One Picture Is Worth a Thousand Words? The Pricing Power of Images in e-Commerce

Author(s):
Naumzik, Christof; Feuerriegel, Stefan

Publication date:
2020-04

Permanent link:
https://doi.org/10.3929/ethz-b-000422268

Originally published in:
https://doi.org/10.1145/3366423.3380086

One Picture Is Worth a Thousand Words? The Pricing Power of Images in e-Commerce

Christof Naumzik
ETH Zurich
Zurich, Switzerland
cnaumzik@ethz.ch

Stefan Feuerriegel
ETH Zurich
Zurich, Switzerland
sfeuerriegel@ethz.ch

ABSTRACT

In e-commerce, product presentations, and particularly images, are known to provide important information for user decision-making, and yet the relationship between images and prices has not been studied. To close this research gap, we suggest a tailored web mining framework, since one must quantify the relative contribution of image content in describing prices *ceteris paribus*. That is, one must account for the fact that such images inherently depict heterogeneous products. In order to isolate the pricing power of image content, we suggest a three-stage framework involving deep learning and statistical inference.

Our empirical evaluation draws upon a comprehensive dataset of more than 20,000 online real estate listings. We find that the image content describes a large portion of the variance in prices, even when controlling for location and common characteristics of apartments. A one standard deviation in the image variable is associated with a 14.45 % increase in price. By utilizing a carefully designed instrumental variables estimation, we further set out to obtain causal estimates. Our empirical findings contribute to theory by quantifying the hedonic value of images and thus establishing a causal link between visual appearance and product pricing. Even though a positive relationship seems intuitive, we provide for the first time an empirical confirmation. Based on our large-scale computational study, we further yield evidence of a picture superiority effect: simply put, a beneficial image corresponds to the same price change as 2856.03 additional words in the textual description.

In sum, images capture valuable information for users that goes beyond narrative explanations. As a direct implication, we aid online platforms and their users in assessing and improving the multimodal presentation of product offerings. Finally, we contribute to web mining by highlighting the importance of visual information.

KEYWORDS

Product presentation; E-commerce; Images; Data mining; Econometrics

ACM Reference Format:
Christof Naumzik and Stefan Feuerriegel. 2020. One Picture Is Worth a Thousand Words? The Pricing Power of Images in e-Commerce. In *Proceedings of The Web Conference 2020 (WWW '20), April 20–24, 2020, Taipei, Taiwan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3366423.3380086

WWW '20, April 20–24, 2020, Taipei, Taiwan

ACM ISBN 978-1-4503-7023-3/20/04.
https://doi.org/10.1145/3366423.3380086

1 INTRODUCTION

Customer decision-making in online settings is made challenging by the fact that a physical examination of products is infeasible. Instead, users can only evaluate product characteristics based on either textual descriptions and product images in order to make an informed purchase decision. Both textual descriptions and images are known to receive different levels of attention from users: users look first at images when entering a website [10] and images also receive more attention in comparison to other website elements [14]. These findings highlight the importance of product images to customer decision-making, and yet the actual relationship between images and online transactions remains subject to research.

The few works on the use of product images in online settings can be grounded as follows. First, there are studies of a descriptive nature, which explore user attention via eye-tracking [e. g., 10, 14]. Second, the presence of online product images is known to affect purchasing. This has been demonstrated by studying both conversion rate [8] and sales volume [37]. Third, product images also link to prices. This was confirmed by assessing buyers reaction to presentations with images in comparison to those in which images were absent [8]. The above studies only quantify the presence of images; however ignore the image content. Specifically, the relationship between visual appearance, i. e., the so-called *image sentiment*, of a product and its price has remained unknown and thus represents the focus of this paper.

We follow related research analyzing image sentiment in a different setting [e. g., 27, 28]. In our case, we expect product images with a more positive image sentiment to have more appeal, thus attracting greater customer interest. In other words, more positive aesthetics would thus provide the basis for quoting a higher price.

Research Question: *What is the pricing power of image content (i. e., image sentiment) in online product presentations?*

There is strong theoretical backing why online settings (as compared to offline settings) are characterized by a dominant role of product images. The infeasible physical examination in online settings is compensated by other information that are subsumed under the so-called theoretical construct of product diagnosticity [9]. By definition, product diagnosticity refers to all product information – both visual and textual – that eventually convey product attributes, i. e., features, functionality, quality, and design. According to prior theory, an increased level of product diagnosticity affects users in multiple ways. For instance, it reduces buyer uncertainty [9], and increases purchase intentions [22]. Even though product diagnosticity should theoretically embrace visual information, actual findings concerning the role of image sentiment are still lacking.

