Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp

Mengxia Zhang, Lan Luo

To cite this article:
Mengxia Zhang, Lan Luo (2022) Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp. Management Science

Published online in Articles in Advance 04 Apr 2022

. https://doi.org/10.1287/mnsc.2022.4359

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–26
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Can Consumer-Posted Photos Serve as a Leading Indicator of Restaurant Survival? Evidence from Yelp

Mengxia Zhang,[a,*] Lan Luo[b]

[a] Marketing, Western University Ivey Business School, London, Ontario N6A 3K7, Canada; [b] Marketing, University of Southern California, Los Angeles, California 90089

*Corresponding author

Contact: mezhang@ivey.ca, https://orcid.org/0000-0002-4901-0635 (MZ); lluo@marshall.usc.edu, https://orcid.org/0000-0003-1002-8237 (LL)

Received: October 26, 2019
Revised: May 4, 2021; October 28, 2021
Accepted: January 22, 2022
Published Online in Articles in Advance: April 4, 2022

https://doi.org/10.1287/mnsc.2022.4359

Abstract. Despite the substantial economic impact of the restaurant industry, large-scale empirical research on restaurant survival has been sparse. We investigate whether consumer-posted photos can serve as a leading indicator of restaurant survival above and beyond reviews, firm characteristics, competitive landscape, and macroconditions. We employ machine learning techniques to extract features from 755,758 photos and 1,121,069 reviews posted on Yelp between 2004 and 2015 for 17,719 U.S. restaurants. We also collect data on restaurant characteristics (e.g., cuisine type, price level) and competitive landscape as well as entry and exit (if applicable) time from each restaurant's Yelp/Facebook page, own website, or Google search engine. Using a predictive XGBoost algorithm, we find that consumer-posted photos are strong predictors of restaurant survival. Interestingly, the informativeness of photos (e.g., the proportion of food photos) relates more to restaurant survival than do photographic attributes (e.g., composition, brightness). Additionally, photos carry more predictive power for independent, young or mid-aged, and medium-priced restaurants. Assuming that restaurant owners possess no knowledge about future photos and reviews, photos can predict restaurant survival for up to three years, whereas reviews are only predictive for one year. We further employ causal forests to facilitate the interpretation of our predictive results. Among photo content variables, the proportion of food photos has the largest positive association with restaurant survival, followed by proportions of outside and interior photos. Among others, the proportion of photos with helpful votes also positively relates to restaurant survival.

History: Accepted by Juanjuan Zhang, marketing.
Funding: The authors thank Nvidia and Clarifai for supporting this research.
Supplemental Material: The online appendix and data are available at https://doi.org/10.1287/mnsc.2022.4359.

Keywords: user-generated content • photos • word of mouth • analytic modeling • computer vision • image analysis • text mining • Yelp • business survival • restaurants

1. Introduction

The restaurant industry has a substantial impact on the U.S. economy. According to the National Restaurant Association, the U.S. restaurant industry generated more than $833 billion in revenue and jobs for 1 in 10 workers in 2018. Meanwhile, this industry is also well known for its high turnover rate. According to Parsa et al. (2005), the first-year turnover rate of restaurants is as high as 26%. Nevertheless, large-scale empirical research on restaurant survival is scarce.

With the extensive use of camera-enabled smartphones and the increasing popularity of various photo-sharing platforms, three billion photos are shared on social media daily (McGrath 2017). The number of photos taken by consumers in 2017 was projected to be 1.3 trillion globally (*The New York Times*, July 29, 2015).[1] Compared with many other industries, the restaurant industry is also unique in that consumers love to share photos of their dining experience online (*The New York Times*, April 6, 2010).[2] Within this context, the primary goal of this research is to investigate whether photos may serve as a leading indicator of restaurant survival above and beyond alternative factors such as restaurant characteristics, competitive landscape, etc.

Historically, it is well-documented in the business survival literature (e.g., Parsa et al. 2005, Lafontaine et al. 2018) that firm characteristics, competition, and macroconditions (listed in Table 1) are the main factors associated with business survival. Many businesses with consumer-posted photos also receive consumer reviews that contain rich information about consumers' descriptions and/or opinions toward their consumption experiences. As such, it is not evident that consumer-posted photos would play a role in predicting business survival after all these alternative factors are controlled for.

Table 1. Review of Prior Literature on Business Survival

Factors		Literature
Company	Firm age	Carroll (1983), Kalleberg and Leicht (1991), Bates (1995), Fritsch et al. (2006)
	Ownership type (e.g., chain versus independent)	Kamshad (1994), Audretsch and Mahmood (1995), Cooper et al. (1994), Bates (1995, 1998), Kalnins and Mayer (2004), Parsa et al. (2005, 2011)
	Type of business (e.g., cuisine types)	Parsa et al. (2005)
	Operation (e.g., price, service, environment)	Kalleberg and Leicht (1991), Audretsch and Mahmood (1995), Audretsch et al. (2000), Tveterås and Eide (2000), Parsa et al. (2010, 2015)
Competition	Concentration	Kalleberg and Leicht (1991), Wagner (1994), Parsa et al. (2005), Fritsch et al. (2006)
	Number of entries	Fritsch et al. (2006)
Macro	National conditions	Audretsch and Mahmood (1995), Boden and Nucci (2000), Fritsch et al. (2006), Parsa et al. (2010)
	Local environment	Kalleberg and Leicht (1991), Fritsch et al. (2006), Haapanen and Tervo (2009), Parsa et al. (2010, 2011, 2015), Kalnins and Lafontaine (2013)

On the flip side, consumer-posted photos may also contain unique information that cannot be directly captured by firm characteristics, competitive landscape, macroenvironment, or consumer reviews. We conjecture that there are two possible routes in which photos may be useful in predicting restaurant survival. First, from the posters' perspective, it takes very little effort to upload photos on social media. As such, photos may serve as a proxy of restaurant popularity and/or reflect a restaurant's situation (e.g., well-managed/deteriorating restaurants may be reflected by customer-posted photos on platforms such as Yelp) above and beyond reviews. Second, from the viewers' perspective, photos may communicate valuable information about the restaurant with great efficiency, hence facilitating a better match between the focal business and its customers. Scientific studies show that human brains can process a photo in as little as 13 milliseconds (*MIT News*, January 16, 2014).[3] The brain also processes visual information 60,000 times faster than text (Vogel et al. 1986). On Yelp and TripAdvisor, businesses with 10+ photos receive double to triple the number of views over businesses with the same number of reviews and no photos.[4] Furthermore, photos visualize rich information about the restaurant (e.g., food items served, ambiance), revealing whether the focal restaurant matches the viewer's horizontal taste. For example, a consumer may not be able to figure out whether the consumer may like a seafood pizza based on a review stating "it is the most unique pizza I had!" But the consumer can clearly visualize the pizza from a photo and discern whether this is a food item that the consumer may like. In a similar vein, Ghose et al. (2012) show that hotel ranking systems incorporating photos can generate recommendations with a better fit compared with those with no photos. Similarly, we conjecture that photos may help viewers better visualize how much they might enjoy the food items and/or the dining experience.

Thus, although the survival of a restaurant might not hinge on the experience of a single consumer who either posts photos or views photos posted by others, consumer-posted photos may collectively provide some useful information that can be used to forecast its survival potential. Given that the restaurant industry is abundant with rich information on consumer reviews, restaurant characteristics, and the competitive landscape, it provides us with an ideal setting to explore answers to the following questions: (1) Can consumer-posted photos serve as a leading indicator of restaurant survival? (2) If so, what aspects of photos are more and less important? (3) Are photos more informative for certain types of restaurants? (4) How long can photos stay predictive in forecasting restaurant survival?

Specifically, we employ machine learning methods to extract various features from 755,758 photos posted on Yelp between October 2004 and December 2015 for 17,719 U.S. restaurants, among which 25.37% went out of business during this time window. To investigate the incremental predictive power of photos, we also extract features from 1,121,069 reviews for these restaurants during the same time window as controls. Built upon the business/restaurant survival literature (e.g., Parsa et al. 2005, Lafontaine et al. 2018), we further collect data on these restaurants' characteristics (e.g., chain status, cuisine type, price level), competitive landscape (e.g., restaurant concentration, new entries/exits, photos and reviews of competitors), entry time (from each restaurant's Yelp/Facebook page, own website, or the Google search engine) along with data on macrofactors (year and zip code) as additional controls. Given that unobserved restaurant quality may impact both consumer-posted photos and survival, we employ deep learning techniques to extract restaurant quality measures along four dimensions (food, service, environment, and price) via deep mining of 1,121,069 reviews.[5] We emphasize these four dimensions of restaurant quality based on the prior literature (e.g., Hyun 2010, Ryu et al. 2012, Bujisic et al. 2014). Such an approach is inspired by the recent trend of utilizing consumer reviews to track product or service quality over time (e.g., Tirunillai and Tellis 2014, Hollenbeck 2018).[6] We then employ an implementation of gradient boosting trees called XGBoost algorithm (Chen and Guestrin 2016) to discern the incremental predictive power of photos with all other variables served as controls.

We discover that consumer-posted photos can significantly improve forecast accuracy for restaurant survival above and beyond reviews, restaurant characteristics, competitive landscape, and macroconditions. We further explore the incremental predictive power of various aspects of photos, including content, photographic attributes, caption, volume, and helpful votes. We learn that the informativeness of photos (e.g., the proportion of food photos) relates more to restaurant survival than do photographic attributes (e.g., composition, brightness). Particularly, the cumulative proportion of food photos turns out to be the most predictive variable of restaurant survival among all photo-related variables. This is potentially because food photos vividly demonstrate the food items served by the focal restaurant, revealing whether the restaurant matches the viewer's private taste. And such a match is often necessary to get consumers in the door. This result echoes prior literature suggesting that food is the most critical aspect of a restaurant (Sulek and Hensley 2004, Duarte Alonso et al. 2013). Additionally, we learn that photos are more informative for independent (versus chain), young or midaged (versus established), and medium-priced (versus low-priced) restaurants. Assuming that restaurant owners do not possess any knowledge about future photos and reviews for both themselves and their competitors, photos can predict restaurant survival for up to three years, whereas reviews are informative only for one year.

Despite its many desirable features (such as the ability to explore nonlinear relationships among a large number of variables), the predictive XGBoost algorithm emphasizes maximizing out-of-sample prediction accuracy rather than obtaining unbiased/consistent parameter estimates quantifying how each independent variable (e.g., the proportion of photos with helpful votes) relates to restaurant survival. Aiming to provide a better understanding of how various photo-related (as well as non–photo-related) factors relate to restaurant survival, we further employ cluster-robust causal forests (Athey and Wager 2019, Athey et al. 2019) to facilitate parameter interpretation of the most informative predictors as suggested by the SHAP feature importance (Lundberg et al. 2020) derived from our XGBoost algorithm. As discussed in Athey and Imbens (2016) and Wager and Athey (2018), the causal forests model can serve as a suitable alternative to conventional propensity score methods in inferring treatment effects from rich observational data such as ours. Furthermore, our cluster-robust causal forests allow for clustered errors within a restaurant to account for time-invariant variables that are unobservable to us (e.g., owner education).[7] We learn that, among photo content variables, the proportion of food photos has the largest positive association with restaurant survival, followed by proportions of outside and interior photos. Among others, the proportion of photos with helpful votes is also positively related to restaurant survival.

To our knowledge, this study is among the first to link consumer-posted photos with business survival. Over the past decade, there has been an extensive literature in marketing (e.g., Archak et al. 2011; Netzer et al. 2012, 2019; Tirunillai and Tellis 2012, 2014; Toubia and Netzer 2016; Puranam et al. 2017; Lee et al. 2019; Liu et al. 2019; Timoshenko and Hauser 2019) that emphasizes extracting managerially relevant information from consumer reviews. More recently, several researchers explore the role of photos in consumption experiences (e.g., Diehl et al. 2016; Barasch et al. 2017a, b), advertising (Xiao and Ding 2014), ranking systems (Ghose et al. 2012), brand perceptions (Liu et al. 2020, Dzyabura and Peres 2021), social media engagement (Ko and Bowman 2020, Li and Xie 2020, Hartmann et al. 2021), lodging demand (Zhang et al. 2018), product returns (Dzyabura et al. 2019), crowdfunding success (Li et al. 2019), and the labor market (Malik et al. 2020, Troncoso and Luo 2020). However, few studies explore the relationship between consumer-posted photos and long-term business prosperity. Our research contributes to the literature by filling this void.

Our research also adds to the business/restaurant survival literature by examining whether consumer-posted photos can predict restaurant survival above and beyond reviews and other known factors related to company, competition, and macroconditions. Historically, research on business survival mainly focuses on company characteristics (e.g., Bates 1995, Lafontaine et al. 2018), competitive landscape (e.g., Wagner 1994, Fritsch et al. 2006), and macroconditions (e.g., Audretsch and Mahmood 1995, Boden and Nucci 2000) (see Table 1 for a review of this literature). We contribute by adding user-generated content (UGC, particularly photos) as an additional component to this business survival literature. Given our extensive efforts in extracting as much information as possible from a large number of restaurants over an extended time period, we believe that our research is perhaps among the most comprehensive studies on restaurant survival to date.

Findings from our research can be beneficial to various stakeholders, including business investors, landlords, online platforms, restaurant owners, and research/trade associations. First, managers/investors can better understand the market before starting a new business or investing in existing businesses. Additionally, our research can help landlords who must make high-stakes decisions, such as deciding whether to begin/renew rental leases to new/existing restaurants. The average investment per restaurant in the United States is about half a million (*Restaurant Engine*, May 4, 2015).[8] As of May 2019, the restaurant industry represents about a $375 billion market capitalization.[9] Nevertheless, restaurants are also known to have the highest turnover rates in the retail sector (Parsa et al. 2005, Fritsch et al. 2006). As per our findings, incorporating photos (especially their content information and helpful votes received) into such analysis can significantly increase the prediction accuracy of these important business decisions.

