0.19	0.2	-0.56	0.576
<i>SMOKE_DETECTOR</i>	0.55	0.52	0.62	0.534
<i>SHAMPOO</i>	0.45	0.44	0.09	0.929
<i>BATHROOM_PHOTO_RATIO</i>	0.22	0.21	1.05	0.295
<i>BEDROOM_PHOTO_RATIO</i>	0.29	0.29	-0.26	0.795
<i>KITCHEN_PHOTO_RATIO</i>	0.1	0.1	-0.15	0.880
<i>LIVINGROOM_PHOTO_RATIO</i>	0.18	0.19	-0.52	0.602
<i>OUTDOOR_PHOTO_RATIO</i>	0.2	0.21	-0.13	0.898
<i>SECURITY_DEPOSIT</i>	202.77	225.19	-0.65	0.517
<i># OPEN DAYS</i>	21.49	22.68	-1.10	0.271
<i>CANCELLATION_STRICT</i>	0.26	0.29	-0.89	0.371
<i># BEDS</i>	1.81	1.96	-1.17	0.242
<i>APARTMENT</i>	0.6	0.61	-0.27	0.786
<i>ENTIRE_HOME</i>	0.61	0.64	-0.64	0.522

2) Sensitivity Analysis of the Propensity Score Method (Rosenbaum Bounds Test)

In the model for estimating the propensity scores, we included a rich set of variables and their interactions. The inclusion of a complete set of covariates reduced the odds that our main results would be affected by variables that were not accounted for when computing the propensity scores.

Despite the long list of included covariates for propensity scores estimation, some variables that may affect one's decision regarding participation in a professional program were omitted. As is commonly acknowledged,

PSs are computed based on observed variables; therefore, there may be a hidden bias if unobserved variables simultaneously affect the selection process (i.e., the treatment assignment) and the outcome variables. To assess the sensitivity of our estimation to potential hidden bias, we implement a widely adopted approach of sensitive analysis—the Rosenbaum bounds test (Rosenbaum 2002).

The logic of the Rosenbaum bounds analysis is as follows. Suppose the participation probability of unit i is $P_i(\text{treated}_i = 1 | x_i, u_i) = f(\beta x_i + \gamma u_i)$, where x_i and u_i are the vectors of observed and unobserved variables, respectively, and γ is 0 if there are no unobserved variables affecting the treatment selection process. Units i and j with $x_i = x_j$ have the same probability of receiving the treatment if and only if $\gamma(u_i - u_j) = 0$. Rosenbaum bounds evaluate the degree of minimum change in the odds ratio (OR) of participation due to unobservables would be required to nullify the treatment effect identified with the PS method. The estimation results inspire more confidence if a greater change in the OR caused by unobservables is needed to overturn the estimated treatment effect.

The results of the Rosenbaum bounds test are provided in Table A10. As our main DiD (Difference in Differences) analysis identified that verified photos had a positive effect on property demand, we were more concerned with potential upward (positive) than downward (negative) bias in the DiD estimator. Therefore, in Table A11, we are mostly interested in the sig+ column. The gamma results suggest that even when a hypothetical unobservable increases the OR by 1.4 times, our causal inference of the positive treatment coefficient of professional images on property demand, identified with the PS method, will be robust at the 95% confidence level (and will remain positively significant at the 90% confidence level until the gamma is increased to 1.55). The results suggest that for a positive estimated treatment effect on property demand to be overturned, the potential unobserved factors affecting the treatment assignment process would have to be large enough to increase the OR of participation by at least 60%. Moreover, if we look at Hodges–Lehmann’s estimates (Rosenbaum 1993), they suggest a more robust result, as the upper (t-hat+) and lower (t-hat) bounds do not contain 0, at least until 1.6.

The results of our sensitivity analysis reveal a similar increase in gamma to that reported in the extant literature (1.2- to 1.6-fold; DiPrete et al. 2004; Li et al. 2016; Manchanda et al. 2015; Sun and Zhu 2014). Although the treatment effect will be insignificant if the unobservable is large enough to change propensities, this does not mean that the PS method is invalid because Rosenbaum bounds analysis gives us a lower bound of confidence when making a causal inference in the worst-case scenario with hypothetical hidden selection bias (note that hidden bias caused by unobservables does not necessarily exist). For this reason, we are confident that our study is, to some extent, robust to hidden bias caused by hypothetical unobserved factors affecting the selection process.

Table A10. Sensitivity Analysis: Rosenbaum Bounds Test

Gamma	Significance Level		Hodges–Lehmann Point Estimate		95% Confidence Interval	
	sig+	sig-	t-hat+	t-hat	CI+	CI-
1	0.000205	0.000205	13.9456	13.9456	5.50612	24.1935
1.05	0.00052	0.000075	12.9032	15.8307	4.83871	25
1.1	0.00119	0.000027	12.5	17.5824	3.22581	26.4631
1.15	0.002482	9.50E-06	11.2903	18.3333	2.2043	27.7778
1.2	0.004777	3.30E-06	10.2716	19.3548	1.25353	29.6461
1.25	0.008569	1.10E-06	9.05707	20.2419	−4.40E-07	30
1.3	0.014442	3.80E-07	8.06452	21.4286	−4.40E-07	30.8405
1.35	0.02304	1.30E-07	6.45162	22.4194	−4.40E-07	32.0079
1.4	0.035005	4.20E-08	6.45161	23.5526	−4.40E-07	33.1349
1.45	0.050917	1.40E-08	5.50612	24.1935	−4.00E-06	33.8095
1.5	0.071236	4.50E-09	5	24.7878	−1.0326	35.0824
1.55	0.096249	1.50E-09	4.07337	25.8064	−2.01613	35.7692
1.6	0.126037	4.70E-10	3.22581	26.7374	−3.10676	36.7374

Note: Gamma log odds of differential assignment because of unobserved factors; sig + (−): upper (lower) bound of the significance level; t-hat + (−): upper (lower) bound of the Hodges–Lehmann point estimate CI + (−): upper (lower) bound of the confidence interval ($\alpha = 0.95$).

3) Addressing the Possibility of an Inflated Long-Term Effect

i. Analyses of a Shorter-Period Sample

This is an alternative approach to address the concern about the possible long-term inflation of the treatment effect if Airbnb’s search algorithm favors properties with professional photos. To address this concern, we

estimate our main DiD specification on a set of subsamples in which we included a shorter period (8 months) for each treated unit.