Hypotheses: This research sets out to study the informativeness of visual content in online product presentations, specifically with respect to pricing. (1) We hypothesize that the appearance of images, i. e., the image sentiment, helps in explaining the variance of prices beyond other product characteristics. In this sense, a higher price should be described by a more positive image sentiment. However, unique to our study is that we make such claims *ceteris paribus*, i. e., after controlling for the potential heterogeneity in product characteristics. (2) We further expect images to capture information beyond narrative description and, hence, compare the informativeness of image sentiment against the length and sentiment of the textual description. (3) We test to whether image sentiment has predictive capacity over prices.

We propose a three-staged causal framework on the basis of deep learning with clear advantages: first of all, it is independent of human raters and their subjectivity. Hence, we refrain from prescribing a specific dimension of what could theoretically characterize pricing power (e. g., for some, aesthetics could be a tidy flat, for others a bright light or a modern design); instead, our classifier *learns* such aesthetics from data. Second, it allows us to conduct a large-scale causal analysis.

2 BACKGROUND

2.1 Images in Online Product Presentations

In contrast to textual descriptions [e. g., 2, 31], research on the role of images in online product presentations is surprisingly sparse.

Prior works have often studied the picture effect on consumer perceptions, yet without analyzing the visual content or confirming its value to hedonic price models. For instance, the presence of pictures have been found to affect customer attitudes towards brands [21]. Further, the imagery-evoking nature of pictures triggers purchase intentions [20]. However, such findings mostly stem from offline settings and, hence, it is unclear to what extent they generalize to online settings.

There have been various works that examine the role of images in online settings. Goswami et al. [11] studied the number of images in eBay listing and further involves human ratings of picture quality. In their work, a correlation analysis was performed but without the rigor of controlling for potential confounding factors. Di et al. [8] extended the analysis to "watching", a particular characteristic of eBay. Various other works examine differences between the presence/absence of images from an economic point of view, as it presents a vehicle for mitigating information asymmetries [e. g., 7, 19]. The work by Zhang et al. [37] investigated changes due to "verified" images via a difference-in-difference estimation at Airbnb, finding that it results in higher demand.

2.2 Picture Superiority Effect

Research on information processing suggests that the format in which information is presented governs the corresponding processing: verbal stimuli evoke mainly discursive processing, while visual stimuli elicit imagery information processing [20]. As a consequence, both are of different importance, which was previously acknowledged in the picture superiority effect.

Actual findings of a picture superiority effect are inconclusive. On the one hand, images have been found to be superior to text with regard to recall and affecting consumers' attitudes [20]. On the other hand, Kim and Lennon [15] performed a study where both visual and verbal stimuli in online product presentations were available; however, only verbal description revealed a significant influence on purchase behavior. Following the notion of a picture superiority effect, we extend it to pricing and thus compare the relative influence from images versus verbal descriptions.

2.3 Image Sentiment

Prior research on image sentiment can be summarized as follows:

Labels are subject to considerable differences and include, for instance, subjective aesthetics [16], objective aesthetics [27], preference towards faces [26], adjective-noun phrases [36], and picture quality [8, 11] amongst others. The variety stems from the fact that there is no universal concept of image sentiment and, as a result, some studies asked subjects to provide ratings of their subjective appeal or predefined dimension, while others used external variables such as clicks or likes. Oftentimes labels are discrete, so that one yields a classification task [11, 16], whereas ours is given by regression task where the predicted variable is itself the result of another machine learning classifier.

Datasets are primarily of heterogeneous nature. That is, they include open-domain images from, e. g., Flickr [e. g., 34, 36], combinations of landscapes and faces [16], or multi-category product images [8, 11]. Hence, the task is to recognize objects that are linked to a certain sentiment (e. g., "spider" = negative; "cat" = positive). In contrast datasets of identical objects as in this work are rare (e. g., "angry cat" = negative; "playful cat" = positive).

Methods were earlier chosen to be feature-based classifiers [e. g., 5, 28], yet are nowadays replaced by deep learning; specifically convolutional neural networks represent the state-of-the-art [e. g., 26, 36]. We later customize pre-trained neural networks by a tailored form of transfer learning to cope with limited data as in our setting.

3 METHODS

3.1 Problem Statement

Let $y_i \in \mathbb{R}$ denote the price variable[1] included in listing $i = 1, \ldots, n$. Each listing further entails an image $x_i \in \mathbb{R}^{H \times W}$, as well as J additional covariates $c_i \in \mathbb{R}^J$ characterizing the attributes of product i. This lets us state our research question: to what extent does x_i describe y_i while simultaneously controlling for all covariates?

Formally, we arrive at an hedonic regression [25] problem

$$y_i = \alpha + \beta\, f_\theta(x_i) + \sum_{j=1}^{J} \gamma_j c_{ij} \tag{1}$$

with unknown coefficients α, β, and $\gamma_1, \ldots, \gamma_J$. The unknown function f_θ parameterized by θ represents the image sentiment. Thus, the term $\beta\, f_\theta$ yields the marginal contribution from the image x_i.