Second, because our research suggests that UGC (especially photos) is useful in forecasting survival, online platforms such as Yelp and TripAdvisor might monetize our work by offering premium analytics reports to their business customers. These platforms already routinely provide business owners with basic analytics reports (e.g., number of consumer-posted photos and reviews). These business owners may also benefit from an enriched report that provides more in-depth information on photos (e.g., photo content, helpful votes received), reviews (e.g., consumer sentiment on food, service, environment, and price), and predicted survival probabilities for both the focal business and all competing businesses within close proximity.

Third, restaurant owners might also leverage our research to advance their competitive intelligence and resource-allocation strategies. As is well-known in the literature (e.g., Kalleberg and Leicht 1991, Parsa et al. 2005, Fritsch et al. 2006), a thorough understanding of the competitive landscape is vital to business/restaurant survival. Based on our predictive model, restaurant owners might decide whether and/or when to initiate aggressive marketing strategies (e.g., offering promotions) upon detecting a decline in the survival probability of a close competitor. Additionally, given that photos can predict restaurant survival for up to three years into the future, our research may also be useful for longer-term strategic planning by restaurant owners.

Finally, our research provides useful insights for research/trade associations such as the National Restaurant Association and the National Tour Association regarding survival probabilities by business characteristics (e.g., chain versus independent, cuisine type, price level), time trends, and macroperformances in the restaurant industry.

Compared with the benefits stated, the cost of utilizing photos in restaurant survival prediction is not high. Extracting information from all photos from tens of thousands of restaurants and calibrating our proposed model only takes a couple of hours or days, depending on the computing engine. Once calibrated, the focal model merely needs to be updated once a year. Given the calibrated model, predicting survival for a new/existing restaurant in the coming year only requires extracting information from recent photos from on the focal restaurant, which also takes minimum computation time.

2. Data

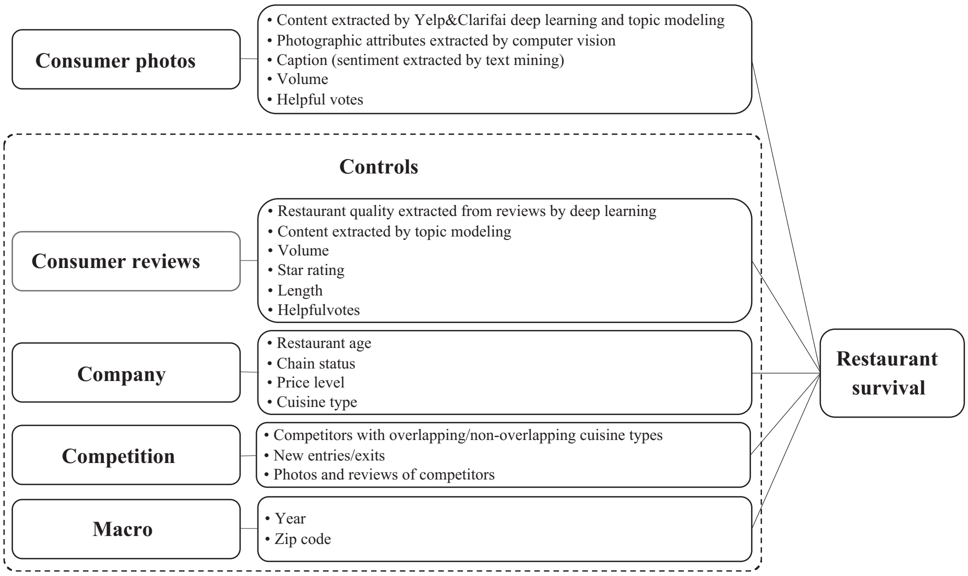
We collected data from 17,719 U.S. restaurants listed on Yelp with a total of 755,758 photos and 1,121,069 reviews between October 12, 2004, and December 24, 2015. Our data set also contains information on firm characteristics, competitive landscape, year, zip code, and entry and exit (if applicable) times for these restaurants. The restaurant and review data come from the Yelp Data Set Challenge round 7. The competitive landscape, zip codes, and exit times are processed from Yelp-provided data. To complement the Yelp data set, we further collected the Yelp photos of all restaurants in our data set. Additionally, we collected data on the entry time of these restaurants from Yelp, the restaurant's website or Facebook page, or via the Google search engine.

Figure 1 depicts the road map of this research. In particular, we focus on studying the role of photos in predicting restaurant survival. All other factors, including reviews, restaurant characteristics, competition, and macroconditions, serve as important control variables in our research. In the following sections, we explain each component of this road map in more detail. Summary statistics of reviews and photos are presented in Table 2. Restaurant-level summary statistics are provided in Table 3. Figure 2 shows the distribution of the restaurant life span broken down by open and closed restaurants in our sample. Table 4 describes all variables related to consumer photos and reviews in our predictive model. Table 5 lists all variables associated with company characteristics, competitive landscape, and macrofactors. In Online Appendix A, we depict survival patterns by cuisine types and states. This online appendix also provides model-free evidence that supports our conjecture that consumer-posted photos may serve as a leading indicator of restaurant survival.

2.1. Photos

The photo information collected from Yelp includes each photo's upload date, helpful votes received, and the photo caption written by the poster as well as its

Figure 1. Research Road Map



general content (i.e., food, drink, interior, outside, menu) as classified by Yelp. Because we only observe the posting date of each photo and the total number of helpful votes received by a photo at the end of our data collection, we use the number of helpful votes normalized by the number of years since upload, that is, yearly helpful votes, to measure the helpfulness of each photo.[10] We describe how we extract additional information for photo content, photographic attributes, and photo caption as follows.

2.1.1. Photo Content. Yelp has developed a deep learning model that yields a general photo classification as follows: food, drink, interior, outside, and menu.[11] Because this general classification may not fully capture the detailed content of a photo (e.g., strawberries, seafood, ocean view), we further extract specific content in each photo using the Clarifai API. Several marketing researchers utilize APIs to identify objects in photos (Ko and Bowman 2020, Li et al. 2019). We chose Clarifai because it is a market leader in photo-content detection. Clarifai was among the top five winners in photo classification at the ImageNet 2013 competition and is particularly useful for our study because it provides highly detailed labels for food. Clarifai's "Food" model recognizes more than 1,000 food items in photos down to the ingredient level.[12] Online Appendix B provides a more detailed description of how we used Clarifai API in our research. The Clarifai API generates 5,080 unique labels for all photos in our data set.

To summarize the content in the photos detected by Clarifai, we calibrated a topic model called latent Dirichlet allocation (LDA) (Blei et al. 2003). The topic model serves as a data-reduction technique, allowing us to use our prediction model to relate LDA

Table 2. Summary Statistics of Photos and Reviews

	Count	Mean	Standard deviation	Minimum	Maximum
General content of photos					
Food	755,758	59.10%			
Interior	755,758	6.09%			
Outside	755,758	5.14%			
Drink	755,758	3.03%			
Menu	755,758	2.30%			
Others	755,758	24.34%			
Length of captions (number of characters)	533,965	31.94	26.13	1	140
Yearly helpful votes for photos[a]	755,758	0.16	0.34	0	38
Star rating of reviews	1,121,069	3.75	1.30	1	5
Length of reviews (number of of characters)	1,121,069	617.45	585.41	1	5,000
Yearly helpful votes for reviews[a]	1,121,069	0.56	1.52	0	133

[a]Yearly helpful votes = $\frac{\text{total \# of helpful votes received at the end of our observation window}}{\text{years since upload}}$.

Table 3. Restaurant Level Summary Statistics

	Count	Mean	Standard deviation	Minimum	Maximum
Age (at the end of year 2015)	17,719	19.89	21.10	1	192
Chain	17,719	25.95%			
Price level	16,932	1.56	0.61	1	4
Top 10 cuisine types					
Mexican	17,719	12.34%			
American (Traditional)	17,719	12.04%			
Pizza	17,719	11.89%			
Nightlife	17,719	10.94%			
Fast Food	17,719	10.80%			
Sandwiches	17,719	9.76%			
American (New)	17,719	8.10%			
Burgers	17,719	7.36%			
Italian	17,719	7.18%			
Chinese	17,719	6.77%			
Other cuisine types	17,719	2.82%			
Total number of photos[a]	17,719	42.65	122.55	0	4,793
Total number of reviews[a]	17,719	63.27	135.76	0	5,040

[a]Total number of photos/reviews is cumulative volume at the end of our observation window.

probabilities of photo content topics to survival probabilities. We vary the number of topics between 2 and 40. We find that the fitted LDA model with 10 topics yields the highest topic coherence score (Röder et al. 2015), a method that is shown to make the resulting topics more interpretable (Chang et al. 2009, Newman et al. 2010). Therefore, we set the number of photo topics to 10 in our study. Online Table A2 in Online Appendix B presents the 10 topics and the most representative words for each topic. Please refer to Online Appendix B for technical details of our topic-modeling procedure.

Given that the variety of objects in a photo might also affect its memorability (Isola et al. 2011) and consumer appetite (Wadhera and Capaldi-Phillips 2014), we follow prior literature on visual complexity (Isola et al. 2011) by counting the number of unique labels in a photo. One challenge we face is that there is no clear structure in the returned 5,080 labels (e.g., returned labels might include both "berry" and "strawberry," but "berry" is a superset of "strawberry"). To solve this problem, we use WordNet (Miller 1995, Fellbaum 1998) to structure the labels. For example, "fruit" is a superset of "berry," which, in turn, is a superset of "strawberry." Thus, in this example, "strawberry" is the leaf node, and we use leaf nodes (the label at the most refined level) among the labels of a photo to measure the variety of objects in the photo. The total number of unique leaf nodes is 5,037.

2.1.2. Photographic Attributes. We further examine whether photographic attributes carry any predictive power for restaurant survival. As suggested by Zhang et al. (2018), photographic attributes may reflect the quality of a photo, which, in turn, may affect demand. Following Zhang et al. (2018), we organize photographic attributes into three categories: (1) color, (2) composition,

Figure 2. (Color online) Distribution of Life Span of Open vs. Closed Restaurants

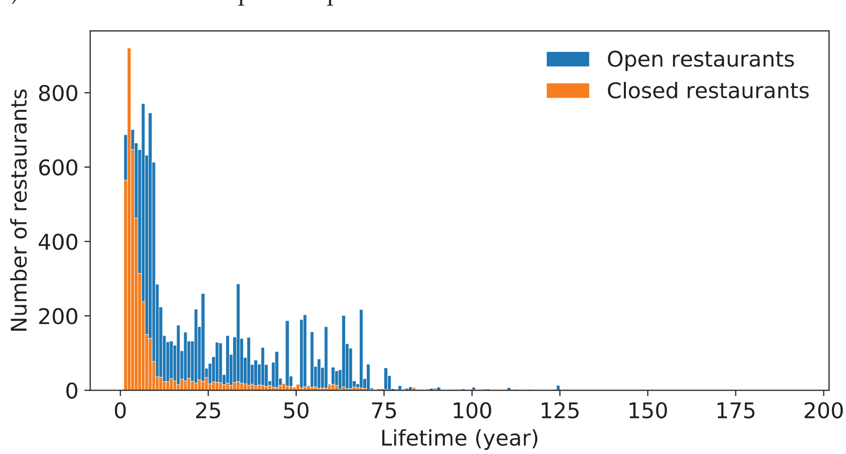


Table 4. Definitions of Independent Variables: Photos and Reviews

Type	Subtype	Variable (Both one-period lag and cumulative until $t-1$)
Photo	Photo content$_{it-1}$	Prop. of photos depicting each of the general content types (food, drink, interior, outside, and menu) provided by Yelp
		Prop. of photos on each of the 10 LDA topics
		Avg. photo variety based on Clarifai labels
		Std. of photo variety based on Clarifai labels
		Prop. of photos equipped with top 33.3% variety in photo content[a]
		Prop. of photos equipped with middle 33.3%variety in photo content
		Prop. of photos equipped with bottom 33.3% variety in photo content
	Photographic attributes$_{it-1}$	Avg. of each attribute (18 attributes of color, composition, figure–ground relationship)
		Std. of each attribute
		Prop. of photos rated equipped with top 33.3% of each attribute
		Prop. of photos rated equipped with middle 33.3% of each attribute
		Prop. of photos rated equipped with bottom 33.3% of each attribute
	Caption$_{it-1}$	Prop. of photos with a caption
		Avg. caption length (number of characters)
		Std. of caption length
		Prop. of photos with a dish name caption
		Prop. of photos with a positive caption
		Prop. of photos with a negative caption
		Avg. sentiment of captions (if caption is not a dish name)
		Std. of caption sentiment (if caption is not a dish name)
	Photo volume$_{it-1}$	Number of photos
	Helpful vote$_{it-1}$	Prop. of photos with helpful votes[b]
		Avg. yearly helpful votes for photos
		Std. of yearly helpful votes for photos
Review	Restaurant quality dimensions$_{it-1}$	Prop. of reviews mentioning each restaurant quality dimension (food, service, environment, price)
		Avg. sentiment of each dimension
		Std. of sentiments of each dimension
		Prop. of reviews equipped with top 33.3% sentiment of each dimension
		Prop. of reviews equipped with middle 33.3% sentiment of each dimension
		Prop. of reviews equipped with bottom 33.3% sentiment of each dimension
	Review content$_{it-1}$	Prop. of reviews on each of the 20 LDA topics
		Avg. review variety based on nouns
		Std. of review variety based on nouns
		Prop. of reviews equipped with top 33.3% variety in review content
		Prop. of reviews equipped with middle 33.3%variety in review content
		Prop. of reviews equipped with bottom 33.3% variety in review content
	Review volume$_{it-1}$	Number of reviews
	Star rating$_{it-1}$	Avg. star rating
		Std. of star ratings
		Prop. of reviews of each star
	Review length$_{it-1}$	Avg. review length (number of characters)
		Std. of review length
	Helpful vote$_{it-1}$	Prop. of reviews with helpful votes
		Avg. yearly helpful votes for reviews
		Std. of yearly helpful votes for reviews

[a]For example, "prop. of photos equipped with top 33.3% variety in photo content" is the prop. of photos from the focal restaurant with the top one third variety in photo content among all photos in our data set.