For each treated unit, the observation spans exactly four periods per property before and after treatment adoption. For the untreated (control) units, we used the full periods (16 months) for estimation. This is because for untreated units, there is no reference period for which we can define the pre- and post-spans.

Table A11 presents the estimation results for the robustness analysis of the selected subsets. As shown, the estimated coefficients of the key variable, *TREATIND*, remain consistently positive and significant. The consistent estimated treatment effect when shorter periods are applied adds confidence that our main finding was not driven by long-term inflation caused by Airbnb's search ranking algorithm.

Table A11. DiD Robustness Tests: Selecting Shorter Periods of Samples (Limited Eight Periods/Month Spans)

Variables	Subset of four pre- and four post-treatment periods for treated units and a whole period for untreated units	
	Estimates	Robust S.E.
<i>TREATIND</i>	10.33***	1.624
<i>log REVIEW_COUNT_{t-1}</i>	10.03***	0.768
<i>NIGHTLY_RATE</i>	−0.175***	0.0262
<i>INSTANT_BOOK</i>	4.436***	1.169
<i>CLEANING_FEE</i>	0.0898***	0.0125
<i>MAX_GUESTS</i>	−2.142*	1.034
<i>RESPONSE_RATE</i>	0.0767*	0.0348
<i>RESPONSE_TIME (minutes)</i>	−0.00152	0.00136
<i>MINIMUM_STAY</i>	0.0279	0.0873
<i>SECURITY_DEPOSIT</i>	0.00298*	0.00143
<i>SUPER_HOST</i>	1.989*	1.003
<i>BUSINESS_READY</i>	0.902	0.919
<i>CANCELLATION_STRICT</i>	0.521	1.045
<i>HAS_RATING</i>	6.380	8.310
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	0.204	1.336
<i>HAS_RATING</i> × <i>ACCURACY</i>	0.00741	1.056
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	−2.015	1.035
<i>HAS_RATING</i> × <i>CHECKIN</i>	−1.142	1.398
<i>HAS_RATING</i> × <i>LOCATION</i>	0.805	1.074
<i>HAS_RATING</i> × <i>VALUE</i>	1.552	0.998
Fixed Effect	Property	
Seasonality	City-Year-Month	
Observations	75,406	
R-squared	0.6943	

Note: The number of observations is smaller than that in the main analyses (Tables 2, 3, 4, and 7 in the main paper) because a shorter period was used for the treated units. Specifically, instead of using the whole panel of 16 periods of data, we used up to eight periods for the estimation in this table.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

ii. **Potential Dynamics in the Treatment Effect (Robustness Test of Possible Long-term Inflation)**

To address the concern that the estimated effect could be inflated in the long term if Airbnb's ranking algorithm favors properties with professional images, we implemented a relative-time model to break the *after* period into a series of short periods, and we estimated the treatment effect for each month following treatment adoption. We again relied on a relative model to estimate the following:

$$DEMAND_{it} = INTERCEPT + \sum_j \beta_j (PRE_{it}(j) \cdot TREAT_i) + \sum_k \beta_k (POST_{it}(k) \cdot TREAT_i) + \lambda CONTROLS_{it} + SEASONALITY_{cym} + PROPERTY_i + \varepsilon_{it}.$$

The inclusion of the period leads, $POST(k)$, allowed us to examine the possible dynamics in the treatment effect across periods after adopting the treatment. That is, if the treatment had a greater impact a certain number of months after the treatment adoption, we expected that the estimated coefficient associated with that period dummy would be greater.

As shown in Table A12, the treatment effect exhibits dynamics in the post-treatment period. Specifically, the effect size of the treatment increases over time before stabilizing (decreasing). We interpret this as follows: in the month in which the treatment was adopted, $(POST(0) * TREAT)$, we may have underestimated the treatment effect, as some bookings may have been made in previous periods. In later periods, the estimated coefficient of the verified photos initially increases before stabilizing. There are two forces that may lead to this change:

- (1) All the demand in these periods is due to the verified photos and not the unverified ones.
- (2) With an increased demand, Airbnb ranking algorithms may have started showing this property higher in search rankings, leading to more demand.

Comparing the effect size in the month following the treatment month (i.e., the coefficient of $POST(1) * TREAT$) to the treatment effect obtained from the main model, we see that the results are consistent (8.958 versus 9.303). The coefficient of interaction term with the last period dummy, $POST(5) * TREAT$, decreased compared with the estimated coefficient for $\{POST(k) * TREAT\}_{k=1}^4$. This is likely because we have fewer properties for which more than four post-treatment periods are observed (for example, a property that adopted treatment in February 2017 contributes to the estimation of coefficients for $\{POST(k) * TREAT\}_{k=1}^4$, with $k = 0-2$ only).

Note that to conclusively rule out or separate any potential effect of Airbnb's ranking system that could be confounded with verified photo adoption, we would like to know whether Airbnb's ranking system considers property images and, if so, how the inclusion of images affects the ranking algorithm. Unfortunately, we do not have information about whether and how (if at all) Airbnb increases the ranking of treated properties, as the algorithm is proprietary. In addition, if the data allow, we would like to use more granular information for investigating whether Airbnb ranks properties with verified photos higher immediately after the adoption of verified photos. For example, using the available data, one could examine whether a property received more

bookings within the next couple of hours or days. Yet, such an analysis is not feasible because our demand and treatment data are at a monthly level. We knew the month in which a property adopted the treatment condition but did not have more granular information about the treatment time (e.g., time stamp, exact date, or exact week). In addition, our demand data were at a monthly level (e.g., number of days booked in a month). As a result, we could not test whether demand increased within a very short period after the properties adopted verified photos. We acknowledge that more granular data are needed to resolve this issue. The results should therefore be interpreted with this caveat in mind.