The image sentiment $f_\theta : \mathbb{R}^{H \times W} \to \mathbb{R}$ takes the pixels from image x_i as input, and maps the image to a single real value. This function f_θ must be high-dimensional and inherently non-linear.

Differences to pure image sentiment analysis: Our estimation problem from Equation (1) shows crisp differences to how image sentiment was previously used: prior literature was concerned with a

[1] In the following, we use the terms price and rent interchangeably.

naïve prediction task, namely, $f'_\theta : x_i \mapsto y'_i$ in the absence of covariates, where image labels y'_i were the variable of interest. In contrast, our objective is statistical significance testing, i. e., $H_0 : \beta = 0$. Notably, a coefficient β would not be present in a pure prediction task. Further, obtaining simple point estimates of the coefficients α, β, and $\gamma_1, \ldots, \gamma_J$ is not sufficient; instead, rigorous statistical inferences with confidence intervals are needed. Here it is essential that we include the covariates, so that the potential heterogeneity among products is properly controlled for.

3.2 Proposed Approach

Our estimation problem from Equation (1) is solved by the following three-staged approach:[2]

Stage (1) computes adjusted prices. These explain the variance in original price that is unexplained by other product attributes. Formally, it fits a linear model regressing the price variable y_i on the control variables $c_{i1}, \ldots, c_{iJ}$. The residuals $\tilde{y}_i = y_i - \hat{y}_i$ are then used to train the function f_θ in the next stage.

Stage (2) trains the function f_θ that maps the image x_i to the image sentiment using the training labels $\tilde{y}_i$ from stage (1). Given a predefined $f_\theta(\cdot)$, it returns the estimated parameters $\hat{\theta}$. For modeling f_θ, we utilize a procedure based on convolutional neural networks (CNN) as in earlier research [e. g., 33, 35]; however, we use a tailored transfer learning approach.

Stage (3) takes the function $f_{\hat{\theta}}$ with estimated parameters $\hat{\theta}$ as input and, based on it, calculates the image sentiment $\sigma_i = f_{\hat{\theta}}(x_i)$ for each listing i. Then, it performs statistical inference based on the rewritten model formulation

$$y_i = \alpha + \beta\, \sigma_i + \sum_{j=1}^{J} \gamma_j c_{ij}, \tag{2}$$

so that we obtain estimates for the coefficients α, β, and γ.

Separate parts of the data are used for stages (1)–(2) vs. stage (3). We further emphasize that different dependent variables appear, namely the price y_i in stage (1), the residual $\tilde{y}_i$ in stage (2), and the actual price y_i again in stage (3). Notably, the control variables find application in all stages: In stage (1), they provide the basis for computing residuals, i. e., the adjusted price, so that the image sentiment can later only explain the variance beyond these controls. Without such a price adjustment, the framework would later learn learn observable product characteristics, such as the apartment size, inside the image sentiment. In stage (2), the controls occur implicitly, as we learn the relationship between images and the residuals, i. e., the price that has previously been adjusted for the controls. In stage (3), the controls describe the between-listing heterogeneity.[3] Later the above approach is further accompanied by various robustness checks including, e. g., different neural network architectures, feature-based classifiers from traditional machine learning, various model configurations, and instrumental variables estimation [1] in order to obtain causal estimates.

Our three-staged computational approach has a clear advantage: we refrain from making specific assumptions of what a positive image content characterizes and circumvent the need for a universal definition. This is in line with earlier findings according to which aesthetics are highly subjective [16]. For instance, some would argue that the quality of the photo itself is important (e. g., no blurriness), while others argue in favor of the aesthetics of the shown content (e. g., a beautiful fireplace). Also, visual appeal could potentially vary with the underlying product (e. g., a college apartment should be designed more hipster than a wild west ranch). In keeping with the previous arguments, we deliberately follow a data-driven approach where we *learn* such characteristics inside f_θ. This is important for our research, since we want to quantify the combined effect of the image content on prices.

3.3 Stage 1: Price Adjustment

The image sentiment $f_\theta(x_i)$ should help to explain the variance in price that is unexplained by other product attribute. To this end, we fit the following linear model

$$y_i = \delta + \sum_{j=1}^{J} \vartheta_j c_{ij}. \tag{3}$$

to the training data. The residuals $\tilde{y}_i = y_i - \hat{y}_i$ are then used in the next stage, i. e., the image sentiment f_θ is trained based on them.

3.4 Stage 2: Image Sentiment

In stage (2), the image sentiment is computed. A pre-trained CNN from computer vision was used, but it was modified in order to fine-tune it to our dataset via transfer learning. The original CNN is given by VGG-16 [29], which represents a state-of-the-art architecture [e. g., 12] for object detection. Originally, the network consisted of 16 layers (13 convolutional layers and 3 fully-connected ones) for the purpose of classification. In contrast, our setting involves a regression over a continuous output. To this end, the network is tailored to our objective of image sentiment analysis: we apply transfer learning during which we modified the architecture to continuous output by adding an additional fully-connected layer with a single neuron to the network. This single neuron eventually outputs the image sentiment as continuous variable.