[b]For example, to calculate percentage of photos with helpful votes till year 2006, suppose that the focal restaurant has a total of 20 photos by the end of 2006, and 10 out of the 20 photos have received helpful votes. Then, the percentage of photos with helpful votes till 2006 is calculated to be 50%.

and (3) figure–ground relationship (see definitions in Online Table A3) based on the prior photography and marketing literature. For example, researchers show that color can affect the aesthetics (Arnheim 1965, Freeman 2007) and attractiveness of food (Nishiyama et al. 2011, Wadhera and Capaldi-Phillips 2014, Takahashi et al. 2016). Composition describes how different subjects and visual elements are arranged within the photo (Krages 2005). Visually balanced photos give the audience the feeling of order and tidiness and minimize cognitive demands (Kreitler and Kreitler 1972). The figure–ground relationship reflects how the central element of the photo

Table 5. Definitions of Independent Variables: Company, Competition, and Macro

Type	Variable	Definitions of variable
Company	Age_{it-1}	Years since the restaurant opened. We include age as a set of dummies (age = 0, 1, 2, ..., 21, 21+) in our prediction model.
	Chain_{i}	Whether the number of restaurants with the same name was greater than five in our data
	Price level_{i}	$, $$, $$$, $$$$
	Cuisine types_{i}	Cuisine types as defined in Table 3
Competition	Number of overlapping $\text{competitors}_{it-1}$	Number of competitors with overlapping cuisine types in the same zip code in period $t-1$
	Number of nonoverlapping $\text{competitors}_{it-1}$	Number of competitors with no overlapping categories in the same zip code in period $t-1$
	Number of new entries_{it-1}	Number of competitors opened in the same zip code in period $t-1$
	Number of new exits_{it-1}	Number of competitors closed in the same zip code in period $t-1$
	Avg. of number of photos_{it-1} per competitor	Avg. of photo volume of competitors in the same zip code (both one-period lag and cumulative until $t-1$)
	Avg. of number of reviews_{it-1} per competitor	Avg. of review volume of competitors in the same zip code (both one-period lag and cumulative until $t-1$)
	Avg. of competitors' avg. star rating_{it-1}	Avg. of avg. star rating across competitors in the same zip code (both one-period lag and cumulative until $t-1$)
Macro	Year_{t-1}	Year dummies
	Location: zip code_{i}	Zip codes with more than 100 restaurants included

is distinguished from the background. Consumer research suggests that ads containing photos with clear figure–ground relationships receive more attention from their audience (Larsen et al. 2004).

In sum, we use an 18-dimensional vector to capture the color, composition, and figure–ground relationship attributes of each photo with each dimension bounded between zero and one. A higher score represents a higher intensity in that dimension. For example, a photo with a 0.9 brightness level is much brighter than one measuring 0.1. Online Appendix B includes more technical details of how we extract these photographic attributes.

2.1.3. Photo Caption. Seventy-one percent of photos in our data set have a caption (i.e., a short description of the photo provided by the poster), generally reflecting either a dish name, such as "strawberry cheesecake," or positive/negative sentiment associated with the photo, such as "giant pretzel is yummy!" or "the tofu wasn't that great."

We use Valence Aware Dictionary and sEntiment Reasoner (VADER) (Hutto and Gilbert 2014) sentiment analysis to analyze photo captions. VADER is a lexicon and rule-based sentiment-analysis tool specifically attuned to sentiments expressed in social media platforms, such as Yelp. VADER is calibrated on multiple sources: a sentiment lexicon built on well-established sentiment word banks (Linguistic Inquiry and Word Count, Affective Norms for English Words, the General Inquirer) and validated by independent human judges, tweets, *The New York Times*, movie reviews from Rotten Tomatoes, and Amazon reviews.

Our analysis generates a sentiment score for each caption, ranging from −1 (most negative) to +1 (most positive). Consistent with our expectation, dish name captions always have a neutral sentiment (sentiment score = 0). We also find that 70% of the captions are neutral dish names, 26% are positive, and only 4% are negative. Please see Online Appendix B for examples and distribution of extracted sentiments from photo captions.

2.2. Reviews

Given that both photos and reviews are user-generated content that reflects consumer experience with the focal restaurant, reviews serve as an important control for us to examine the incremental predictive power of photos. Because prior literature (e.g., Chevalier and Mayzlin 2006, Liu 2006) suggests that review volume, star rating, and length can impact demand, we include these variables in our analysis. To the extent feasible, we also make the analyses of photos and reviews as comparable as possible. For example, similar to helpful votes of photos, we use yearly helpful votes of a review to measure its helpfulness. In the following, we explain how we extract restaurant quality dimensions and content information from reviews.

2.2.1. Restaurant Quality Dimensions. Restaurant quality is unobservable to us, but it is a crucial factor for restaurant survival. Based on prior literature on restaurant quality dimensions (Hyun 2010, Ryu et al. 2012, Bujisic et al. 2014), we extract the following four quality dimensions from reviews: food, service, environment, and price. Compared with the time-invariant price level given by Yelp, the quality dimension measure of price based on reviews can track consumer sentiment regarding prices over time.

One challenge of evaluating restaurant quality dimensions based on reviews is that Yelp reviews do not provide separate numerical ratings for each quality dimension. For each review in our data set, we had to determine whether each quality dimension is mentioned and, if so, the corresponding sentiment. Because we have more than a million reviews in our data set, it is challenging to manually label all of them. Therefore, we randomly selected 10,000 reviews from our review data set and recruited 4,051 consumers from Amazon Mechanical Turk (Mturk) to each read and provide labels for 20 reviews, which provides an average of eight labels per review. Please see the details of our Mturk survey in Online Appendix C.

In line with the stream of literature that utilizes deep learning to extract managerially relevant information from reviews (e.g., Lee et al. 2019, Liu et al. 2019, Timoshenko and Hauser 2019), we use a text-based multitask convolutional neural network (CNN) to extract restaurant quality dimensions from reviews. The 10,000 reviews and their labels are used to calibrate the text-based CNN. We randomly split the 10,000 reviews into 80% for calibration and 20% for out-of-sample testing. In the calibration process, the text of each review is treated as model inputs, and the outputs are the eight quality dimension scores labeled by the Mturk survey. We use the area under the receiver operating characteristic (ROC) curve (AUC) (Hanley and McNeil 1982) to measure prediction accuracy in the holdout test data set. AUC is proven to be useful in evaluating prediction accuracy in many machine learning applications (Bradley 1997; Fawcett 2004, 2006; Netzer et al. 2019). AUC ranges between zero and one. The larger the AUC, the better. A random model generates an AUC = 0.5. The text-based CNN yields AUC scores greater than 0.88 for all eight tasks, as shown in Online Table A7. We then extrapolate the calibrated CNN to each review in our data set to extract the quality dimension scores, an eight-dimensional vector of scores bounded between zero and one. A higher score represents a higher probability that the review mentioned the corresponding quality dimension (for the first four scores in the vector) or better sentiment (for the second four scores in the vector) in the corresponding dimension. More technical details of this text-based CNN are provided in Online Appendix C.

2.2.2. Review Content. Although the primary purpose of extracting restaurant quality dimensions is to control for the valence of each dimension, these four dimensions may not fully capture the detailed content of a review (e.g., dessert served, waiting time, happy hour). Additionally, whereas certain specific content (e.g., dessert) may exist in both photos and review texts, there may be rich information contained in review content (e.g., dissatisfaction for waiting time, love for a restaurant) that are not reflected in photos. Therefore, we further extract specific content information from reviews using topic modeling.

Similar to our topic modeling on photo content, we calibrated an LDA model to analyze the review content. We vary the number of topics between 2 and 40 and find that the fitted LDA model with 20 topics yields the highest topic coherence score (Röder et al. 2015). Online Table A9 presents the 20 topics and the most representative words for each topic. Please refer to Online Appendix C for technical details of our topic modeling procedure for reviews.

Moreover, because the variety of objects in photos might also be discussed in reviews, we capture the variety of objects in reviews as an additional control. Because nouns are natural units of objects in reviews, we count the number of unique nouns in a review. We use WordNet (Miller 1995, Fellbaum 1998) to structure the nouns (e.g., "strawberry" is a leaf node of "berry"). Then, we use leaf nodes among the nouns of a review to measure the variety of objects in the review. The total number of unique leaf nodes is 30,296.

2.3. Restaurant Characteristics

Based on the extant business survival literature (e.g., Audretsch and Mahmood 1995, Bates 1995, Parsa et al. 2005), we include several restaurant characteristics, including age, chain status, cuisine type, and price level, as additional controls in our survival model. Prior studies (e.g., Carroll 1983) often suggest that less-established businesses are more prone to failure than their longer established counterparts. Because not all restaurants include the opening year information on Yelp, we took the following steps to collect data on age for restaurants in our sample. First, we wrote a Python program to check and collect birth year information from each restaurant's Yelp page. Second, for restaurants with no birth year listed on Yelp, we collected the URL link of its Facebook page from a restaurant's website (if listed on Yelp). Third, we then collected birth year information from the respective Facebook page. Finally, for the restaurants that do not report birth years on Yelp or Facebook, two research assistants manually checked birth years on restaurants' websites (if available) or through the Google search engine. In total, we collected birth year information for 10,368 restaurants, accounting for 59% of all restaurants in our sample. For the remaining 41% of restaurants, we used the date of the first photo or first review as a proxy for their birthdates.

We further conducted two robustness checks regarding the age measure. First, we predicted restaurant survival using only the 10,368 restaurants with accurate age information (Online Appendix D, Table A18). Second, we account for the fact that the first photo/review on Yelp is truncated in year 2004. As such, when birthdate = 2004 is inferred from the year of the first photo/review on Yelp, such age information might be less accurate. Hence, we added a dummy (one when birthyear = 2004 based on UGC, zero otherwise) as an additional variable in the model as a robustness check (Online Appendix D, Table A26).[13] The results of both robustness checks are consistent with those in Table 6.

Following Parsa et al. (2011), chain status is determined by whether the number of restaurants with the same name in our data set is greater than five. Cuisine types are based on the cuisine type labels provided by Yelp (e.g., fast food, bars, pizza). Yelp categorizes restaurant cuisines into 260 types. A restaurant may have more than one cuisine type (e.g., McDonald's belongs to both fast food and burger categories). Because many cuisine types include only a few restaurants in our sample, we examine the survival probabilities of the top 10 cuisine types and group the remaining types as "others." The price level is measured by one to four dollar signs as indicated on Yelp. As shown in Table 3, 26% of restaurants in our sample are chain restaurants, and our restaurant sample includes a wide variety of restaurants with different cuisine types and price levels.

2.4. Competitive Landscape

According to the business survival literature (Kalleberg and Leicht 1991, Wagner 1994, Parsa et al. 2005, Fritsch et al. 2006), the intensity of competition plays a vital role in the survival of a business enterprise. Consequently, we take into account competitor concentration, new entries/exits in each period, and UGC of competitors in our survival model. Following Parsa et al. (2005), we consider all restaurants operating in the same zip code and year as competitors of the focal restaurant. Given that restaurants from the same cuisine types often compete for a similar customer base, we further group competitors into competitors with overlapping/nonoverlapping cuisine types. Overlapping means that the competing restaurant shares at least one cuisine type with the focal restaurant; nonoverlapping means that they have no cuisine type in common. A large number of competitors within the same cuisine type in an area can indicate the popularity of that cuisine type and/or fierce competition in that area (Fritsch et al. 2006). We also consider the number of new entries and exits (we discuss how we identify exit status and exit time in Section 2.6) in each year and zip code. Many exits may indicate decreased demand and/or less competition in that area, whereas many new entries may indicate increased demand and/or more competition (Fritsch et al. 2006). Finally, we account for the average volume of photos and reviews per competitor and the average star rating of competitors in each year and zip code because the UGC of competitors may also reflect the fierceness of competition.

2.5. Macroconditions

We account for macroconditions, such as time trends by year dummies and local environments by zip code dummies. The interactions between year dummies and zip code dummies capture local condition changes over the years. We do not need to include year–zip code–specific dummies because the models we use (the XGBoost algorithm in Section 3 and causal forests in Section 4) account for interactions between year and zip code dummies automatically.

2.6. Exit Time

To study restaurant survival, we need to know whether and/or when a restaurant went out of business. For each closed restaurant, Yelp has a salient banner on the restaurant page indicating that "Yelpers report this location has closed" (see Online Figure A1). During the 11 years in our observation window, 4,495 restaurants (25.37% of all restaurants in our sample) went out of business.

However, Yelp does not provide information regarding the exit time. For each closed restaurant, we scanned for the earliest review that mentions a restaurant's closure and use the date of that review to approximate its exit time. We use keyword matching to identify the earliest review mentioning the closure of a restaurant. To obtain a dictionary of keywords, we recruited a research assistant to read all reviews of 200 randomly chosen closed restaurants in order to identify words and phrases representing the permanent closure status of a restaurant. This dictionary of keywords was then used to identify exit time for all closed restaurants in our data set. Online Figure A1 shows an example of a review that mentions a restaurant closure. If a closed restaurant has no reviews mentioning its closure, we use the date of the last photo or last review to approximate its exit time.