Table A12. Falsification Checks of Pre-Treatment Trends: Relative-Time Model

Variables	DiD Model (Relative Time)	
	Estimates	Robust S.E.
<i>PRE (-4) *TREAT</i>	2.674	2.494
<i>PRE (-3) *TREAT</i>	-0.687	3.163
<i>PRE (-2) *TREAT</i>	1.706	3.102
<i>PRE (-1) *TREAT</i> (reference month)	--	--
<i>POST (0) *TREAT</i>	6.858*	3.130
<i>POST (1) *TREAT</i>	9.303**	2.917
<i>POST (2) *TREAT</i>	12.78***	2.932
<i>POST (3) *TREAT</i>	12.92***	2.918
<i>POST (4) *TREAT</i>	13.00***	2.927
<i>POST (5) *TREAT</i>	8.818***	2.609
Property (Non-Photo) Characteristics		
<i>log REVIEW_COUNT_{t-1}</i>	9.367***	0.931
<i>NIGHTLY_RATE</i>	-0.145***	0.0319
<i>INSTANT_BOOK</i>	4.059**	1.364
<i>CLEANING_FEE</i>	0.0808***	0.0183
<i>MAX_GUESTS</i>	0.0581	1.098
<i>RESPONSE_RATE</i>	0.0705	0.0433
<i>RESPONSE_TIME (minutes)</i>	-0.000424	0.00162
<i>MINIMUM_STAY</i>	0.132	0.130
<i>SECURITY_DEPOSIT</i>	0.00162	0.00198
<i>SUPER_HOST</i>	3.698*	1.489
<i>BUSINESS_READY</i>	1.809	0.982
<i>CANCELLATION_STRICT</i>	1.068	1.270
<i>HAS_RATING</i>	13.86	12.19

$HAS_RATING \times COMMUNICATION$	-0.183	1.425
$HAS_RATING \times ACCURACY$	1.028	1.208
$HAS_RATING \times CLEANLINESS$	-1.509	1.136
$HAS_RATING \times CHECKIN$	-2.039	1.523
$HAS_RATING \times LOCATION$	-0.686	1.176
$HAS_RATING \times VALUE$	2.057	1.180
<hr/>		
Fixed Effect	Property	
Seasonality	City-Year-Month	
<hr/>		
Num. Observations	76,901	
R-squared	0.6616	
<hr/>		
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.		

4) Adding Interaction Terms with Meaningful Amenities

One concern regarding the self-selection issue is that properties with particular amenities may be more likely to adopt the treatment. For example, if some amenities make the properties more attractive in particular seasons (e.g., a pool or beach in summer), and the hosts adopt the treatment at that time, then some of the increase in demand could be brought about by those attractive amenities. The effects of these amenities on demand (which are fixed characteristics of the property) cannot be fully accounted for by the property's fixed effect terms, as amenities' effects may be time varying (e.g., a pool has a greater effect in summer than in winter).

To address this concern, when estimating the PSs for PSW, we obtained additional data on the properties' amenities and incorporated amenity information (e.g., AC, pool, whether the property is close to a beach) to account for the possible factors that may be correlated with both property demand and the hosts' treatment adoption decision in particular seasons but that cannot be captured by property fixed effects. In addition, in the model specification, we added interaction terms for the dummy *AFTER* and meaningful amenities (e.g., pool, beach, AC) to account for the higher effect of these time-invariant variables after the treatment or in particular seasons.

In Table A13, we present the estimation results. Column (1) presents the results of Equation (2a), which is our main specification, and column (2) presents the results of Equation (2b), in which we incorporated the interaction of *AFTER* with meaningful amenities. The consistency of the estimated results confirms a positive and significant treatment coefficient of more than an 8% increase in the property occupancy rate after controlling for area-specific seasonality, as well as the time-varying effect of particular time-invariant property amenities. In the panel "Interacting with Meaningful Property Amenities," we present the coefficients of the interaction terms. As can be seen, all coefficients are insignificant.

$$\begin{aligned}
\text{DEMAND}_{it} &= \text{INTERCEPT} + \alpha_3 \text{TREATIND}_{it} + \lambda \text{CONTROLS}_{it} + \text{SEASONALITY}_{cym} \\
&\quad + \text{PROPERTY}_i + \varepsilon_{it}. \quad (2a) \\
\text{DEMAND}_{it} &= \text{INTERCEPT} + \alpha_3 \text{TREATIND}_{it} + \lambda \text{CONTROLS}_{it} + \vartheta_1 \text{AFTER} \times \text{POOL} \\
&\quad + \vartheta_2 \text{AFTER} \times \text{BEACH} + \vartheta_3 \text{AFTER} \times \text{AC} + \text{SEASONALITY}_{cym} \\
&\quad + \text{PROPERTY}_i + \varepsilon_{it}. \quad (2b)
\end{aligned}$$

Table A13. DiD Model: Regressing Property Demand on Verified Photos

Variables	Main DiD Model (Equation 2a)		Interacting with Amenities (Equation 2b)	
	Estimates	Robust S.E.	Estimates	Robust S.E.
<i>TREATIND</i> (<i>TREAT</i> × <i>AFTER</i>)	8.985***	1.660	8.668***	1.747
Property (Non-Photo) Characteristics				
<i>log REVIEW_COUNT_{t-1}</i>	9.375***	0.930	9.403***	0.933
<i>NIGHTLY_RATE</i>	−0.146***	0.0320	−0.149***	0.0320
<i>INSTANT_BOOK</i>	4.156**	1.361	4.168**	1.362
<i>CLEANING_FEE</i>	0.0808***	0.0184	0.0814***	0.0185
<i>MAX_GUESTS</i>	0.260	1.117	0.305	1.111
<i>RESPONSE_RATE</i>	0.0699	0.0430	0.0693	0.0430
<i>RESPONSE_TIME (minutes)</i>	−0.000477	0.00161	−0.000591	0.00161
<i>MINIMUM_STAY</i>	0.133	0.131	0.136	0.131
<i>SECURITY_DEPOSIT</i>	0.00177	0.00201	0.00171	0.00199
<i>SUPER_HOST</i>	3.801*	1.494	3.781*	1.494
<i>BUSINESS_READY</i>	1.806	0.985	1.791	0.984
<i>CANCELLATION_STRICT</i>	1.016	1.271	1.040	1.271
<i>HAS_RATING</i>	14.32	12.25	14.56	12.24
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	−0.212	1.420	−0.145	1.421
<i>HAS_RATING</i> × <i>ACCURACY</i>	0.878	1.211	0.837	1.211
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	−1.344	1.133	−1.323	1.134
<i>HAS_RATING</i> × <i>CHECKIN</i>	−2.060	1.526	−2.118	1.526
<i>HAS_RATING</i> × <i>LOCATION</i>	−0.757	1.183	−0.767	1.183
<i>HAS_RATING</i> × <i>VALUE</i>	2.141	1.176	2.142	1.173
Interacting with Meaningful Property Amenities				
<i>AFTER</i> × <i>POOL</i>			6.157	4.692
<i>AFTER</i> × <i>BEACH</i>			−10.77	10.83