Fine-tuning to our dataset was achieved by training the existing fully connected layers, as well as the new additional layer while keeping the convolutional layers fixed. Here, the weights of the original output layer were reset and then randomly initialized. The weights of all other fully-connected layers were then trained with a lower learning rate. The model is trained by minimizing the Euclidean loss between $f_\theta(x_i)$ and the price residual $\tilde{y}_i$. Transfer learning is required to learn a new task with relatively few samples as in our research. Owed to it, we benefit from the pre-trained weights that were obtained on large-scale computer vision datasets.

3.5 Stage 3: Hedonic Regression

Based on the price y_i, we conduct a so-called hedonic regression [25]: according to it, products are differentiated by attributes that describe the overall price, yet where each attribute is not a product

[2] The R code for the analyses is available in the supplements and at https://github.com/cfnaumzik/ImageSentiment.git.

[3] The differences across stages also distinguish our approach from double machine learning as a way to establish causality.Likewise, our framework, despite causal, differs from traditional causal modeling, as we are not interested in estimating the treatment effect but infer the unbiased coefficient β.

itself, but where its implicit contribution to the overall price can be modeled [25]. This allows us to isolate the contribution of image content to the price and compare its marginal effect to that of other product attributes. We reiterate that the image sentiment variable is only the contribution of the image *beyond* the controls that describe the between-listing heterogeneity. As our baseline estimator, we chose ordinary least squares (OLS). Following best practice, we tested for auto-correlation and used heteroskedastic-consistent estimators[4] for the standard errors of the parameter estimates [32].

4 DATA

4.1 Real Estate Listings

Our empirical findings are based on real estate listings, since potential tenants have named images as the most important feature for online platforms.[5] Furthermore, this setting is based on a homogenous products (i. e., only real estate) with a large variety in appearance (i. e., the interior of no two apartments will look identical, thus allowing us to collect observations of image sentiment from a single product category at scale). This is different from alternative online platforms such as eBay, which usually feature a heterogeneous products (i. e., images of variable quality, yet featuring identical items such as the same smartphone).

We collected rental offers in the metropolitan area of Boston, MA. We chose this particular city for two reasons. First, real estate pricing in Boston has been serving as a baseline in academic research, since the inaugural work by Harrison and Rubinfeld [13]. Furthermore, the market in this city is highly competitive and, hence, the offered price should be close – if not identical – to the settled price. This is also confirmed by the fact that the listed rent is comparable to the official US Market Rent by the US Department of Housing and Urban Development [3]. This should rule out potential biases (i. e., that the reported price deviates from the settled one).

We collected a dataset consisting of a total of 26,461 apartment listings with at least one image. The monthly rent ranged between 600 [$] and 7250 [$], while apartments were located in one of 113 districts in the Boston metropolitan area. The dataset was then randomly partitioned into two subsets, namely, a training set with 80 % of the samples (i. e., 21,168) and a test set with the remaining 20 % (i. e., 5,293). The training set was used during the first two stages of our approach, while the test set was used exclusively in stage (3).

4.2 Variables

Each apartment listing is accompanied by the following covariates: (i) the price in form of the monthly rent (in US$), (ii) the size (in sqft), (iii) the number of bath-/bedrooms and, (iv) and the locations as districts. As noted in prior literature, prices are largely determined by the aforementioned baseline variables [e. g., 13]. Hence the need to use the price adjustments for training becomes evident, since, otherwise, the image sentiment in stage (2) would learn observable product characteristics from the image (e. g., the size of the apartment). Additionally, our sample contains (v) the number of images, and a (vi) user-generated description in narrative form (as a concatenation of both title and the additional free text). Table 1 lists summary statistics for the key variables.

Table 1: Descriptive statistics of key variables.

	Mean	SD	Median
Dependent variable:			
Monthly rent (in US$)	2958.22	1003.92	2800.00
Covariates/controls:			
Size (in sqft)	1099.54	435.03	1017.00
Number of bedrooms	2.29	1.11	2.00
Number of bathrooms	1.37	0.54	1.00
Number of images	8.14	4.80	7.00
Description length (in words)	144.97	85.98	131.00

SD = standard deviation

Following common practice in OLS estimation, the dependent variable was set to the monthly rent in log values. Since prices tend to be log-normally distributed, the log-transformation is supposed to reduce the risk of heteroskedasticity (non-constant variance of the errors); cf. Wooldridge [32]. Similarly, the size of the apartment was also subject to a log transformation. We refer to apartment size, number of bedrooms/bathrooms, and district dummies as controls.