As a robustness check, we collect reports about restaurant closure on Eater.com (a website with local culinary news that reports restaurant openings and closings). Eater.com only covered restaurant closure information for selected cities in our data set from 2012 to 2015. We use the information provided by the subset of closed restaurants that we could identify on Eater.com to cross-check restaurant exit time. We use the exact closure date if it is provided in the Eater.com report; if no exact date is reported, we use the date of

Table 6. One-Year-Ahead Prediction: Comparison Between Photos and Reviews

	Out of sample					In sample
	AUC	KL divergence	Sensitivity	Specificity	Balanced accuracy	Pseudo R^2
Baseline (i.e., no UGC)	0.7020	0.1973	0.6484	0.6284	0.6384	0.1373
	(0.0047)	(0.0035)	(0.0056)	(0.0092)	(0.0046)	(0.0026)
Baseline + review	**0.7152**	0.1955	**0.7301**[a]	0.5614	0.6457	**0.2100**
	(0.0048)	(0.0035)	(0.0054)	(0.0095)	(0.0042)	(0.0062)
Baseline + photo	**0.7612**[a]	**0.1877**[a]	**0.6735**	**0.6936**[a]	**0.6835**[a]	**0.2324**
	(0.0066)	(0.0027)	(0.0087)	(0.0086)	(0.0051)	(0.0029)
Baseline + review + photo	**0.7660**[a]	**0.1857**[a]	**0.7027**	**0.6820**[a]	**0.6924**[a]	**0.2673**[a]
	(0.0065)	(0.0028)	(0.0081)	(0.0089)	(0.0052)	(0.0046)
Total obs.				89,384		

Notes. Total obs. is the number of all observations (including training and testing sets) used to calculate the predictive performance. When an observation is used multiple times in yearly predictions, we only count it once. To see the number of observations used in each yearly prediction, please refer to Online Table A13. Baseline model includes restaurant characteristics, competitive landscape, and macroconditions. For sensitivity, specificity, and balanced accuracy, the training data are reweighted so that the total weights of open and closed observations are equal. Results are averaged over years and cross-validation iterations. Standard errors are provided in parentheses. Bold numbers indicate significant improvement over the baseline model at the 0.05 level with a two-sided test.

[a]Best in the column or not significantly different from best in the column at the 0.05 level with a two-sided test.

the report itself as the exit time. The average discrepancy between the closure date identified via Yelp reviews and that identified by Eater.com is 60.95 days. Because we use a year as the unit of analysis, the difference is negligible.

3. Predicting Restaurant Survival

In Section 3.1, we describe a predictive model of restaurant survival based on the XGBoost algorithm (Chen and Guestrin 2016), a scalable implementation of gradient boosting trees (Friedman 2001) using all input variables shown in Tables 4 and 5. We first apply this model for a one-year-ahead prediction and explore what factors are the most informative in forecasting restaurant survival in Section 3.2. We then examine whether consumer-posted photos are more informative for certain types of restaurants in Section 3.3. In Section 3.4, we investigate how long photos can remain predictive in forecasting restaurant survival.

3.1. Restaurant Survival Model

Our restaurant survival model is based on gradient boosting trees. In particular, we apply the XGBoost algorithm (Chen and Guestrin 2016) because of its excellent performance in many predictive tasks. The superior predictive performance of this algorithm is widely recognized across many machine learning challenges held by Kaggle and KDD cups (Chen and Guestrin 2016). Several prior marketing studies also use XGBoost (e.g., Rafieian and Yoganarasimhan 2021, Rajaram et al. 2021) or similar gradient boosting trees methods (Yoganarasimhan 2019) for complex customer behavior prediction problems.

In our context, the XGBoost algorithm has the following four properties that are particularly desirable. First, XGBoost works well in handling a large number of predictor variables that may correlate with each other (Chen et al. 2018). This property is particularly helpful in our context because many of our predictors are correlated. Second, given that our predictive model includes a very high number of independent variables, XGBoost is also particularly suitable as it automatically selects the most informative variables for prediction (Chen and Guestrin 2016). Third, XGBoost's flexibility in handling potentially higher order interactions among predictors (Friedman 2001) is advantageous in our context (e.g., photos might be more indicative of the survival of an independent restaurant than for a chain restaurant). Finally, the XGBoost algorithm's ability to efficiently process sparse data (Chen and Guestrin 2016) is also helpful in our setting as our data contain many dummy variables (e.g., cuisine types, zip codes, years) as well as variables with missing values. To discern the predictive ability of the XGBoost algorithm in our specific context, we compare the predictive performance of XGBoost with random forests and support vector machine in Online Appendix D. The comparison results show that XGBoost outperforms the benchmark models.

Following Figure 1, we include consumer-posted photos and a set of control variables (i.e., consumer reviews, company, competition, and macrofactors) as inputs and restaurant survival as the outcome variable for XGBoost. Let i denote restaurant and t denote period. We define each period as a calendar year (for example, in year 2014, $t = 2014$) because yearly (compared with monthly) predictions provide a longer lead time for managers and investors to plan out their resource-allocation decisions. Furthermore, because some restaurants may have no newly posted photos or reviews within a given month, using year rather than month as the unit of analysis also reduces sparsity of data and improves the stability of our empirical model.

In what follows, we outline the basic intuition behind our predictive model. Following Chen and Guestrin (2016), the XGBoost algorithm in our predictive model aims to minimize the objective function outlined in Equation (1) consisting of two parts: a loss function $loss(\theta)$ (a negative log-likelihood) and a regularization term $\Omega(\theta)$. θ is the set of parameters to be calibrated through the model (i.e., the functions f_g, each containing the structure of the tree and the leaf scores ω_{gl}). In Equation (1), y_{it} is the observed binary dependent variable. $y_{it} = 1$ if restaurant i survives at period t. $y_{it} = 0$ if restaurant i is closed at period t. $\boldsymbol{X}_{it-1} = (\boldsymbol{x}_{it-1}, \boldsymbol{x}_i, \boldsymbol{x}_{t-1})$ is the vector of independent variables as shown in Tables 4 and 5 with $\boldsymbol{x}_{it-1}$ being the time-variant variables including both one-period and cumulative variables (e.g., photo volume in $t-1$ and cumulative photo volume until $t-1$), $\boldsymbol{x}_i$ representing time-invariant variables (e.g., cuisine types), and $\boldsymbol{x}_{t-1}$ being restaurant-invariant variables (e.g., year dummies). $\widehat{y_{it}}$ is the predicted survival probability between zero and one. Please refer to Online Appendix D for more technical details of this predictive model and how we fine-tune the hyperparameters (e.g., γ and λ).

$$\min_{\theta} \; loss(\theta) + \Omega(\theta)$$
$$loss(\theta) = -\sum_{it \in training\ sample} [y_{it} \ln \widehat{y_{it}} + (1 - y_{it}) \ln (1 - \widehat{y_{it}})]$$
$$\Omega(\theta) = \sum_{g=1}^{G} \left[\gamma L_g + \frac{1}{2} \lambda \sum_{l=1}^{L_g} \omega_{gl}^2 \right], \tag{1}$$

where $\widehat{y_{it}} = \frac{e^{\sum_{g=1}^{G} f_g(X_{it-1})}}{1 + e^{\sum_{g=1}^{G} f_g(X_{it-1})}}$; $f_g: \boldsymbol{X}_{it-1} \rightarrow \omega_{gl}$, $l = 1, 2, \ldots, L_g$; g indexes trees; l indexes leaves.

To ensure that our predictive model of restaurant survival can be generalized to restaurants not in our data set, we predict the survival of out-of-sample restaurants in out-of-sample time periods. Specifically, we shuffle the restaurants and split them into 10 even buckets for tenfold cross-validation. Then, we use data up to period $t-1$ of nine buckets of restaurants (i.e., the training sample is $\{(y_{is}, X_{is-1}) \mid i \in$ the nine buckets of restaurants, $s \le t-1\}$) to calibrate the model. We further test the model's predictive performance by applying the calibrated model to predict the survival of the remaining bucket of restaurants in period t (if they are open in period $t-1$). Namely, we plug X_{it-1} into the calibrated XGBoost to generate $\widehat{y_{it}}$ prediction for each open restaurant i as of period $t-1$ in the remaining bucket. For example, we calibrate the model using observations from 90% of all restaurants until year 2009. Among the remaining 10% of restaurants in the holdout sample, we first remove restaurants that did not survive until the end of year 2009. We then use the calibrated model to predict the survival of the remaining restaurants in 2010 (i.e., use $\boldsymbol{X}_{i2009}$ to predict y_{i2010}). As such, we carry out six predictive tasks to predict survival in years 2010–2015, respectively (e.g., using data up to 2009 to predict survival of out-of-sample restaurants in 2010, using data up to 2010 to predict survival of out-of-sample restaurants in 2011, etc.).[14] Given that we use tenfold cross-validation, we conduct 10 predictive tasks for each yearly prediction. Namely, there are a total of 60 predictions for each model specification. All tables in this section are based on average predictive results across years 2010–2015 and across the tenfold cross-validation.

As a robustness check, we also carry out an alternative forecasting scenario by using data up to period $t-1$ to predict the survival of all open restaurants in period t. For example, we calibrate the model using observations of all restaurants till year 2009. We then predict survival in year 2010 for all open restaurants in year 2009. The results (in Online Tables A22 and A24) are consistent with what we observe in Tables 6 and 7. Although this alternative forecasting scenario may appear to be an easier predictive task, we observe that the two prediction approaches have similar predictive performance. Our conjecture is that, because a restaurant's prior years have no variation in survival (i.e., survival in each year), prior survival of a restaurant contains little information about whether the same restaurant would continue to survive in a future year. We also provide yearly predictive results for both prediction approaches in Online Appendix D, Tables A13 and A23. The pattern is, by and large, consistent across years.[15]

Given our primary interest in predicting restaurant survival in the next period, we lag t for one period for time-variant independent variables. In our data, 1,723 restaurants have only one year of observation. These restaurants are dropped if we lag t for one period. Nevertheless, these observations can be helpful for us to better understand restaurant survival. First, in practice, an investor may want to predict whether a restaurant will survive in its first year. Second, having the volume of UGC being zero in the prior year might be an informative factor to predict survival. Finally, the non-UGC characteristics of these restaurants enable us to preserve additional variations among non-UGC-related variables in our data. Therefore, we decide not to drop these restaurants in our analysis. Instead, we include a period zero in which the UGC volume of all focal restaurants in the period prior to when they were listed on Yelp is set to zero.[16] In total, we have 89,384 restaurant–year observations.

We conduct two robustness checks for this procedure. First, we perform predictions by dropping period zero, which, in turn, removes the 1,723 restaurants from our restaurant sample. The results (Online Appendix D, Table A20) are qualitatively consistent with the results in Table 6. Second, although all restaurants don't have UGC in period zero, such zero UGC can have different meanings for new and existing restaurants. For example, restaurant A was founded in 2010 (restaurant A's period 1), and thus, it was not eligible to receive any UGC in 2009 (age = 0 in restaurant A's period 0). Differently, restaurant B was founded in 2008, but it received its first review in 2010 (restaurant B's period 1) because it did not receive any UGC in 2009 (age = 2 in restaurant B's period 0). In such cases, whereas both restaurants have zero UGC in period 0, the reasons are different. Hence, we conduct another robustness check by dropping observations with age = 0. In such cases, restaurant A's age = 0 in the period 0 observation is dropped because it cannot receive UGC in period 0. But restaurant B's age = 2 in the period 0 observation remains because zero UGC is meaningful to be accounted for in such cases. The results (Online Appendix D, Table A25 and Figure A13; Online Appendix E, Table A29) are also qualitatively similar to results in Table 6, Figure 4, and Table 9, respectively. More importantly, we would like to clarify that, because our focal model already accounts for age = 0 and age > 0 when UGC = 0, the main model is sufficient to address such differences. The second robustness check is for a pressure test only.

Given that the literature on restaurant survival (e.g., Carroll 1983, Kalleberg and Leicht 1991, Fritsch et al. 2006) shows that organization age is a critical factor for business survival, we include age as a set of dummies (age = 0, 1, 2, … , 21, 21+) to explicitly examine the potential nonlinear relationship between age and restaurant survival. Because consumers can post photos without posting reviews and vice versa, Yelp separates photos (on top) and reviews (on bottom) on each restaurant page. Consequently, we included photo and review variables as two separate batches of independent variables.

Because recent photos and reviews may carry more predictive power compared with earlier ones, we include both the one-period lag variables that capture photos and reviews posted in the last period (referred to as $OnePeriod_{t-1}$) and the cumulative variables that encompass all photos and reviews cumulated until the end of last period (referred to as Cum_{t-1}) whenever applicable. We include Cum_{t-1} because (1) cumulative values capture the entire history of a restaurant, and (2) in each period, consumers see summary statistics of cumulated photos and reviews on Yelp (e.g., the volume of photos and average star rating). Therefore, the results reported in the paper are based on a predictive model that includes both $OnePeriod_{t-1}$ and Cum_{t-1} variables.

As robustness checks, we also explore the following alternative specifications of one-period and cumulative variables: (1) $OnePeriod_{t-1} + Cum_{t-2}$ to avoid overlap of period $t-1$ in the model, (2) $OnePeriod_{t-1} + OnePeriod_{t-2} + Cum_{t-3}$ to capture both one- and two-year lags, and (3) $OnePeriod_{t-1} + Cum_{t-1} + Change_{t-1}$ to explicitly capture changes in UGC. Online Appendix D provides details on these alternative model specifications. Multiple-year lags are required in all alternative models described. As such, we are constrained to drop restaurants that survived for only a short time period (one year for specifications 1 and 3; one or two years for specification 2), the inclusion of which can be crucial in our understanding of restaurant survival. Online Tables A14 to A16 show that our main model specification performs similarly or slightly better compared with these alternative specifications, possibly because the main specification includes the highest number of observations along with relatively fewer predictors. Additionally, comparisons for the incremental predictive power of photos are consistent across these different model specifications, which also provides us with robustness checks for our Table 6 findings to be discussed.

3.2. One-Year-Ahead Survival Prediction: What is the Most, Somewhat, or Not at All Important

Given that both photos and reviews reflect consumer experience, they may overlap in their incremental predictive power over all non-UGC variables. Hence, in Table 6, we compare the predictive power of the following model specifications: (1) baseline (including all non-UGC variables related to company, competition, and macrofactors), (2) baseline + review (including all non-UGC variables and variables related to reviews only), (3) baseline + photo (including all non-UGC variables and variables related to photos only), and (4) baseline + review + photo.

We employ the following metrics to gauge the performance of these models: (1) AUC, (2) Kullback–Leibler (KL) divergence, (3) sensitivity, (4) specificity, (5) balanced accuracy, and (6) pseudo R^2. Our predicted $\widehat{y_{it}}$ is a probability bounded between zero and one, which is used to calculate AUC, KL divergence, and pseudo R^2. Because metrics (3)–(5) can only be calculated using a predicted binary outcome, we converted the survival probability to a dummy variable using 0.5 as the cutoff. We use metrics (1)–(5) for out-of-sample validation and metric (6) for in-sample fit measurement. KL divergence (Kullback and Leibler 1951) is an information theory measure of divergence between predicted distribution and observed distribution, which is widely applied in the marketing literature (e.g., Dzyabura and Hauser 2011, Hauser et al.