<i>AFTER</i> × <i>AC</i>		1.034	3.990
Fixed Effect	Property	Property	
Seasonality	City-Year-Month	City-Year-Month	
No. of Observations	76,901	76,901	
R-squared	0.6608	0.6609	
Note: * <i>p</i> < 0.05, ** <i>p</i> < 0.01, *** <i>p</i> < 0.001.			

5) Testing Changes in the Property or the Host's Unobserved Quality (via Multidimensional Ratings)

It is possible that properties or hosts self-select the adoption of professional Airbnb images when there are substantial changes in the quality of the experience delivered to guests. Such changes could include an upgrade of the house/room or the delivery of more friendly/better services to guests. This would introduce an upward bias in the estimated coefficient, as it could happen at the same time as the treatment adoption. Although we were not able to observe and control for everything, we adopted a few measures to help control for or alleviate this issue.

First, in the demand model, we added the measurement of host responsiveness (host response rate and host response time) to address the concern that hosts are more responsible in the post-treatment period.

Second, in the demand model, we added a complete set of multidimensional guest ratings to capture and account for any potential changes in the stay experience or hosting quality.

Third, as we show below, for the properties for which ratings are available, we compared the average ratings in terms of multidimensional aspects for a few periods before and after the treatment. The goal is to examine whether there are substantial changes in the property characteristics that were unobserved by us but captured in guests' ratings. To implement this robustness test, we estimated the following:

$$Rating_{it} = INTERCEPT + \alpha_3 TREATIND_{it} + \lambda CONTROLS_{it} + SEASONALITY_{cym} + PROPERTY_i + \varepsilon_{it},$$

where the specification is the same as that in our main demand model, and the dependent variable is replaced with one of the following guest ratings capturing, to some extent, how good a stay or host was: communication, accuracy, cleanliness, and check-in. The metrics for cleanliness capture potential changes in the property, whereas those for communication capture the quality of communication between a host and a guest. The coefficient α_3 captures changes in ratings along a particular dimension after treatment adoption.

As shown in Table A14, the coefficients of *TREATIND* are insignificant across all specifications, suggesting that for the treated units, there was no substantial change in the stay or hosting quality delivered to guests before and after adopting verified photos.

Table A14. Robustness Test: Changes in Multidimensional Guest Ratings after Verified Photo Adoption

Variables	DV: Multidimensional Guest Review Rating			
	Communication	Accuracy	Cleanliness	Check-in
<i>TREATIND</i>	−0.0155 (0.0118)	0.0175 (0.0167)	−0.0261 (0.0193)	−0.0245 (0.0177)
<i>log REVIEW_COUNT_{t-1}</i>	−0.0371 (0.0189)	0.0277 (0.0265)	−0.0582* (0.0286)	0.0261 (0.0285)
<i>NIGHTLY_RATE</i>	0.00189*** (0.000297)	−0.000621 (0.000344)	0.00123* (0.000546)	0.000400 (0.000394)
<i>INSTANT_BOOK</i>	−0.0134 (0.0153)	0.0154 (0.0187)	−0.0174 (0.0197)	0.0162 (0.0158)
<i>CLEANING_FEE</i>	−0.000325 (0.000208)	0.000604 (0.000378)	−0.000343 (0.000665)	−0.000104 (0.000318)
<i>MAX_GUESTS</i>	−0.0101 (0.00728)	0.00335 (0.00814)	0.0449*** (0.0125)	0.0106 (0.00988)
<i>RESPONSE_RATE</i>	−0.000951** (0.000332)	0.000425 (0.000367)	−0.000782 (0.000504)	0.000130 (0.000357)
<i>RESPONSE_TIME (minutes)</i>	−0.00000573 (0.0000131)	0.000000617 (0.0000126)	−0.00000225 (0.0000163)	−0.0000180 (0.0000133)
<i>MINIMUM_STAY</i>	−0.00129** (0.000395)	−0.000554 (0.000612)	−0.000550 (0.000739)	0.000187 (0.000565)
<i>SECURITY_DEPOSIT</i>	0.00000253 (0.0000107)	0.0000174 (0.0000112)	0.0000104 (0.0000168)	−0.00000397 (0.0000101)
<i>SUPER_HOST</i>	−0.0246* (0.0117)	0.0169* (0.00826)	0.0188 (0.0154)	0.00510 (0.00905)
<i>BUSINESS_READY</i>	−0.00618 (0.00943)	0.0166 (0.0109)	0.0270 (0.0140)	0.00700 (0.00982)
<i>CANCELLATION_STRICT</i>	−0.0281 (0.0147)	0.0284 (0.0164)	−0.0152 (0.0200)	0.0236 (0.0194)
Fixed Effect	Property	Property	Property	Property
Seasonality	City-Year-Month	City-Year-Month	City-Year-Month	City-Year-Month
Observations	45,386	45,386	45,386	45,386

R-squared	0.8771	0.8931	0.9155	0.8562
-----------	--------	--------	--------	--------

Note: The number of observations (45,386) is lower than that in our main analyses (76,901). This is because the set of analyses presented here used a subset of samples that included only properties with guest ratings presented on the property page. Note that Airbnb computes and presents ratings only when a property has more than three guest reviews. As a result, when a property had less than three reviews in a given period, that observation was automatically dropped from the series of analyses presented in this table.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

6) Alternative Specifications: Logging All Numeric Variables

We performed a log transformation of all numeric variables (security deposit, price, etc.) and reported the results of the estimation of the main DiD regression (Table 2 in the main paper). Table 3 presents the estimation results. As can be seen, the estimated coefficient of the verified photos is consistent with the treatment effect reported in the main paper.