5 EMPIRICAL FINDINGS

5.1 Relationship between Image Sentiment and Price

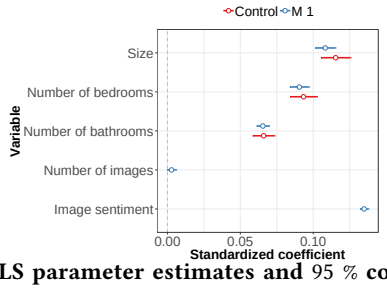

Figure 1: OLS parameter estimates and 95 % confidence intervals (based on heteroskedastic-consistent standard error estimates). Excluding district dummies for readability.

Figure 1 reports the empirical results (obtained for the test set). Model M1 presents our model of interest from Equation (1), based on which we find supporting evidence for our hypothesis: image sentiment is positively associated with the price variable (p-value < 0.01 %). The standardized coefficient amounts 0.135 and can be interpreted as follows: a one standard deviation higher image sentiment corresponds – ceteris paribus – to an increase in the monthly rent by $e^{0.135} - 1 \approx 14.45\%$. Model M1 additionally includes the number of images in the listing. We find that it is positively linked to the price variable (p-value < 0.01 %). However, the standardized coefficient for the number of images is significantly lower than the coefficient for image sentiment (0.135 compared to 0.003); hence, suggesting a more important role of image sentiment.

As a comparison, Figure 1 also lists a control model that considers only the control variables (i. e., apartment size, number of

[4]Estimation was implemented in R using the package `sandwich`.
[5]https://www.nar.realtor/reports/real-estate-in-a-digital-age

bathrooms/bedrooms, and district dummies). The control model was included to confirm that the coefficient estimates remain robust.

In sum, we point out that the standardized coefficient belonging to image sentiment is the largest in size. We also note that per Table 3 model M1 is preferred when comparing the models in terms of Akaike information criterion (AIC), Bayesian information criterion (BIC), and explained variance (adjusted R^2). In fact, omitting the image sentiment from model M1 reduces the adjusted R^2 by 16.7 percentage points. Hence the results confirm our hypothesis that the image sentiment helps in explaining the variance in price beyond other product characteristics.

5.2 Instrumental Variables Estimation

We now estimate the causal effect of image sentiment σ on price y. It is known that, for an endogenous variable, the OLS estimates are not consistent and produce a biased estimate, not reflecting the true causal effect of the variable [e. g., 1]. We address possible issues of endogeneity with respect to the image sentiment variable $\hat{\sigma}$ by conducting an instrumental variables (IV) estimation [32, Ch. 6].[6] It corrects a potential bias by considering an instrument z for σ. Loosely speaking, an instrument z is a variable that affects y not directly but only through its effect on σ. This construction allows one to obtain an estimate of the causal effect outside of a controlled experiment: If one finds a correlation between the instrument z and price variable y, this may be seen as evidence that σ has a causal effect on y, since z can effect y by construction only through σ [32].

We experimented with two different instruments due to challenges of finding a strong instrument for our research and each alleviating different concerns. Eventually, we decided upon two choices, namely (a) the average blue color across all pixels of the image and (b) the image sentiment σ_{i-1} from the previous listing $i-1$. Instrument (a) has no semantic meaning and should thus be largely random without direct impact on y, i. e., only through the image sentiment variable. Instrument (b) is informed by common practice in social sciences [e. g., 24] and, in our case, is obviously unrelated to the outcome y, as it stems from a different apartment ($i \neq i-1$) and is thus independent.

We followed common checks in instrumental variable estimations [32, Ch. 6].[7] We then compared the IV estimates for β to the OLS estimate. Most importantly, we find that the coefficients remain statistically significant. Also, the relative ordering of the variables and the size of the coefficients remain robust, i. e., we obtained a image sentiment coefficient of 0.146 for instrument (a), and 0.167 for instrument (b).

We also ran a third instrumental variables estimation with both (a) and (b) as instruments.[8] The estimated coefficient for image sentiment remains statistically significant at the 0.01 % level. The coefficient amounts to 0.156 which is similar to the OLS estimate (0.135). Together with the results from the Wu-Hausmann test, we can conclude that the estimated coefficients are robust to endogeneity and image sentiment has a causal effect on the price variable.

[6] We used the function `ivreg` from the `R` package `AER`.

[7] We first discern weak instruments. For this purpose, we used a heteroskedasticity-corrected Wald test. We obtained statistically significant F-values for both instruments, i. e., the F-statistic amounts to 57.41 for instrument (a) and 49.95 for (b). This confirms the strength of both instruments. We then tested for exogeneity of the image sentiment variable by running a regression-based Wu-Hausmann test. For instrument (a), the resulting p-value was 0.471, while, for instrument (b), the p-value numbered to 0.046.