Table 7. One-Year-Ahead Prediction: Different Aspects of Photos

	Out of sample					In sample
	AUC	KL divergence	Sensitivity	Specificity	Balanced Accuracy	Pseudo R^2
Baseline	0.7020	0.1973	0.6484	0.6284	0.6384	0.1373
	(0.0041)	(0.0030)	(0.0049)	(0.0080)	(0.0046)	(0.0023)
Baseline + photographic attributes	0.7005	0.1975	**0.6861**[a]	0.5838	0.6350	**0.1767**
	(0.0048)	(0.0035)	(0.0055)	(0.0093)	(0.0042)	(0.0043)
Baseline + photo caption	0.7010	0.1972	**0.6646**	0.6060	0.6353	**0.1460**
	(0.0047)	(0.0035)	(0.0046)	(0.0096)	(0.0048)	(0.0027)
Baseline + photo volume	0.7051	0.1967	0.6581	0.6222	0.6401	0.1407
	(0.0048)	(0.0035)	(0.0048)	(0.0086)	(0.0043)	(0.0027)
Baseline + helpful votes	**0.7157**	0.1962	0.6601	0.6339	0.6470	**0.1670**
	(0.0046)	(0.0034)	(0.0098)	(0.0115)	(0.0038)	(0.0034)
Baseline + photo content	**0.7539**[a]	**0.1876**[a]	**0.6698**	**0.6884**[a]	**0.6791**[a]	**0.2014**[a]
	(0.0075)	(0.0028)	(0.0061)	(0.0094)	(0.0065)	(0.0020)
Total obs.				89,384		

Notes. Total obs. is the number of all observations (including training and testing sets) used to calculate the predictive performance. Baseline model includes restaurant characteristics, competitive landscape, and macroconditions. For sensitivity, specificity, and balanced accuracy, the training data are reweighted so that the total weights of surviving and closed observations are equal. Results are averaged over years and cross-validation iterations. Standard errors are provided in parentheses. Bold numbers indicate significant improvement over the baseline model at the 0.05 level with a two-sided test.

[a]Best in the column or not significantly different from best in the column at the 0.05 level with a two-sided test.

2014, Huang and Luo 2016). A smaller KL divergence value represents better accuracy. In our context, sensitivity is the hit rate among open restaurant–year observations, and specificity is the hit rate among closed restaurant–year observations. Because sensitivity focuses on open observations and specificity emphasizes closed observations, the training data are reweighted so that the total weights of open and closed observations are equal (Seiffert et al. 2008). The balanced accuracy is the arithmetic average of sensitivity and specificity (Brodersen et al. 2010, Bekkar et al. 2013). For AUC and KL divergence, the training data do not need to be reweighted because they work well for imbalanced samples (Kotsiantis et al. 2006, Dzyabura and Hauser 2011). Pseudo R^2 (bounded between zero and one) measures the percentage improvement in log-likelihood over the null model.

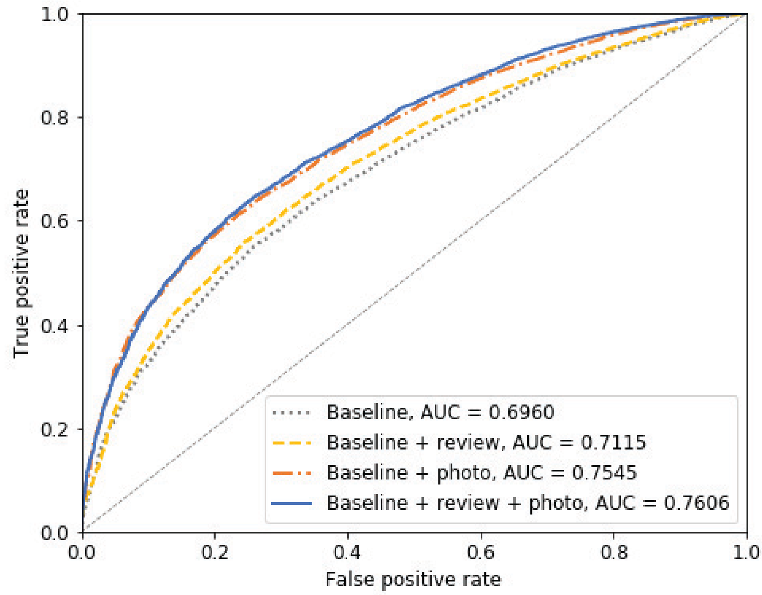
To statistically test the incremental predictive power across the four different model specifications, we use tenfold cross-validation to predict survival in each year and report the average performance of these models in Table 6. Specifically, we shuffle the restaurants and divide them into 10 even buckets. Then, we use data till year $t-1$ of nine buckets (i.e., 90% of restaurants) to train the XGBoost algorithm and calculate in-sample pseudo R^2 for measuring in-sample fit. In order to test the out-of-sample predictive performance of these models, among all open restaurants as of year $t-1$ in the holdout bucket (i.e., 10% of restaurants), we predict their survival in year $t-1$. In particular, we calculate metrics (1)–(5) as indicated in Table 6. We then calculate the mean and the standard deviation of model performances across years ($t = 2010, 2011, \ldots, 2015$) and across 10 cross-validation iterations.

Table 6 suggests that photos are strong predictors of restaurant survival. In particular, the "baseline + photo" model exhibits significantly greater performance compared with the baseline according to all metrics. Additionally, the "baseline + photo" model also performs better than the "baseline + review" model on most performance metrics with the exception of one metric (sensitivity, i.e., the hit rate among open restaurant–year observations).[17] In our particular context, there are many fewer closed observations than open observations. Therefore, predicting restaurant closure (i.e., specificity) is a considerably harder task. The specificity measure shows that the "baseline + photo" model does a much better job than the "baseline + review" model. Moreover, the "baseline + review + photo" model does not improve much from the "baseline + photo" model, indicating that prediction improvement from UGC mainly stems from photos. Such comparisons are further confirmed by the ROC curves in Figure 3.

By and large, photos appear to be a leading indicator of restaurant survival above and beyond reviews and other known factors. Recall that we summarize the content of photos and reviews by topic modeling in our main model specification as described. In Online Appendix D, we use the top 100 Clarifai labels to capture photo content and the top 100 nouns to capture review content. The results are qualitatively the same.

We further explore what aspects of photos are the most informative in predicting survival. Specifically, we compare the incremental predictive power of various aspects of photos as defined in Table 4 (photographic attributes, caption, volume, helpful votes, and content).

Figure 3. (Color online) Performance AUC Visualization for One-Year-Ahead Prediction: Photos Carry More Predictive Power Than Reviews



Notes. Because we cannot directly average ROC curves, each ROC curve in this figure is plotted by pooling together testing observations across years and cross-validation iterations. In contrast, we report the average AUC values across years and cross-validation iterations in Table 6. Hence, the AUC numbers in this figure are slightly different from those in Table 6.

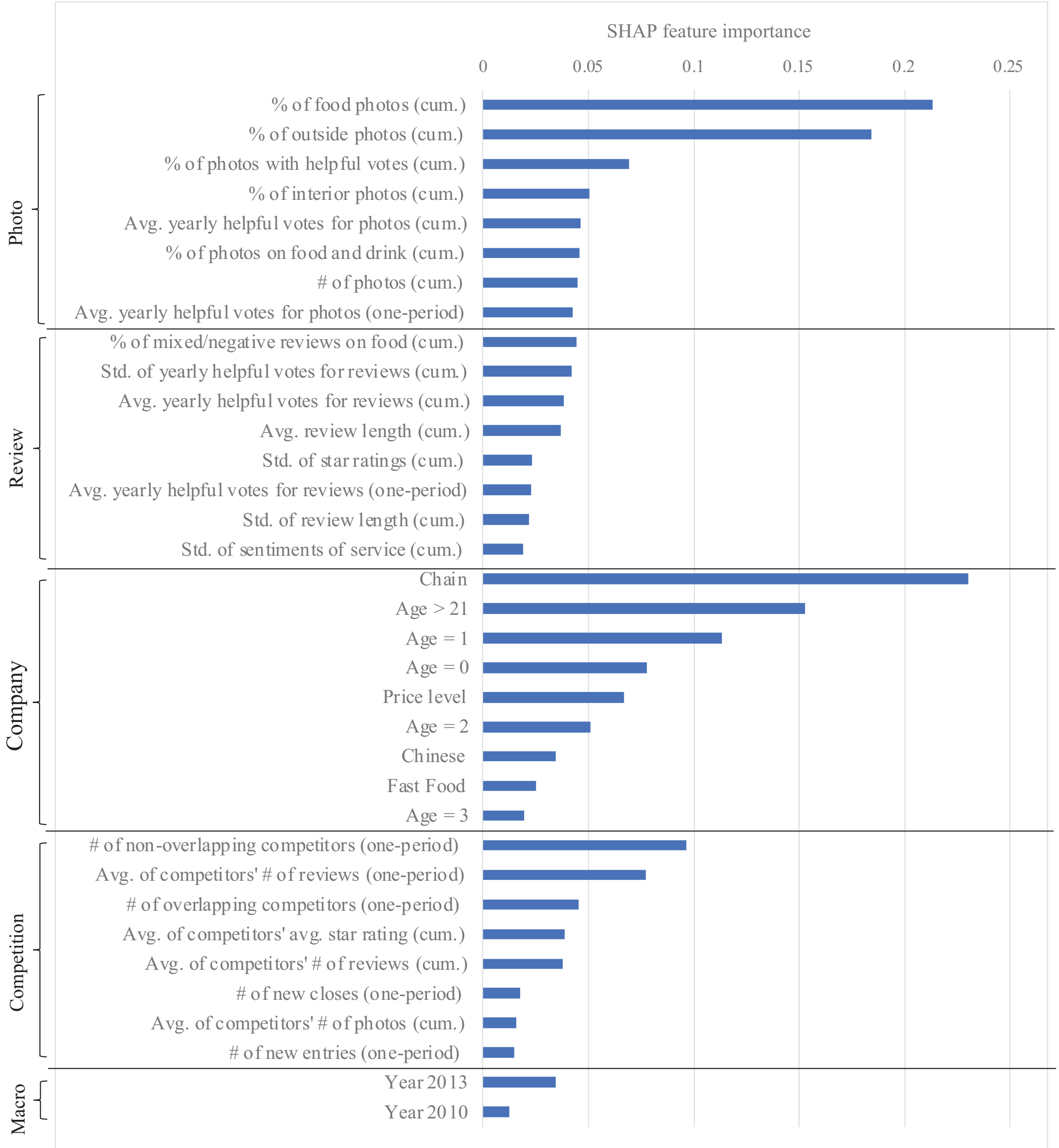
Table 7 shows that the predictive power of photos mainly stems from their content, followed by helpful votes. Although photo volume increases prediction accuracy above the baseline based on most performance metrics, such improvements are not statistically significant. Interestingly, photo captions and photographic attributes do not improve prediction accuracy over the baseline model according to most metrics with the exception that they significantly increase the sensitivity (i.e., the hit rate among open restaurant–year observations) over the baseline model. Again, there are many more open observations than closed observations in our context, and thus, predicting opening is a relatively easier task. Overall, our findings suggest that photo content and helpfulness play a more prominent role than do photo volume or photo aesthetics.

Linking these findings to our earlier conjecture as to why consumer-posted photos may predict restaurant survival, if the posters' perspective were at play, photo volume would play an important role in predicting survival. Meanwhile, if the viewer's perspective prevails by facilitating a better horizontal match between restaurants and photo viewers, photo content provided by photo viewers would be the most informative predictors of survival. When it comes to helpful votes of photos, on one hand, the helpfulness of photos may reveal a better horizontal match between the focal restaurant and viewers' private tastes. On the other hand, as popular restaurants may attract more helpful votes, helpful votes for photos may also be a proxy of potential demand. Given that we found photo content plays the biggest role in the preceding comparison, whereas both perspectives might be at play, the viewer's perspective may play a more salient role in explaining why consumer-posted photos may help predict restaurant survival.

We further examine the top 35 predictors with the highest SHAP feature importance weights (bounded between zero and one) (Lundberg and Lee 2017, Lundberg et al. 2020) for the "baseline + review + photo" model. We choose this model because it contains all factors related to restaurant survival, and thus, we can compare the importance weights of all variables. Essentially, the SHAP feature importance is the marginal contribution of one predictor considering all possible combinations of other predictors (Molnar 2021). So it takes into account the correlations across predictors. Figure 4 shows the top 35 variables and their SHAP feature importance.[18]

First and foremost, we find that restaurant age is among the most significant predictors of restaurant survival. Among the top 35 most informative predictors, we observe five age dummy variables (i.e., age > 21, age = 1, age = 0, age = 2, age = 3, in descending order of SHAP feature importance). Such a finding provides some external validity of our findings because prior research (e.g., Carroll 1983, Kalleberg and Leicht 1991, Fritsch et al. 2006) has long shown that organizational age highly correlates with survival. We conjecture that restaurants with ages $\leq$ 3

Figure 4. (Color online) Top 35 Predictors of Restaurant Survival in One-Year-Ahead Prediction



Notes. Importance weights are based on predicting survival in 2015. Within each type of factor (e.g., photo), variables are ordered by their predictive power.

have smaller chances to survive and that restaurants with ages > 21 are more likely to survive. We formally analyze the relationship between age and survival in Section 4.

Furthermore, we observe that 8 out of the top 35 variables relate to photos. Specifically, the three most informative predictors related to photos are (1) the proportion of photos depicting food, (2) the proportion of photos depicting outside, and (3) the proportion of photos with helpful votes. The results are consistent with prior literature suggesting food is the most critical aspect of a restaurant (Sulek and Hensley 2004, Duarte Alonso et al. 2013) and outdoor attractions are a vital factor for tourism regions (Getz and Brown 2006). Certain other photo content (i.e., the proportion of photos depicting the interior, LDA topic on food and drink based on Clarifai labels) and the total volume of photos are also among the top 35 variables. Interestingly, photo captions and photographic attributes are not in the top 35 predictors, indicating again that content may be more informative for restaurant survival than caption or photographic attributes.