Table A15. Alternative DiD Model Specification: Regressing Property Demand on Verified Photos

Variables	DiD Model	
	Estimates	Robust S.E.
<i>TREATIND</i>	9.135***	(1.647)
<i>log REVIEW_COUNT_{t-1}</i>	9.044***	(0.867)
<i>logNIGHTLY_RATE</i>	-6.182***	(1.145)
<i>INSTANT_BOOK</i>	3.725**	(1.388)
<i>logCLEANING_FEE</i>	1.898***	(0.491)
<i>logMAX_GUESTS</i>	-1.125	(5.412)
<i>logRESPONSE_RATE</i>	5.569***	(1.368)
<i>logRESPONSE_TIME (minutes)</i>	-0.0876	(0.258)
<i>logMINIMUM_STAY</i>	1.480	(1.429)
<i>logSECURITY_DEPOSIT</i>	0.280	(0.318)
<i>SUPER_HOST</i>	3.937**	(1.493)
<i>BUSINESS_READY</i>	1.666	(0.992)
<i>CANCELLATION_STRICT</i>	1.426	(1.298)
<i>HAS_RATING</i>	13.42	(12.26)
<i>HAS_RATING × COMMUNICATION</i>	-0.0991	(1.423)
<i>HAS_RATING × ACCURACY</i>	0.975	(1.224)
<i>HAS_RATING × CLEANLINESS</i>	-1.328	(1.118)
<i>HAS_RATING × CHECKIN</i>	-2.237	(1.520)
<i>HAS_RATING × LOCATION</i>	-0.666	(1.174)

$HAS_RATING \times VALUE$	2.267	(1.174)
Fixed Effect	Property	
Seasonality	City-Year-Month	
Num. Observations	76,901	
R-squared	0.6613	
Note: $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.		

7) Including Pre-treated Units in the Control Group

As a robustness test, we included pre-treated units in the control group (which we excluded from our main analyses for the sake of conceptual clarity, as these pre-treated units had verified photos prior to the start of our observation window). As such, we estimated the change in property demand when the units to be treated (i.e., properties that were untreated prior to our observation window and adopted verified photos during our observation window) used verified photos compared with the pre-treated and untreated units. The logic is that the properties (or hosts) in the pre-treated and to-be-treated groups might be more similar in terms of characteristics in obtaining verified photos. Therefore, including pre-treated units in the comparison group can serve as a falsification check, as the treatment effect should occur only after a property changes from untreated to treated. A similar logic was used in previous work (Manchanda et al. 2015; Narang and Shankar 2019) as a robustness test when self-selection for the treatment was a concern.

Table A16 presents the estimation results. As can be seen, we obtained a consistent estimated treatment effect (the coefficient for *TREATIND* is 7.977 and statistically significant) when we included the pre-treated units in the control group¹⁷.

Table A16. DiD Model: Regressing Property Demand on Verified Photos on Property Demand, Including Pre-treated Properties in the Control Group

Variables	DiD Model	
	Estimates	Robust S.E.
<i>TREATIND</i>	7.977***	(1.756)
<i>log REVIEW_COUNT_{t-1}</i>	8.082***	(0.882)
<i>NIGHTLY_RATE</i>	−0.126***	(0.0354)
<i>INSTANT_BOOK</i>	4.595***	(1.230)
<i>CLEANING_FEE</i>	0.107***	(0.0237)
<i>MAX_GUESTS</i>	1.385	(1.215)

¹⁷ Note that although that the effect size was 11% lower (compared with the value of 8.985 that we obtained from the main DiD regression), the estimated treatment effect was very close, and it remained at the same level. For example, Narang and Shankar (2019) reported that the estimated effects from their robustness test were 16%–35% lower than their main results.

<i>RESPONSE_RATE</i>	0.0735	(0.0448)
<i>RESPONSE_TIME (minute)</i>	−0.000490	(0.00172)
<i>MINIMUM_STY</i>	0.198	(0.171)
<i>SECURITY_DEPOSIT</i>	0.00313	(0.00223)
<i>SUPER_HOST</i>	3.346*	(1.495)
<i>BUSINESS_READY</i>	0.969	(0.963)
<i>CANCELLATION_STRICT</i>	1.362	(1.305)
<i>HAS_RATING</i>	12.37	(16.05)
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	−0.637	(1.482)
<i>HAS_RATING</i> × <i>ACCURACY</i>	0.712	(1.236)
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	−0.437	(1.299)
<i>HAS_RATING</i> × <i>CHECKIN</i>	−1.162	(1.610)
<i>HAS_RATING</i> × <i>LOCATION</i>	−1.165	(1.376)
<i>HAS_RATING</i> × <i>VALUE</i>	1.632	(1.305)
Fixed Effect	Property	
Seasonality	City-Year-Month	
Observations	123186	
R-squared	0.6611	
Note: The number of observations here is larger than that in the main analyses (76,901). This is because the regression presented in this table was performed on a PSW-matched sample that included pre-treated units in the control group.		
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$		

Using Only Pre-treated Units as the Control Group

In Table A8, we present the robustness checks by including the pre-treated units in the control group. In this subsection, we use a stronger control—using the pre-treated units only. Similar to the robustness test in the studies of Manchanda et al. (2015) and Narang and Shankar (2019), this analysis provides a conservative view by comparing the dependent variable among *adopters* only.

Table A17 presents the estimation results. As can be seen, we obtained consistent results (the estimated coefficient of *TREATIND* is 6.77 and significant at the 0.05 significance level) when using only the pre-treated units as the control¹⁸. This robustness analysis, along with Table A8, suggests that our PSW combined with DiD can mitigate the self-selection issue in the verified photo adoptions.

¹⁸ Note that despite the effect size being 24% lower, the estimated effect was very close and remained at the same level (e.g., Narang and Shankar [2019] reported that the estimated effects from this robustness analysis were 16%–35% lower than their main results across multiple DVs of interest).