[8] Again, the Wald test confirmed the strength of the instruments, yielding an F-value of 53.817. The Wu-Hausmann test returned a p-value of 0.058. Since we used two instruments simultaneously, we confirmed their validity using a Hansen-Sargan test. This resulted in a p-value of 0.334.

5.3 Predictive Power of Image Sentiment

Next we analyze the predictive power of image sentiment. To this end, we randomly partitioned the test data into two subsets of 3730 (75 % of the test set) and 1241 observations respectively. The first set was used to train a range of models using the formulation in Equation (1), namely, a linear model (LM) fitted via OLS, a support vector regression (SVR), and a random forest (RF).[9] The second subset of the data was then used to evaluate the out-of-sample performance. The results of the predictions are detailed in Table 2. Evidently, the inclusion of the image sentiment variable greatly improves the predictive power of each model. Compared to the baseline model, the inclusion of the image sentiment variable improves the root mean squared error by 8.11 % for the random forest and by 26.22 % for the linear model. Even the image sentiment alone reduces the error below a naïve baseline such as the sample mean (0.33). This confirms the capacity of image sentiment for predicting prices from it.

Table 2: Out-of-sample performance across models in root mean squared error for predicting logarithm of monthly rent.

Model	Variables		Models		
	Controls	Image sentiment	LM	SVR	RF
Relative change from including image sentiment					
Controls	✓		0.188	0.176	0.166
Model M1	✓	✓	**0.139**	**0.137**	**0.153**
Ablation study: image sentiment only					
Sensitivity check		✓	0.297	0.295	0.275

The lowest error in each column is highlighted in bold.

5.4 Comparison to Textual Descriptions

Product information is also available in textual form in addition to the images. Therefore, we extract the information in the product description by following previous research [e. g., 4, 30] and analyze the text sentiment. Formally, we implemented the machine learning approach from Pröllochs et al. [23]: it preprocesses the text in order to extract polarized language and then trains a LASSO that should be robust against overfitting (i. e., when there are more terms than observations). Analogous to image sentiment, the machine learning classifier was trained on the residuals from Equation (3). We refer to the resulting variable as text sentiment.

We now study model M2 in which image sentiment is compared to text sentiment. The results are reported in Figure 2. We find that the resulting variable for text sentiment is statistically significant at the 0.01 % level and links positively with the logarithm of the monthly rent. However, its standardized coefficient (0.029) is considerably smaller than that belonging to image sentiment which remains robust (0.125). Hence, while a one standard deviation change in image sentiment corresponds to a price increase by

[9] All model hyperparameters were tuned via 10-fold cross-validation.

Table 3: Model comparison across AIC, BIC, and adjusted R^2. Best fit per subgroup is highlighted in bold.

		Model fit		
Model	**Variables**	Adj. R^2	AIC	BIC
M0	Controls	71.32	-3041.20	-2207.74
M1	Controls, image sentiment, number of images	**88.07**	**-7402.60**	**-6556.12**
M1a	Controls, image sentiment	**88.07**	-7401.75	-6561.79
M1b	Controls, number of images	71.33	-3042.85	-2202.88
M2	Controls, image sentiment, number of images, description sentiment, description length	**88.72**	-7676.37	-6816.87
M2a	Controls, image sentiment, description sentiment	88.71	-7673.58	**-6827.10**
M2b	Controls, image sentiment, description sentiment, description length	88.72	**-7677.32**	-6824.33

Controls include the apartment size, the number of bathrooms, the number of bedrooms, and the district dummies.

13.28 %, a one standard deviation increase in text sentiment is only associated with an increase by 2.94 %.

Additionally, we compare both text and image sentiment against the length of the textual description. The corresponding coefficient is positive (as expected) and further statistically significant (p-value < 5 %). We further quantify the relative importance of image sentiment in comparison to the length of the textual description. Here we find a considerably larger standardized coefficient for image sentiment (0.125) as compared to description length (0.004). Based on it, we can compute that a one standard deviation increase in image sentiment equals to the same increase in the monthly rent as an additional 2856.03 words in the textual description. Hence, the previous results point towards a picture superiority effect.

Figure 2: OLS parameter estimates and 95 % confidence intervals (based on heteroskedastic-consistent standard error estimates). Excluding controls for readability.

Finally, we also comment on the overall model fit in Table 3. When including description sentiment and length, the adjusted R^2 improves only slightly, e. g., from 0.881 to 0.887. This highlights that the text sentiment of the description and its length play only a minor role in explaining the variance of prices, as opposed to the decisive role of the image sentiment.

5.5 Robustness Checks

Classifiers: We utilized various alternative classifiers in order to ensure the robustness of our results. Specifically, we used a customized version of AlexNet [18] as a different pre-trained neural network that was developed for image classification and which we modified for analyzing image sentiment.[10]

We further compared image sentiment variables arising form the different classifiers inside the regression from stage (3).[11] We find consistent evidence that the image sentiment variable is statistically significant (p-value < 0.1 %). Overall, the size of the coefficient remains stable; the adjusted R^2 shows only minor variations of below 3 percentage points with the VGG-16 model being ranked at the top; finally, both AIC/BIC prefer the model based on VGG-16.