With regard to control variables, we learn that mixed/negative reviews on food (a review content topic), helpful votes received by reviews, review length, star rating, and consumer sentiment of service also appear among the top 35 most informative predictors for restaurant survival. Our findings regarding review valence and length are, by and large, consistent with prior research linking consumer reviews with demand or stock performance (e.g., Chevalier and Mayzlin 2006, Tirunillai and Tellis 2012). Regarding restaurant characteristics, consistent with prior literature on business survival (e.g., Bates 1990, Audretsch and Mahmood 1995, Kalnins and Mayer 2004, Parsa et al. 2005), we observe that chain status, price level, and cuisine type are all predictive of restaurant survival besides age. Finally, 8 of the top 35 predictors relate to competition, including the number of overlapping/nonoverlapping competitors, new entries/exits, review volume, star rating, and photo volume of competitors, suggesting the importance of competitive landscape in restaurant survival.

3.3. For Which Types of Restaurants Are Photos More Informative in Predicting Survival?

We further explore whether the incremental predictive power of photos differs for heterogeneous restaurants. Specifically, we focus on the following three dimensions of restaurant heterogeneity: (1) chain versus independent restaurants, (2) restaurants of different ages, and (3) restaurants of varying price levels. We focus on these dimensions and conjecture that their survival may be differently associated with photos because prior literature on business survival suggests that these restaurants differ in their survival rates (Carroll 1983, Audretsch and Mahmood 1995, Lafontaine et al. 2018).

Table 8 provides the results of such comparisons. To fully calibrate the model for each type of restaurant, we retrain our predictive model separately using the subset of data associated with each restaurant type. The results shown here are based on AUC. We also examine other metrics, and results are qualitatively consistent. We learn that, whereas photos significantly increase prediction accuracy for all subtypes of restaurants, photos carry more incremental predictive power for independent (versus chain), young or midaged (versus established), and medium-priced (versus low-priced) restaurants.

Photos are more predictive for the survival of independent restaurants compared with chain restaurants possibly because consumers are, in general, less uncertain about chain restaurants because of their brand reputation and franchise regulations (Kalnins and Mayer 2004, Lafontaine et al. 2018). For ease of exposition, we group restaurants by 33.3% and 66.7% percentiles of age in the second comparison, that is, young ($1 \leq$ age ≤ 3), midaged ($3 <$ age ≤ 21), and established (age > 21).[19] Table 8 demonstrates that photos carry more predictive power for young and midaged restaurants compared with established restaurants possibly because the latter are well-known and rely less on consumer-posted photos. Given that restaurants with three or four dollar signs on Yelp represent only 5% of our data, we include only medium-priced (price level = 2) and low-priced (price level = 1) restaurants in the third comparison to be conservative. Table 8 suggests that photos carry more predictive power for medium- than for low-priced restaurants. It is possible that consumers may care more about their dining experiences when paying more, and photos might be more useful in facilitating the horizontal match in this case.

3.4. Multiple-Year Survival Prediction: How Long Can Photos Stay Predictive in Forecasting Restaurant Survival?

We now explore the ability of photos to predict multiple-year restaurant survival. This question is particularly important because it can be challenging to foresee survival beyond one year, and a longer term forecast for survival can be useful for restaurants' resource allocation and strategic planning decisions.

We aim to predict restaurant survival during the future Δt years, assuming that restaurant owners do not observe future photos and reviews (e.g., using

Table 8. Comparison of Incremental Predictive Power of Photos by Restaurant Type (AUC)

	Chain	Independent	$1 \leq$ age ≤ 3	$3 <$ age ≤ 21	Age > 21	Price level = 1	Price level = 2
Baseline + review	0.6700	0.6698	0.6290	0.6261	0.6580	0.7142	0.6960
	(0.0117)	(0.0054)	(0.0084)	(0.0108)	(0.0126)	(0.0059)	(0.0053)
Baseline + review + photo	0.7106	0.7323	0.7128	0.7320	0.7080	0.7582	0.7624
	(0.0157)	(0.0075)	(0.0093)	(0.0085)	(0.0132)	(0.0064)	(0.0075)
AUC increase by photo	**0.0406**	**0.0626**[b]	**0.0837**[a]	**0.1059**[a]	**0.0501**	**0.0440**	**0.0664**[a]
	(0.0119)	(0.0059)	(0.0084)	(0.0111)	(0.0103)	(0.0053)	(0.0060)
Percentage increase by photo	6.06	9.34	13.31	16.91	7.61	6.16	9.54
Total obs.	25,320	64,064	20,654	30,260	29,476	42,013	40,153

Notes. Total obs. is the number of all observations (including training and testing sets) used to calculate the predictive performance. Baseline includes restaurant characteristics, competitive landscape, and macroconditions. Results are averaged over years and cross-validation iterations. Standard errors are provided in parentheses. Bold numbers indicate that the incremental predictive power of photos is significantly different from zero at the 0.05 level with a two-sided test.

[a]Best in the comparison (row) or not significantly different from best in the comparison (row) at the 0.05 level with a two-sided test.

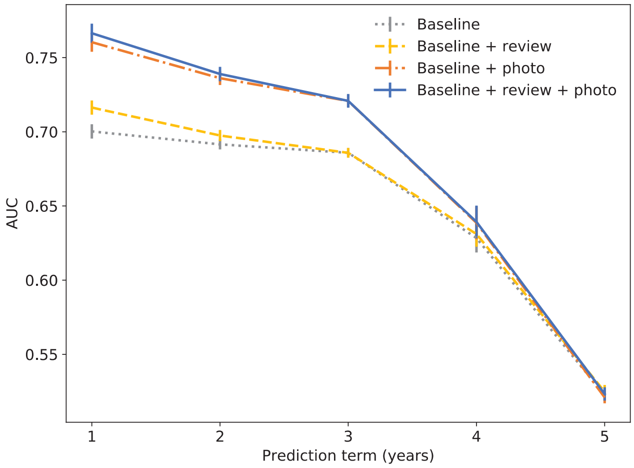
[b]Best in the comparison (row) or not significantly different from best in the comparison (row) at the 0.10 level with a two-sided test.

photos/reviews in 2010 predicting whether a restaurant would survive until the end of 2013). Therefore, we redefine the dependent variable as $\widetilde{y_{it+\Delta t}} = 0$ if any of $y_{it+1}, y_{it+2}, \ldots, y_{it+\Delta t}$ is 0 (recall that 0 means exit); otherwise, $\widetilde{y_{it+\Delta t}} = 1$. We use $\boldsymbol{X}_{it}$ to predict $\widetilde{y_{it+\Delta t}}$ with $\Delta t = 1$, 2, 3, 4, 5 years. We remove all restaurants whose next Δt-year periods are censored at the end of year 2015 as we do not know their survival status in the next Δt-year period. For each Δt, we train separate models (baseline, baseline + review, baseline + photo, baseline + review + photo) with data till t and tenfold cross-validation. To be consistent with our main analysis, we predict the survival of out-of-sample restaurants in out-of-sample time periods.[20] We report average performances of these models across t and across tenfold cross-validation in Figure 5.

As expected, the predictive power of all models declines with forecast duration. Interestingly, the AUC curve of the "baseline + photo" model is significantly above that of the baseline model for up to three years. Nevertheless, the "baseline + review" model is significantly above the baseline model for only one year. These findings indicate that consumer-posted photos can predict restaurant survival for up to three years, whereas reviews are only predictive for one year. Moreover, the "baseline + photo" model is almost as good as the "baseline + review + photo" model, indicating that improvement in multiple-year survival prediction mainly stems from photos.

As robustness checks, we also explore alternative specifications for multiple-year survival predictions in Online Appendix D. Compared with the main specification in which we predict survival during the future Δt years

Figure 5. (Color online) Multiple-Year Survival Prediction Results (AUC): Predictive Power of Photos Lasts Longer Than Reviews



Notes. Baseline model includes restaurant characteristics, competitive landscape, and macroconditions. Results are averaged over years and cross-validation iterations. The error bars represent ±1 times the standard error of each point estimate.

(e.g., forecasting whether a restaurant would survive until the end of 2013), these robustness checks break down the prediction for each year (e.g., predicting whether a restaurant would survive in 2011, 2012, or 2013). To make these robustness checks comparable with the main specification, we multiply the predicted survival probability for each future year to derive the predicted survival probability during the future Δt years. For instance, $\widetilde{y_{i2013}} = \widehat{y_{i2011}} * \widehat{y_{i2012}} * \widehat{y_{i2013}}$. Specifically, we try out three different model specifications within this setup: (1) use $\boldsymbol{X}_{it}$ to predict $y_{it+\Delta t}$, (2) predict $y_{it+\Delta t}$ assuming $x_{i,t+\Delta t-1} = \ldots = x_{i,t+1} = x_{it}$ for one-period variables with cumulative variables in future periods derived from the cumulative variables in the current period and the respective one-period variables in future periods,[21] and (3) predict survival in year $t + \Delta t$ based on distributed lag model (Mela et al. 1997) (i.e., use $\boldsymbol{X}_{it}$ to predict y_{it+1}, use $\widehat{y_{it+1}}$ and $\widehat{\boldsymbol{X}_{it+1}}$ to predict y_{it+2}, use $\widehat{y_{it+2}}$ and $\widehat{\boldsymbol{X}_{it+2}}$ to predict y_{it+3} with $\widehat{\boldsymbol{X}_{it+1}}$ and $\widehat{\boldsymbol{X}_{it+2}}$ derived in the same fashion as in specification 2). Findings from all robustness checks are consistent with what we learn from the main specification.

Our multiple-year survival prediction is related to the literature on long-term effects of marketing (e.g., Pauwels et al. 2002, Ataman et al. 2010, Montabone and Soto 2010, Brüggen et al. 2011). There are two common approaches to capture long-term effects: (1) relate the treatment to the long-term outcome directly (Brüggen et al. 2011, Liu et al. 2017) and (2) derive the long-term effect from the intermediate effect quantified with a model (e.g., the distributed lag model in Mela et al. 1997, the hidden Markov model in Montoya et al. 2010, and the vector autoregressive model in Pauwels et al. 2002; see reviews in Ataman et al. 2010 and Pauwels et al. 2002). Our main model specification and first alternative specification are consistent with the first approach that does not make any functional form assumptions. The other two alternative specifications belong to the second approach.

4. Cluster-Robust Causal Forests for Parameter Interpretation

Whereas it generates a kernel of interesting and managerially relevant insights, the XGBoost algorithm coupled with SHAP values as described is not ideal for obtaining unbiased/consistent parameter estimates quantifying how each independent variable relates to restaurant survival. Within our context, the key benefit of implementing a cluster-robust causal forests model (Athey and Wager 2019, Athey et al. 2019) is that causal forests emphasize obtaining consistent estimates of treatment effects rather than maximizing out-of-sample prediction accuracy as in many prediction-based machine learning models. Although we can plot each independent variable against its SHAP value across observations to check whether it is positively or negatively associated with survival (Lundberg et al. 2020), the interpretations may be subject to bias because of common issues in predictive machine learning models, such as regularization and overfitting (Chernozhukov et al. 2018). The gradient-boosted trees model as described in Section 3 is not an exception. Therefore, findings from our causal forests model can help us better gauge the extent to which each of the top 35 predictors associates with restaurant survival. Similar causal forests models are used in some recent marketing papers (e.g., Narang et al. 2019, Guo et al. 2021). In Section 4.1, we describe how we apply the cluster-robust causal forests model in our setting. In Section 4.2, we report results from the model.

4.1. Model: Cluster-Robust Causal Forests

As discussed in Athey and Imbens (2016) and Wager and Athey (2018), the causal forests model can serve as a good alternative to conventional propensity score methods in inferring treatment effects from rich observational data such as ours. Similar to propensity score matching (Rubin 1973, Rosenbaum and Rubin 1983) in spirit, cluster-robust casual forests are built upon an outcome variable (y_{it}), a treatment variable (W_{it-1}), and a set of control variables ($\boldsymbol{X}_{it-1}$). For instance, when estimating the treatment effect of the proportion of photos with helpful votes (W_{it-1}) on restaurant survival (y_{it}), we include a vector of control variables ($\boldsymbol{X}_{it-1}$), consisting of the number of photos and reviews, photo content, review-related variables, restaurant characteristics, competition, year dummies, and zip codes. The relationship between the outcome, treatment, and controls is modeled in Equation (2) (Athey et al. 2019):

$$y_{it} = W_{it-1}\tau(\boldsymbol{X}_{it-1}) + \mu(\boldsymbol{X}_{it-1}) + \varepsilon_{it}. \tag{2}$$

This model aims to estimate the treatment effect $\tau(\boldsymbol{X}_{it-1})$ conditional on $\boldsymbol{X}_{it-1}$ with $\mu(\boldsymbol{X}_{it-1})$ capturing effects only from controls. The treatment variable is assumed to be independent of potential outcome conditional on controls (Athey et al. 2019): $\{Y_{it}^{(W)}\} \perp W_{it-1} \mid \boldsymbol{X}_{it-1}$. As discussed in Wooldridge (2009), one way to mitigate endogeneity bias resulting from unobservables is to include as many variables as possible in the model to reduce the correlation between the treatment/control variables and the error term ε_{it}. As shown in Online Figure A19, we include as many as 95 controls in our causal forest estimations. Similar to Hollenbeck (2018), the inclusion of these controls is crucial to our empirical strategy, which may be viewed as descriptive while attempting to come close to causality.

Compared with conventional propensity score matching, cluster-robust causal forests utilize a flexible nonparametric data-driven approach to determine similarity across observations. As Athey et al. (2019) discuss, this approach is particularly advantageous with a large number of controls as in our study. Additionally, as Fong et al. (2018) mention, the estimation of traditional propensity score methods is often sensitive to the model specification, especially when the treatment variable is continuous. The causal forests are immune to such problems because the building of an honest tree (the building block of causal forests) does not rely on any particular functional form.