Table A17 Regressing Property Demand on Verified Photos on Property Demand Using Only Pre-treatment Properties as a Control Group

VARIABLES	DiD Model	
	ESTIMATES	Robust S.E.
<i>TREATIND</i>	6.769***	(1.011)
<i>log REVIEW_COUNT_{t-1}</i>	5.943***	(0.345)
<i>NIGHTLY_RATE</i>	−0.0433***	(0.0102)
<i>INSTANT_BOOK</i>	4.259***	(0.488)
<i>CLEANING_FEE</i>	0.0228*	(0.0107)
<i>MAX_GUESTS</i>	−0.185	(0.297)
<i>RESPONSE_RATE</i>	0.0245	(0.0169)
<i>RESPONSE_TIME (minute)</i>	−0.000291	(0.000650)
<i>MINIMUM_STY</i>	0.00502	(0.0503)
<i>SECURITY_DEPOSIT</i>	−0.0000970	(0.000710)
<i>SUPER_HOST</i>	1.393**	(0.487)
<i>BUSINESS_READY</i>	0.402	(0.354)
<i>CANCELLATION_STRICT</i>	−1.046	(0.844)
<i>HAS_RATING</i>	5.406	(6.542)
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	0.150	(0.617)
<i>HAS_RATING</i> × <i>ACCURACY</i>	−0.118	(0.490)
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	−0.302	(0.424)
<i>HAS_RATING</i> × <i>CHECKIN</i>	−0.299	(0.586)
<i>HAS_RATING</i> × <i>LOCATION</i>	−0.275	(0.430)
<i>HAS_RATING</i> × <i>VALUE</i>	0.709	(0.456)
Fixed Effect	Property	
Seasonality	City-Year-Month	
Observations	51918	
R-squared	0.6814	
Note: The number of observations is different from that in the main analysis (76,901), as this regression was performed on a different sample, in which the 4,932 pre-treated units (i.e., properties that adopted verified photos prior to our observation) formed the control group. The DV is the property monthly occupancy rate.		
* <i>p</i> < 0.05, ** <i>p</i> < 0.01, *** <i>p</i> < 0.001		

8) Additional Analyses

i. Regressing Property Price as the DV

We examined how price changed in the post-treatment period. We replicated our DiD regression but replaced the DV with the property's nightly price to estimate the coefficient of verified photos on price. As shown in Table A18, the insignificant coefficient of *TREATIND* is positive, suggesting an upward change in price. However, the coefficient is insignificant at a significance level of 95%. The results imply that after controlling for the property fixed effects and city-specific seasonality and yearly effects, a property's price does not change the price based on having obtained verified photos.

Note that this result does not mean that a property's price cannot increase or change during the post-treatment period. For example, a property's price may increase when it has accumulated more reviews or the host has gained more experience, or as a response to seasonality effects. Rather, the insignificant coefficient of *TREATIND* in Table A18 should be interpreted as indicating that controlling for all other factors (time-varying property characteristics, city-specific seasonality), a property's price would not increase solely because of the use of verified photos. A possible explanation is that changing photos is not automatically associated with improvements in other aspects of the property and host. Thus, everything else being equal, a host would not increase the price after adopting verified photos. Another explanation is that many Airbnb hosts lack rich information about demand and price, so they rely on Airbnb's smart pricing algorithm to automatically set prices. As a result, it is possible that the price would be higher because of the use of verified photos if the smart pricing algorithm did not include the property pictures as an input. Of course, Airbnb does not disclose the details of its proprietary algorithm. Our understanding is that the algorithm considers many factors, such as seasonality, market popularity, and review history. In Airbnb's report on its algorithm, property pictures were not included or described as one of the many factors at play in the algorithm.¹⁹

Table A18. Regressing Property Price on the Treatment Indicator: DiD Model

Variables	Regressing Property Price as the DV	
	Estimates	Robust S.E.
<i>TREATIND</i>	4.388	(3.047)
<i>log REVIEW_COUNT_{t-1}</i>	-14.29***	(2.245)
<i>INSTANT_BOOK</i>	-6.833**	(2.145)
<i>CLEANING_FEE</i>	0.721***	(0.0514)
<i>MAX_GUESTS</i>	4.898	(2.821)

¹⁹ See <https://blog.airbnb.com/smart-pricing/>.

<i>RESPONSE_RATE</i>	0.241**	(0.0900)
<i>RESPONSE_TIME (minute)</i>	0.000684	(0.00445)
<i>MINIMUM_STY</i>	−0.947**	(0.310)
<i>SECURITY_DEPOSIT</i>	0.0443	(0.0237)
<i>SUPER_HOST</i>	−4.432**	(1.569)
<i>BUSINESS_READY</i>	1.909	(2.344)
<i>CANCELLATION_STRICT</i>	−8.532***	(2.318)
<i>HAS_RATING</i>	−142.2***	(21.38)
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	−3.999	(2.071)
<i>HAS_RATING</i> × <i>ACCURACY</i>	3.298	(2.604)
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	7.547***	(1.775)
<i>HAS_RATING</i> × <i>CHECKIN</i>	3.812	(2.279)
<i>HAS_RATING</i> × <i>LOCATION</i>	11.37***	(1.899)
<i>HAS_RATING</i> × <i>VALUE</i>	−8.340**	(2.677)
Fixed Effect	Property	
Seasonality	City-Year-Month	
Observations	76901	
R-squared	0.8810	

Note: The DiD model was estimated by regressing property price on the same set of variables as the main DiD model, except for instrumented price on the right-hand side, which was excluded. Regression was performed on a dataset including properties that had at least one open day in a month. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

ii. Including Review Probability in the PSW and DiD Model

Zhang et al. (2019) found that image quality affects the probability that a customer (after the stay) will write a review for that property. We follow Zhang et al. (2019) and investigate the aspect of review probability as a key variable affecting property demand. To do so, we match properties based on *review probability* to control for this pre-treatment characteristic. In the DiD model, we also control for review probability (the review probability from the previous month). The review probability for a property in each month is computed as the number of received reviews divided by the reservations in that month.

As can be seen in Table A19, the estimated coefficient of *TREATIND* ($b = 7.271$, $p < 0.001$) is consistent with our main results²⁰. In addition, the estimated coefficient of *REVIEW_PROBABILITY* is insignificant at the 0.05 significance level.

²⁰ Note that this regression is estimated on those observations in which *review probability* can be defined and computed, that is, the observations when a property in a month had at least one booking. By including review probability in the model, we automatically excluded those observations with zero booking (*demand* is 0) in a month, which is more likely for untreated properties than for treated properties. As a result, the coefficient of *TREATIND* in this analysis is a conservative estimate.