Dependent variable: Our previous results are based on the log-transformed monthly rent as our price variable. We additionally experimented with other dependent variables, namely, the monthly rent as absolute values and a relative rent per sqft. Still, the image sentiment variable consistently evinced to be statistically significant at the 0.1 % level, thus bolstering the robustness of our findings.

Quantile regression: A reasonable assumption is that the effect of image sentiment may vary in certain price segments. Hence, we employ a quantile regression [17] where coefficients are estimated at different quantiles of the price. We find that the effect size of the image sentiment variable is more pronounced for more expensive apartments. While, for instance, a one standard deviation change in image sentiment corresponds to an increase in the absolute rent by 407.39 $ for an apartment at the median, it attains an increase by 430.87 $ for the top-10 % quantile of the price distribution.

Non-linear relationships: We inspected a potential relationship by including higher-order terms with respect to image sentiment. However, this did not improve the model fit. Including only the quadratic or cubic term of the image sentiment even yielded an inferior fit, so that such a model specification should be discouraged.

Error correction: We accounted for the error of our predictive models that score the image sentiment of apartments. For this reason, we applied the SIMEX procedure as proposed by Cook and Stefanski [6].[12] The estimates from model M1 using the SIMEX algorithm are consistent with the results from OLS. The parameter β remains statistically significant at the 0.01 % level. Again, the model fit is considerably improved over the control model when including the image sentiment.

Model comparison: Table 3 compares the model fit across different specifications. We report the AIC and BIC as information criteria, as well as the adjusted R^2 measuring the explained variance.

[10] AlexNet was modified and fine-tuned analogous to VGG-16. We added a fully-connected layer to the existing architecture. Fine-tuning occurred via the fully-connected layers, where the last layer of the original network was reset and randomly initialized. Compared to VGG-16, the AlexNet architecture entails only 8 instead of 16 layers, which should theoretically reduce the risk of overfitting but also limit flexibility.

[11] We further explored classifiers from traditional machine learning where no pre-training was involved, i. e., a random forest and support vector regression (SVR) with a radial kernel. The preprocessing for these classifiers had to be adapted slightly: data augmentation was omitted as it is usually inherent to deep learning; however, overfitting was prevented by an additional conversion to grayscale and setting the size to 100×100. Hyperparameters were tuned via 10-fold cross-validation.

[12] We use the `R` package `simex`. A brief introduction to the SIMEX method and its application can be found in https://cran.r-project.org/doc/Rnews/Rnews_2006-4.pdf.

Across all metrics, the results highlight the overall importance of image sentiment in explaining the price variable. In contrast, the textual description is hardly able to improve the model fit.

REFERENCES

[1] Joshua D Angrist, Guido W Imbens, and Donald B Rubin. 1996. Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* 91, 434 (1996), 444–455.

[2] Nikolay Archak, Anindya Ghose, and Panagiotis G Ipeirotis. 2007. Show me the money! Deriving the pricing power of product features by mining consumer reviews. In *KDD*.

[3] Geoff Boeing and Paul Waddell. 2017. New insights into rental housing markets across the United States: Web scraping and analyzing Craigslist rental listings. *Journal of Planning Education and Research* 37, 4 (2017), 457–476.

[4] Johan Bollen, Huina Mao, and Alberto Pepe. 2011. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *ICWSM*.

[5] Damian Borth, Rongrong Ji, Tao Chen, Thomas Breuel, and Shih-Fu Chang. 2013. Large-scale visual sentiment ontology and detectors using adjective-noun pairs. In *MM*. 223–232.

[6] J R Cook and L A Stefanski. 1994. Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.* 89, 428 (1994), 1314–1328.

[7] Michaël Dewally and Louis Ederington. 2006. Reputation, certification, warranties, and information as remedies for seller–buyer information asymmetries: Lessons from the online comic book market. *Journal of Business* 79, 2 (2006), 693–729. https://doi.org/10.1086/499169

[8] Wei Di, Neel Sundaresan, Robinson Piramuthu, and Anurag Bhardwaj. 2014. Is a picture really worth a thousand words?. In *WSDM*.

[9] Angelika Dimoka, Yili Hong, and Paul A Pavlou. 2012. On product uncertainty in online markets: Theory and evidence. *MIS Quarterly* 36, 2 (2012), 395–426.

[10] Soussan Djamasbi, Marisa Siegel, and Tom Tullis. 2010. Generation Y, web design, and eye tracking. *International Journal of Human-Computer Studies* 68, 5 (2010), 307–323.