We describe the basic intuition of the model herein and provide more technical details in Online Appendix E. In order for the causal forests model to properly identify the treatment effects based on observation data, the correlations between the treatment variable and the control variables should not be very high (i.e., the "overlap" assumption as stated in Wager and Athey 2018). Consequently, built upon the top 35 predictors from XGBoost, we define the treatment and control variables in our casual forests model as follows: First, we consolidate highly correlated variables. For instance, we combine volume of photos and volume of reviews into one variable because these two variables are highly correlated (see Online Table A27 for a complete list of consolidated variables). Next, we decide to use cumulative rather than one-period lag variables for photos and reviews in our causal forests because (1) these two sets of variables are often highly correlated and (2) our XGBoost algorithm suggests that cumulative variables oftentimes are more predictive of restaurant survival than one-period lag variables. Finally, we include all general content types of photos (i.e., proportions of food, outside, inside, drink, and menu photos), six age brackets (≤1, 2–3, 4–7, 8–21, 22–42, >42, each covering about 16.67% of observations),[22] and the top 10 cuisine types for the completeness of the comparison. We also include average star rating based on prior literature (Chevalier and Mayzlin 2006). See Table 9 for a complete list of our treatment variables. To obtain the effect of each treatment variable, we estimate a separate causal forest model with all other treatment variables plus restaurant quality dimensions, zip code, and year as controls (see the complete list of controls in Table 9).[23]

We then employ orthogonalization (Athey et al. 2019), also used in double machine learning (Chernozhukov et al. 2018), to further reduce correlations between treatment variables and controls. Specifically, as in Equation (3), we regress out the main effect of $\boldsymbol{X}_{it-1}$ on W_{it-1} and y_{it} using regression forests and retain the residuals, W_{it-1}^* and y_{it}^*, to build causal forests. The idea is similar in spirit to propensity score adjustment for continuous treatment (Hirano and Imbens 2004) with the goal that, after such adjustment, the treatment variable becomes more independent of controls. Online Table A28 demonstrates that the correlations between the treatment variables and controls are considerably reduced after orthogonalization.

$$W_{it-1}^* = W_{it-1} - \widehat{W}_{it-1}^{(-i)}(\boldsymbol{X}_{it-1}),\ y_{it}^* = y_{it} - \widehat{y}_{it}^{(-i)}(\boldsymbol{X}_{it-1}). \tag{3}$$

Note that, even with the orthogonalization step, there are still two types of unobservables that could potentially

Table 9. Results of Cluster-Robust Causal Forests

	Treatment variable	Parameter estimate		Standard error
Photo and review	Number of photos and reviews (cum.)	0.0118		0.1301
Photo	Percentage of food photos (cum.)	0.0473	***	0.0056
	Percentage of outside photos (cum.)	0.0327	**	0.0101
	Percentage of interior photos (cum.)	0.0297	*	0.0120
	Percentage of dink photos (cum.)	−0.1958		0.1066
	Percentage of menu photos (cum.)	−0.0272		0.0420
	Percentage of photos on food and drink (cum.)	−0.0109		0.0131
	Percentage of photos with helpful votes (cum.)	0.0538	***	0.0037
Review	Percentage of mixed/negative reviews on food (cum.)	−0.0731	***	0.0129
	Avg. yearly helpful votes for reviews (cum.)	0.0351	***	0.0041
	Avg. review length (cum.)	−0.0036	***	0.0005
	Avg. star rating (cum.)	0.0047	**	0.0016
	Std. of star ratings (cum.)	−0.0229	***	0.0046
Company	Chain	0.0243	***	0.0024
	Age ≤ 1	−0.0719	***	0.0051
	Age 2-3	−0.0189	***	0.0031
	Age 4-7	0.0057	*	0.0026
	Age 8-21	0.0255	***	0.0017
	Age 22-42	0.0226	***	0.0019
	Age > 42	0.0237	***	0.0025
	Price level	−0.0062	***	0.0016
	Mexican	0.0157	***	0.0027
	American (traditional)	0.0066	**	0.0024
	Pizza	0.0060	*	0.0028
	Nightlife	0.0152	***	0.0026
	Fast Food	0.0095	***	0.0028
	Sandwiches	0.0057	*	0.0027
	American (new)	−0.0034		0.0032
	Burgers	−0.0018		0.0031
	Italian	−0.0006		0.0031
	Chinese	0.0331	***	0.0029
Competition	Number of competitors (one-period)	−0.0189	***	0.0044
	Avg. of competitors' number of photos and reviews (cum.)	−0.0718	***	0.0196
	Avg. of competitors' avg. star rating (cum.)	−0.0212	*	0.0093
Additional controls	Restaurant quality dimensions including: (1) Prop. of reviews mentioning each restaurant quality dimension (food, service, environment, price) (cum.) and (2) Avg. sentiment of each dimension (cum.)			
	Zip codes			
	Year dummies			

Notes. A positive sign means positive association with restaurant survival. Each row is a separate cluster-robust causal forests estimation. We rescaled "avg. review length (cum.)," "Number of competitors (one-period)," and "avg. of competitors' number of photos and reviews (cum.)" by 1/100. Average sentiment of each quality dimension is not included as controls when "avg. star rating" is the treatment variable because of high correlation.

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

threaten our attempts to properly identify the treatment effects in our setting: (1) time-invariant restaurant-level unobservables (e.g., owner education) and (2) time-variant unobservables (e.g., changes in finances or chef). Our cluster-robust causal forests account for time-invariant restaurant-level unobservables by allowing clustered errors (Athey et al. 2019). All observations from one restaurant are considered as a cluster. Intuitively the restaurant–year observations are not entirely independent and subject to unobserved restaurant-level effects that might influence survival. Essentially, utilizing clustered errors with causal forests is analogous to a nonparametric random-effect model that does not make distributional assumptions (Athey and Wager 2019).[20]

We try to mitigate potential bias caused by time-variant unobservables via our control variables. Based on directed acyclical graph theory (Pearl 2014, Morgan and Winship 2015), the inclusion of controls is often adequate to correct such bias as long as control variables fully absorb the impact of time-variant unobservables on the outcome variable. Within our context, for any time-variant unobservables whose effects on survival can be fully captured by our restaurant quality measures, controlling restaurant quality is adequate to correct the bias caused by such unobservables. Namely, if any change in chef impacts a restaurant's survival by affecting its quality in food, service, environment, or price, we should be able to absorb such impact by including our restaurant quality metrics in the controls.

Within this context, we follow Athey et al. (2019) by building honest trees to weight the similarity between a set of values for controls x and an arbitrary $\boldsymbol{X}_{it-1}$ denoted by $\alpha_{it}(\boldsymbol{x})$. Then, we can calculate $\widehat{\tau}(\boldsymbol{x})$ (the conditional treatment effects at x) and $\widehat{\tau}$ (the average treatment effect) as in Equation (4) (Athey and Wager 2019, Athey et al. 2019):

$$\widehat{\tau}(\boldsymbol{x}) = \frac{\sum_{it}\alpha_{it}(\boldsymbol{x})(W^*_{it-1} - \overline{W^*_\alpha})(y^*_{it} - \overline{y^*_\alpha})}{\sum_{it}\alpha_{it}(\boldsymbol{x})(W^*_{it-1} - \overline{W^*_\alpha})^2}, \text{ where}$$
$$\overline{W^*_\alpha} = \sum_{it}\alpha_{it}(\boldsymbol{x})W^*_{it-1}, \ \overline{y^*_\alpha} = \sum_{it}\alpha_{it}(\boldsymbol{x})y^*_{it}$$
$$\widehat{\tau} = \frac{1}{\sum_{it}(T_i - T_i^0 + 1)}\sum_{it}\widehat{\tau}(\boldsymbol{X}_{it-1}). \tag{4}$$

4.2. Results: Cluster-Robust Causal Forests

Table 9 presents results from our cluster-robust causal forests model.[25] As discussed, each row is a separate cluster-robust causal forests estimation, in which we treat the focal variable as treatment and all other observable factors (including other variables listed in Table 9, restaurant quality measures, zip code dummies, and year dummies) as controls. A positive sign means a positive association with restaurant survival.

Among photo content variables, the proportion of food photos has the largest positive association with restaurant survival, followed by the proportions of outside and interior photos. Such a result is not surprising because prior literature suggests that food is the most critical aspect of a restaurant (Sulek and Hensley 2004, Duarte Alonso et al. 2013). Food photos could be what get consumers in the door. Outside photos show outdoor attractions, a vital factor for tourism regions (Getz and Brown 2006). Consumers may also use outside photos to check the neighborhood and find a restaurant's gate. Inside photos depict the ambiance of a restaurant. Overall, these findings indicate that photos may offer useful cues to provide a better horizontal match between restaurants and consumer preferences.

Moreover, the proportion of photos with helpful votes is also positively associated with restaurant survival. In practice, helpful votes may also reveal negative aspects of a restaurant (e.g., stains on the table). Nevertheless, we learn from our photo caption analysis that most photos carry neutral or positive sentiment. Mudambi and Schuff (2010) suggest that a review's helpful votes reflect its information diagnosticity in facilitating the consumer's purchase decision process. In a similar spirit, we conjecture that helpful votes on photos might (1) reflect the degree of diagnosticity associated with photos, which, in turn, may facilitate a better match between consumers and the focal business, and/or (2) serve as a proxy of demand and, hence, positively relate to survival. Interestingly, after all other factors controlled, the cumulative number of photos and reviews is not significantly associated with restaurant survival.

For reviews, we find that restaurant survival is negatively associated with the proportion of mixed/negative reviews on food, review length, and standard deviation of star ratings, whereas it is positively associated with helpful votes for reviews and average star rating. The results on review length and star rating are consistent with prior literature studying the relationship between consumer text reviews and demand (e.g., Chevalier and Mayzlin 2006, Zhu and Zhang 2010). Table 9 also reveals the types of restaurants with better chances of survival. We observe a general trend that more established restaurants have a higher survival chance, resonating with Carroll (1983). In particular, the first three years are the hardest for restaurants to survive, and the survival rate is quite stable after the first seven years. Additionally, mainstream cuisine-type restaurants (e.g., Mexican, American traditional, pizza, nightlife, fast food, sandwiches, and Chinese categories), lower-priced restaurants, and chain restaurants have better odds of survival. Finally, we find that restaurants are less likely to survive when there are more competitors or when competitors receive more photos and reviews or have better star ratings. Restaurants usually concentrate in locations with more foot traffic and, consequently, may face stiffer competition. These findings indicate that the competition effect outweighs the demand effect regarding competitor concentration, reinforcing that monitoring competitors is important for restaurant managers.

5. Conclusions

We investigate whether consumer-posted photos can forecast business survival above and beyond reviews, restaurant characteristics, competitive landscape, and macroconditions. Utilizing a predictive algorithm called XGBoost, we discover that photos provide significant incremental predictive power for restaurant survival. Specifically, the informativeness of photos (i.e., photo content) carries the most predictive power for restaurant survival compared with other aspects of photos, such as photographic attributes (e.g., composition or brightness). In particular, the cumulative proportion of food photos is the most predictive among all photo-related variables, echoing prior literature that suggests food is the most critical aspect of a restaurant. We also find that photos are more informative in predicting the survival of independent, young or midaged, and medium-priced restaurants. Moreover, assuming that restaurant owners possess no knowledge of future UGC for the focal business and its competitors, current photos can predict restaurant survival for up to three years, whereas reviews are predictive for only one year. We further develop cluster-robust causal forests with the aim to obtain unbiased/consistent parameter estimates quantifying the relationship between the set of most informative predictors (based on our XGBoost algorithm) and restaurant survival. Among photo content variables, the proportion of food photos has the largest positive association with restaurant survival, followed by proportions of outside and interior photos. The proportion of photos with helpful votes is also positively related to restaurant survival. To the best of our knowledge, this is the first empirical study that explores whether consumer photos can serve as a leading indicator of long-term business prosperity. Our research is also among the most comprehensive empirical studies on restaurant survival to date.

In our modern era of rapidly increasing computing power, the execution cost of utilizing photos in restaurant survival prediction is rather modest compared with the great potential underlying its usefulness in practice. Multiple stakeholders can readily apply our proposed framework to improve their decision-making processes. Business investors and landlords can employ our research results to obtain a better evaluation of the market as well as to monitor restaurant survival likelihood. Our models can significantly increase survival prediction accuracy for better-informed capital investment/lease decisions on restaurants. Online platforms can also utilize our method for premium business intelligence. Furthermore, our findings regarding consumer-posted photos can provide foresight to managers and investors in terms of survival likelihood for up to three years, which can be highly valuable for competitive intelligence, resource allocation, and longer term strategic planning. Finally, our research offers insights into the macroperformance of the restaurant industry for marketing research firms and trade associations.

Our research is also subject to several limitations and provides some fruitful directions for future research. First, future research may follow Hollenbeck (2018) by collecting restaurant reviews from multiple platforms (e.g., TripAdvisor, Google Reviews) to obtain a better measure of restaurant quality. Under our approach, the quality metrics for less popular restaurants might be noisier than those of popular restaurants because the former have fewer reviews. Collecting reviews from multiple platforms provides a larger pool of reviews and may potentially mitigate the measurement error of restaurant quality from Yelp only. Based on simulation studies, we learn that, as long as the quality measure is unbiased, even if the magnitude of its noise (i.e., variance of noise) is larger for less popular restaurants, our estimates of the treatment effects would remain consistent. Second, future research may consider collecting census data to gather possibly richer information on these restaurants to further enrich the restaurant survival study. Third, given that we do not have a time stamp of when a photo/review receives a helpful vote, we cannot directly quantify helpful votes received in each year for each photo/review. Future research may utilize data with a time stamp for each vote to better measure time-variant helpful votes information. Finally, whereas we emphasize the relationship between consumer-posted photos and restaurant survival, future research may examine how such photos relate to other outcomes related to long-term business prosperity, such as consumer attitudes toward the focal business and brand loyalty.