Our interpretation is that although in our main analyses, we did not explicitly include review probability in the PSW and DiD analyses, this variable was indirectly controlled for. This is because review probability can be computed from (as a function of) a set of covariates, including image quality, host response time, price, and other property characteristics (Zhang et al. 2019). These covariates were exactly used in the PSW model and the DiD model. Furthermore, when we thought about how review probability may influence property demand, we found that review probability should not have an impact on the present demand. This is because review probability affects the evolution of *Number of Reviews*, which is a key driver in attracting bookings. Following the arguments and findings of Zhang et al. (2019), review probability affects a property's future demand by affecting the *speed* of accumulating reviews. We believe that review probability alone should not directly affect demand (in fact, Zhang et al. (2019) did not include this variable in their demand model). Although review probability is related to property quality and guest satisfaction, it is unobserved by guests. We can compute review probability because we obtain the # reviews and # bookings in each month. For an average Airbnb guest, however, such information is unknown (a guest observes the total number of reviews but does not know how many bookings a property has received). As a result, guests cannot use review probability as an important variable to aid their booking decisions. It then follows that review probability alone should not have an impact on the present demand for a property.

Table A19 Including Review Writing Probability in Demand Regression

VARIABLES	DiD Model	
	ESTIMATES	Robust S.E.
<i>TREATIND</i>	7.271***	(2.022)
<i>log REVIEW_COUNT_{t-1}</i>	7.362***	(1.249)
<i>NIGHTLY_RATE</i>	-0.149***	(0.0374)
<i>INSTANT_BOOK</i>	5.775***	(1.501)
<i>CLEANING_FEE</i>	0.0779**	(0.0279)
<i>MAX_GUESTS</i>	-0.312	(1.263)
<i>RESPONSE_RATE</i>	0.120	(0.0670)
<i>RESPONSE_TIME (minute)</i>	-0.00152	(0.00245)
<i>MINIMUM_STY</i>	0.227	(0.158)
<i>SECURITY_DEPOSIT</i>	0.00486	(0.00261)
<i>SUPER_HOST</i>	2.094	(1.576)
<i>BUSINESS_READY</i>	0.503	(1.171)
<i>CANCELLATION_STRICT</i>	0.666	(1.557)

<i>HAS_RATING</i>	21.64	(20.22)
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	−1.632	(1.842)
<i>HAS_RATING</i> × <i>ACCURACY</i>	−0.430	(1.588)
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	−0.865	(1.336)
<i>HAS_RATING</i> × <i>CHECKIN</i>	0.885	(1.807)
<i>HAS_RATING</i> × <i>LOCATION</i>	−1.783	(1.620)
<i>HAS_RATING</i> × <i>VALUE</i>	1.932	(1.292)
<i>REVIEW_PROBABILITY</i> _{<i>t-1</i>}	2.06	(1.507)
Fixed Effect	Property	
Seasonality	City-Year-Month	
Observations	40590	
R-squared	0.5797	
Note: The DV is the property monthly demand. The number of observations in this regression is different from that in the main analysis (76,901). This is because this regression includes review probability in the previous month as a control variable. Therefore, the model was regressed on observations in which a property had at least one booking in the previous month (other: in addition to this, the property had at least one open day in the current month).		
* <i>p</i> < 0.05, ** <i>p</i> < 0.01, *** <i>p</i> < 0.001		

9) Testing Host Behavior – Changes in the Number of Open Days

One self-selection concern is that hosts may change their behavior (in a way that will increase the property booking) along with the treatment adoption. To test this, we used the number of open days to serve as a falsification test for endogeneity. Note that the treatment and control properties differ in the observed variables (the treatment is endogenous in our setting). To address this, we performed a propensity score-based adjustment to make the treatment and control groups comparable on the observed measures. If the propensity score-weighted treatment and the control properties were to exhibit a different number of nights open subsequent to the treatment, then it would indicate that part of the treatment may be driven by hosts' desire to open for more nights.

To check this, we produced the following results. Here, we regressed open nights as a function of the treatment indicator and other controls (the dependent variable was log (#open nights + 1)). We found that the treatment had no impact on the number of open nights. Hosts' adoption of treatment did not appear to be driven by the intent to open the property for more days in the future. As shown in Table A20, the insignificant coefficients across the two model specifications suggest that the increase in demand was not introduced by hosts being more likely to open nights after adopting Airbnb professional photos. Note that in the model specification in column (2), we added the number of reservation days in the previous period as an additional control.

As another robustness test, we include # Open Days in the DiD demand model, control for possible changes in the number of open days. As presented in Table A21, the estimated coefficients are similar to our main DiD estimation results. Specifically, the estimated coefficient of *TREATIND* on demand was 8.984 ($p < 0.001$).

Table A20 Robustness Test: Changes in # of Open Days along with Verified Photo Adoption

VARIABLES	DV: log # Open Days in Period t	
	Model (1)	Model (2)
<i>TREATIND</i>	0.0201 (0.0339)	-0.0267 (0.0317)
$\log \# \text{ RESERVATION_DAYS}_{t-1}$		0.194*** (0.00831)
$\log \text{ REVIEW_COUNT}_{t-1}$	-0.00463 (0.0181)	-0.0650*** (0.0176)
<i>NIGHTLY_RATE</i>	0.00140* (0.000554)	0.00184*** (0.000517)
<i>INSTANT_BOOK</i>	0.0431 (0.0226)	0.01000 (0.0213)
<i>CLEANING_FEE</i>	-0.000765* (0.000359)	-0.00103** (0.000346)
<i>MAX_GUESTS</i>	0.0112 (0.0232)	0.00981 (0.0204)
<i>RESPONSE_RATE</i>	0.000359 (0.00106)	0.000376 (0.000983)
<i>RESPONSE_TIME (minutes)</i>	-0.0000379 (0.0000331)	-0.0000210 (0.0000307)
<i>MINIMUM_STAY</i>	-0.00824** (0.00296)	-0.00884** (0.00269)
<i>SECURITY_DEPOSIT</i>	-0.0000389 (0.0000398)	-0.0000309 (0.0000369)
<i>SUPER_HOST</i>	-0.0588 (0.0342)	-0.0759* (0.0297)