[11] Anjan Goswami, Sung H Chung, Naren Chittar, and Atiq Islam. 2012. Assessing product image quality for online ohopping. In *Image Quality and System Performance IX (SPIE Proceedings)*.

[12] Jinyoung Han, Daejin Choi, Jungseock Joo, and Chen-Nee Chuah. 2017. Predicting popular and viral image cascades in Pinterest. In *ICWSM*.

[13] David Harrison and Daniel L Rubinfeld. 1978. Hedonic housing prices and the demand for clean air. *Journal of Environmental Economics and Management* 5, 1 (1978), 81–102.

[14] Sylvia Jansen, Harry Boumeester, Henny Coolen, Roland Goetgeluk, and Eric Molin. 2009. The impact of including images in a conjoint measurement task: Evidence from two small-scale studies. *Journal of Housing and the Built Environment* 24, 3 (2009), 271–297.

[15] Minjeong Kim and Sharron Lennon. 2008. The effects of visual and verbal information on attitudes and purchase intentions in Internet shopping. *Psychology and Marketing* 25, 2 (2008), 146–178.

[16] Won-Hee Kim, Jun-Ho Choi, and Jong-Seok Lee. 2018. Objectivity and subjectivity in aesthetic quality assessment of digital photographs. *IEEE Transactions on Affective Computing* (2018).

[17] Roger Koenker and Gilbert Bassett. 1978. Regression quantiles. *Econometrica* 46, 1 (1978), 33–50.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *NIPS*.

[19] Gregory Lewis. 2011. Asymmetric information, adverse selection and online disclosure: The case of eBay motors. *The American Economic Review* 101, 4 (2011), 1535–1546.

[20] Deborah J MacInnis and Linda L Price. 1987. The role of imagery in information processing: Review and extensions. *Journal of Consumer Research* 13, 4 (1987), 473–491.

[21] Andrew A Mitchell and Jerry C Olson. 1981. Are product attribute beliefs the only mediator of advertising effects on brand attitude? *Journal of Marketing Research* 18, 3 (1981), 318–332.

[22] Paul A Pavlou and Mendel Fygenson. 2006. Understanding and predicting electronic commerce adoption: An extension of the theory of planned behavior. *MIS Quarterly* 30, 1 (2006), 115–143.

[23] Nicolas Pröllochs, Stefan Feuerriegel, and Dirk Neumann. 2018. Statistical inferences for polarity identification in natural language. *PLOS ONE* 13, 12 (2018), e0209323.

[24] William Robert Reed. 2015. On the practice of lagging variables to avoid simultaneity. *Oxford Bulletin of Economics and Statistics* 77, 6 (2015), 897–905.

[25] Sherwin Rosen. 1974. Hedonic prices and implicit markets: Product differentiation in pure competition. *Journal of Political Economy* 82, 1 (1974), 34–55.

[26] Rasmus Rothe, Radu Timofte, and Luc van Gool. 2016. Some like it hot: Visual guidance for preference prediction. In *CVPR*.

[27] Rossano Schifanella, Miriam Redi, and Luca M Aiello. 2015. An image is worth more than a thousand favorites: Surfacing the hidden beauty of Flickr pictures. In *ICWSM*.

[28] Stefan Siersdorfer, Enrico Minack, Fan Deng, and Jonathon Hare. 2010. Analyzing and predicting sentiment of images on the social web. In *MM*.

[29] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (sep 2014). arXiv:1409.1556 http://arxiv.org/abs/1409.1556

[30] Andranik Tumasjan, Timm O Sprenger, Philipp G Sandner, and Isabell M Welpe. 2010. Predicting elections with Twitter: What 140 characters reveal about political sentiment. In *ICWSM*.

[31] Jinpeng Wang, Wayne Xin Zhao, Yulan He, and Xiaoming Li. 2015. Leveraging product adopter information from online reviews for product recommendation. In *ICWSM*.

[32] Jeffrey M Wooldridge. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

[33] Jufeng Yang, Ming Sun, and Xiaoxiao Sun. 2017. Learning visual sentiment distributions via augmented conditional probability neural network. In *AAAI*.

[34] Yilin, Wang, Yuheng, Hu, Subbarao, Kambhampati, Baoxin, and Li. 2015. Inferring sentiment from web images with joint inference on visual and social cues: A regulated matrix factorization approach. In *ICWSM*.

[35] Quanzeng You, Hailin Jin, and Jiebo Luo. 2017. Visual sentiment analysis by attending on local image regions. In *AAAI*.

[36] Quanzeng You, Jiebo Luo, Hailin Jin, and Jianchao Yang. 2015. Robust image sentiment analysis using progressively trained and domain transferred deep networks. In *AAAI*.

[37] Shunyuan Zhang, Dokyun Lee, Param Singh, and Kannan Srinivasan. 2016. How much is an image worth? An empirical analysis of property's image aesthetic quality on demand at Airbnb. In *ICIS*.