<i>BUSINESS_READY</i>	0.00305 (0.0199)	−0.00470 (0.0177)
<i>CANCELLATION_STRICT</i>	−0.0156 (0.0270)	−0.0277 (0.0254)
<i>HAS_RATING</i>	−0.157 (0.239)	−0.255 (0.219)
<i>HAS_RATING</i> × <i>COMMUNICATION</i>	−0.114*** (0.0304)	−0.100*** (0.0276)
<i>HAS_RATING</i> × <i>ACCURACY</i>	0.0581* (0.0267)	0.0439 (0.0257)
<i>HAS_RATING</i> × <i>CLEANLINESS</i>	−0.000146 (0.0224)	0.0128 (0.0207)
<i>HAS_RATING</i> × <i>CHECKIN</i>	0.0426 (0.0317)	0.0531 (0.0280)
<i>HAS_RATING</i> × <i>LOCATION</i>	0.0242 (0.0243)	0.0230 (0.0221)
<i>HAS_RATING</i> × <i>VALUE</i>	−0.00595 (0.0233)	−0.0193 (0.0214)
<i>INTERCEPT</i>	2.913*** (0.144)	2.729*** (0.134)
Fixed Effect	Property	Property
Seasonality	City-Year-Month	City-Year-Month
Observations	76,901	76,901
R-squared	0.4444	0.5092
Note: The difference between the two specifications is that in Model (2), we control for the number of reservation days in the previous period; robust standard errors are in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.		

Table A21 Robustness Test: Including # Open Days in Demand Model

VARIABLES	Main DiD Model (Eq. 2)	
	ESTIMATES	Robust S.E.
<i>TREATIND</i>	8.984***	(1.659)
<i>log REVIEW_COUNT_{t-1}</i>	9.377***	(0.931)

NIGHTLY_RATE	-0.147***	(0.0320)
INSTANT_BOOK	4.150**	(1.360)
CLEANING_FEE	0.0809***	(0.0185)
MAX_GUESTS	0.260	(1.117)
RESPONSE_RATE	0.0699	(0.0430)
RESPONSE_TIME (<i>minutes</i>)	-0.000472	(0.00161)
MINIMUM_STAY	0.133	(0.131)
SECURITY_DEPOSIT	0.00178	(0.00201)
SUPER_HOST	3.808*	(1.492)
BUSINESS_READY	1.806	(0.985)
CANCELLATION_STRICT	1.016	(1.271)
HAS_RATING	14.32	(12.24)
HAS_RATING \times COMMUNICATION	-0.203	(1.419)
HAS_RATING \times ACCURACY	0.873	(1.208)
HAS_RATING \times CLEANLINESS	-1.344	(1.133)
HAS_RATING \times CHECKIN	-2.063	(1.526)
HAS_RATING \times LOCATION	-0.761	(1.182)
HAS_RATING \times VALUE	2.143	(1.176)
# OPEN DAYS	0.00610	(0.0583)
Fixed Effect	Property	
Seasonality	City-Year-Month	
Num. Observations	76,901	
R-squared	0.6608	
Note: the D.V. is computed as a function of <i>Open Days</i> (occupancy rate = #reservation days/#open days).		
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.		

VI. Data Description and Sample Construction

We started with over 13,000 unique Airbnb properties in seven US cities, collecting images from the property pages in each month from January 2016 to April 2017. Of the full sample, 4,932 properties were pre-treated (i.e., had verified photos prior to our observation in January 2016). For the analyses presented in this study, we removed these properties from our sample for the sake of conceptual clarity (in an additional robustness test, we included these units in the control group and obtained consistent results; see Section V.7) in this Appendix for detailed results and discussions). The remaining properties totaled 8,858 (13,790 – 4,932). As described in the main paper, we removed missing data points and properties for which there were web page errors during data collection. Errors may have occurred because Airbnb blocked our data collection requests, and/or hosts unlisted their properties from Airbnb temporarily or permanently.

Next, using the valid units, we performed a propensity score analysis to match the treated and untreated units by comparison (matching) based on a set of observed covariates. Of the 7,825 valid properties identified during the observational window, there were 231 treated units and 7,594 untreated units before PSW. Next, we performed a propensity score estimation to find properties in the two groups that were close or matched with each other in terms of the property and host characteristics in the pre-treatment period. We first estimated the propensity score as a logit function of a set of property, host, and neighborhood covariates measured at the start

of the first period (i.e., the pre-treatment period). The set of variables used in the propensity score estimation included time-invariant characteristics (e.g., number of bedrooms, property type, property amenities) and time-varying dynamic characteristics (e.g., property price, number of reviews).

We first denoted the set of variables for the units to be matched on as X . Following prior literature that applies the propensity score approach, we computed the propensity score as a logit function of X . Unit i has an estimated propensity score of $ps_i = f(X_i)$. Then, as we described in the paper, we computed a weight for each unit based on its estimated propensity score. Specifically, we calculated weight as $w_i = \frac{1}{ps_i}$ if i was a treated unit and as $w_i = \frac{1}{1-ps_i}$ if i was an untreated unit. As we described in the paper, this is a widely used PSW approach called inverse probability treatment weighting. Later, we used the sample weights for DiD regressions. Furthermore, to address the concern that a very low estimated probability can result in extremely large (and possibly unstable) weights, we followed common practice in PSW and used trimmed weights, excluding those that were outside 1% and 99% of the distribution (Austin and Stuart 2015; Cole and Hernan 2008). The PSW generated a sample that included 7,211 of the 7,594 untreated units and 212 of the 231 treated units. This matched sample was used in the main regressions (Tables 2, 3, 4, and 7 in the main paper). The total number of observations used in the analyses was 76,901²¹.

²¹ Note that the total observations in the regressions was fewer than # units * # periods because for properties that did not have open days in a month, the dependent variable—property demand (occupancy rate)—would be indefinable.

VII. Example of Property Images with 1 SD of Improvement in Key Image Features

As an example, we show what an increase of 1 SD may look like visually in Figure A1. The figure shows the original photos and a 1 SD increase for two example attributes—brightness and saturation.

Figure A1 Examples of Images with a 1 SD Increase in Image Attributes

Attribute	Original Photo	+1 <i>SD</i> in Attribute
Brightness		
Saturation	