Acknowledgments
The authors thank participants at the Advanced Research Techniques Forum, the INFORMS Society for Marketing Science Doctoral Dissertation Award Special Session at the INFORMS Marketing Science Conference, the University of Texas at Dallas Frontiers of Research in Marketing Science Conference, the Product and Service Innovation Conference, Carnegie Mellon University Conference on Digital Marketing and Machine Learning, Erasmus University Workshop on Machine Learning & Multi-bandit Algorithms, GERAD Montreal, University of Miami Marketing Research Camp, MIT Marketing Group Doctoral Workshop, and seminar attendants at Chinese University of Hong Kong, City University of Hong Kong, Erasmus University, Hong Kong University of Science and Technology, University of Hong Kong, University of Iowa, National University of Singapore, Ohio State University, Peking University, Penn State University, Santa Clara University, Shanghai University of Finance and Economics, Tongji University, Tsinghua University, University of Western Ontario, and University of Wisconsin-Madison for their helpful comments.

Endnotes

[1] See https://www.nytimes.com/2015/07/23/arts/international/photos-photos-everywhere.html.

[2] See https://www.nytimes.com/2010/04/07/dining/07camera.html?hp=&pagewanted=print.

[3] See https://news.mit.edu/2014/in-the-blink-of-an-eye-0116.

[4] See https://searchengineland.com/business-profile-review-best-practices-tripadvisor-yelp-276247 and https://www.tripadvisor.com/ForRestaurants/wp-content/uploads/2018/01/Dinerengagement_us_en.pdf.

[5] Because consumers self-select into writing reviews, we fully acknowledge that Yelp reviews are not a perfect measure of restaurant quality. We have explored alternative measures of restaurant quality, such as Zagat or the Michelin Guide. Nevertheless, these sources suffer from severe restaurant-selection bias with Zagat only focusing on the most popular restaurants and Michelin mostly emphasizing high-end, upscale dining options. Because our research requires time-varying measures of restaurant quality for a large number of restaurants, we resort to text mining of Yelp reviews to discern trajectories of restaurant quality over time.

[6] Tirunillai and Tellis (2014) find that quality metrics mined from consumer reviews and those from *Consumer Reports* correlate

between 0.61 (for footwear) to 0.81 (for computers). Using 15 years of hotel reviews from multiple platforms, Hollenbeck (2018) discovers that consumer reviews reduce information asymmetry of quality between sellers and buyers and serve as a substitute of brand name (or chain affiliation) for independent hotels.

[7] In Section 4.1, we discuss in detail how we mitigate possible time-varying restaurant unobservables (e.g., changes in finances or chef) via our restaurant quality measures based on the directed acyclical graph theory (Pearl 2014, Morgan and Winship 2015).

[8] See https://restaurantengine.com/startup-restaurants-typically-overspend/.

[9] The number is calculated by summing up market capitalization of all publicly traded restaurant companies listed on https://markets.on.nytimes.com/research/markets/usmarkets/industry.asp?industry=53312 accessed on 05/10/2019.

[10] For example, a photo received 100 helpful votes at the end of year 2015, and the photo was posted on the focal restaurant in 2005. Then, the yearly helpful votes for this photo is $100/(2015 - 2005) = 10$.

[11] See https://engineeringblog.yelp.com/2015/10/how-we-use-deep-learning-to-classify-business-photos-at-yelp.html.

[12] See https://clarifai.com/models/food-image-recognition-model-bd367be194cf45149e75f01d59f77ba7.

[13] We observed only eight restaurants with birth year = 2004 in the entire sample. Among these, four restaurants have their birth years inferred from Yelp reviews or photos, and these four restaurants only have 36 restaurant-year observations in total.

[14] We did not perform such tasks to forecast survival before year 2010 because there are very little data for model calibration (less than 20% of total restaurant–year observations were from years before 2010).

[15] We take within-restaurant variation into account in Section 4 when our primary goal is to interpret parameter estimates rather than maximizing out-of-sample prediction accuracy.

[16] We use the following demonstration to illustrate how we predict survival in the first year. For example, suppose restaurant i was founded in 2013. When predicting survival in 2013 and when restaurant i is in the 10% holdout restaurants, we use $\boldsymbol{X}_{i2012}$ to predict y_{i2013}. In this case, year 2012 is essentially defined as period zero for the focal restaurant, and $\boldsymbol{X}_{i2012}$ is defined as follows: (1) the UGG volume of restaurant i is set to be zero; (2) age is zero; (3) chain status, cuisine types, and price level are set at x_i; (4) for competition variables, competitors would be all open restaurants in 2012 in the same zip code with all competition-related variables computed based on this set of restaurants; and (5) year dummy set as one for year 2012 and zip code being the location of restaurant i.

[17] In Table 6, we use the superscript "a" to indicate whether there exists a statistically significant difference between the best performing model and other models. For example, in the "specificity" column of Table 6, the (baseline + photo) model has an "a," but (baseline + reviews) does not, indicating that the (baseline + photo) model has a statistically superior performance compared with the (baseline + review) model.

[18] The top 35 predictors are generally consistent when forecasting survival from 2010 to 2015. The results in Figure 4 are based on the scenario of using data until 2014 to predict survival in year 2015. We choose to report results from this scenario because it covers the most observations and contains the most recent and comprehensive data in our sample.

[19] We exclude age = 0 in this comparison because restaurants have no photos at age = 0. We also tried other age cutoff values (e.g., $1 \leq \text{age} \leq 5$, $5 < \text{age} \leq 30$, and age > 30), and the results were consistent. In Online Appendix D, we further report photos' incremental predictive power by each age for restaurants younger than or equal to five years old in order to better understand photos' predictive power because younger restaurants tend to be more susceptible to failure.

[20] For example, when using information until year 2010 to predict whether a restaurant would survive until the end of year 2013 (i.e., using $\boldsymbol{X}_{i2010}$ to predict $\widetilde{y_{i2013}}$), for each cross-validation iteration, we use data till year 2010 of nine buckets of restaurants to calibrate the model. Then, to test the model's performance, for all open restaurants at the end of year 2010 in the remaining bucket, we predict whether they would survive until the end of year 2013.

[21] For example, we assume # of photos_{it} (one-period) = # of photos_{it-1} (one-period). And # of photos_{it} (cum.) = # of photos_{it-1} (cum.) + # of photos_{it} (one-period).

[22] We did not use age dummies (age = 0, 1, 2, ..., 21, 21+) for causal forests because a single age can have very few restaurant observations, making it difficult to satisfy the "overlap" assumption in causal forests model. This assumption requires that there should be enough observations in both treated and untreated conditions near any set of values of controls, which also implies the correlations between the treatment variable and the control variables should not be very high (Wager and Athey 2018).

[23] When an age bracket dummy (e.g., $4 \leq \text{age} \leq 7$) is the treatment variable, all other age brackets are combined as the untreated group because it provides a relatively larger pool of untreated units to satisfy the "overlap" assumption.

[24] We consider fixed effects to account for unobserved restaurant characteristics. Nevertheless, because our dependent variable is censored and nonrepeating (i.e., each restaurant can only fail once), simply adding fixed effects to such nonlinear models leads to inconsistent and biased coefficient estimates (Greene 2004).

[25] Given that our focus here is to interpret how each independent variable relates to restaurant survival, we use all data to estimate the causal forests because there is no longer a need to separate out the training and testing samples.

References

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

Arnheim R (1965) *Art and Visual Perception: A Psychology of the Creative Eye* (University of California Press, Berkeley, CA).

Ataman MB, Van Heerde HJ, Mela CF (2010) The long-term effect of marketing strategy on brand sales. *J. Marketing Res.* 47(5): 866–882.

Athey S, Imbens G (2016) Recursive partitioning for heterogeneous causal effects. Shiffrin RM, ed. *Proc. Natl. Acad. Sci. USA* 113(27): 7353–7360.

Athey S, Wager S (2019) *Estimating Treatment Effects with Causal Forests: An Application* (Observational Studies), 5(2):37–51.

Athey S, Tibshirani J, Wager S (2019) Generalized random forests. *Ann. Statist.* 47(2):1148–1178.

Audretsch DB, Mahmood T (1995) New firm survival: New results using a hazard function. *Rev. Econom. Statist.* 77(1):97–103.

Audretsch DB, Houweling P, Roy Thurik A (2000) Firm survival in the Netherlands. *Rev. Indust. Organ.* 16(1):1–11.

Barasch A, Zauberman G, Diehl K (2017a) How the intention to share can undermine enjoyment: Photo-taking goals and evaluation of experiences. *J. Consumer Res.* 44(6):1220–1237.

Barasch A, Diehl K, Silverman J, Zauberman G (2017b) Photographic memory: The effects of volitional photo taking on memory for visual and auditory aspects of an experience. *Psych. Sci.* 28(8):1056–1066.

Bates T (1990) Entrepreneur human capital inputs and small business longevity. *Rev. Econom. Statist.* 72(4):551–559.

Bates T (1995) A comparison of franchise and independent small business survival rates. *Small Bus. Econom.* 7(5):377–388.

Bates T (1998) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *J. Bus. Venturing* 13(2): 113–130.

Bekkar M, Djemaa HK, Alitouche TA (2013) Evaluation measures for models assessment over imbalanced data sets. *J. Inform. Engrg. Appl.* 3(10):27–38.

Blei DM, Ng AY, Jordan MI (2003) Latent Dirichlet allocation. *J. Machine Learn. Res.* 3:993–1022.

Boden RJ Jr, Nucci AR (2000) On the survival prospects of men's and women's new business ventures. *J. Bus. Venturing* 15(4): 347–362.

Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30(7):1145–1159.

Brodersen KH, Ong CS, Stephan KE, Buhmann JM (2010) The balanced accuracy and its posterior distribution. *IEEE 20th Internat. Conf. Pattern Recognition*, 3121–3124.

Brüggen EC, Foubert B, Gremler DD (2011) Extreme makeover: Short-and long-term effects of a remodeled servicescape. *J. Marketing* 75(5):71–87.

Bujisic M, Hutchinson J, Parsa HG (2014) The effects of restaurant quality attributes on customer behavioral intentions. *Internat. J. Contemporary Hospitality Management* 26(8):1270–1291.

Carroll GR (1983) A stochastic model of organizational mortality: Review and reanalysis. *Soc. Sci. Res.* 12(4):303–329.

Chang J, Gerrish S, Wang C, Boyd-Graber J, Blei D (2009) Reading tea leaves: How humans interpret topic models. *Adv. Neural Inform. Processing Systems* 22:288–296.

Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. *Proc. 22nd ACM Sigkdd Internat. Conf. Knowledge Discovery Data Mining*, 785–794.

Chen T, Benesty M, He T (2018) *Understand Your Dataset with Xgboost* (R Document).

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double/debiased machine learning for treatment and structural parameters. *Econom. J.* 21:C1–C68.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* 43(3):345–354.

Cooper AC, Gimeno-Gascon FJ, Woo CY (1994) Initial human and financial capital as predictors of new venture performance. *J. Bus. Venturing* 9(5):371–395.

Diehl K, Zauberman G, Barasch A (2016) How taking photos increases enjoyment of experiences. *J. Personality Soc. Psych.* 111(2):119–140.

Duarte Alonso A, O'Neill M, Liu Y, O'Shea M (2013) Factors driving consumer restaurant choice: An exploratory study from the Southeastern United States. *J. Hospitality Marketing Management* 22(5):547–567.

Dzyabura D, Hauser JR (2011) Active machine learning for consideration heuristics. *Marketing Sci.* 30(5):801–819.

Dzyabura D, Peres R (2021) Visual elicitation of brand perception. *J. Marketing* 85(4):44–66.

Dzyabura D, El Kihal S, Hauser JR, Ibragimov M (2019) Leveraging the power of images in managing product return rates. Preprint, submitted September 4, http://dx.doi.org/10.2139/ssrn.3209307.

Fawcett T (2004) ROC graphs: Notes and practical considerations for researchers. *Machine Learn.* 31(1):1–38.

Fawcett T (2006) An introduction to ROC analysis. *Pattern Recognition Lett.* 27(8):861–874.

Fellbaum C (1998) *WordNet: An Electronic Lexical Database* (MIT Press, Cambridge, MA).

Fong C, Hazlett C, Imai K (2018) Covariate balancing propensity score for a continuous treatment: Application to the efficacy of political advertisements. *Ann. Appl. Statist.* 12(1):156–177.

Freeman M (2007) *The Photographer's Eye: Composition and Design for Better Digital Photos* (Routledge, New York).

Friedman J (2001) Greedy function approximation: A gradient boosting machine. *Ann. Statist.* 29(5):1189–1232.

Fritsch M, Brixy U, Falck O (2006) The effect of industry, region, and time on new business survival—A multi-dimensional analysis. *Rev. Indust. Organ.* 28(3):285–306.

Getz D, Brown G (2006) Critical success factors for wine tourism regions: A demand analysis. *Tourism Management* 27(1):146–158.

Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Sci.* 31(3):493–520.

Greene W (2004) The behaviour of the maximum likelihood estimator of limited dependent variable models in the presence of fixed effects. *Econom. J.* 7(1):98–119.

Guo T, Sriram S, Manchanda P (2021) The effect of information disclosure on industry payments to physicians. *J. Marketing Res.* 58(1):115–140.

Haapanen M, Tervo H (2009) Self-employment duration in urban and rural locations. *Appl. Econom.* 41(19):2449–2461.

Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1):29–36.

Hartmann J, Heitmann M, Schamp C, Netzer O (2021) The power of brand selfies in consumer-generated brand images. *J. Marketing Res.* 58(6):1159–1177.

Hauser JR, Dong S, Ding M (2014) Self-reflection and articulated consumer preferences. *J. Product Innovation Management* 31(1): 17–32.

Hirano K, Imbens GW (2004) The propensity score with continuous treatments. Gelman A, Meng X-L, eds. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (John Wiley & Sons, Ltd., West Sussex, England), 73–84.

Hollenbeck B (2018) Online reputation mechanisms and the decreasing value of chain affiliation. *J. Marketing Res.* 55(5):636–654.

Huang D, Luo L (2016) Consumer preference elicitation of complex products using fuzzy support vector machine active learning. *Marketing Sci.* 35(3):445–464.

Hutto CJ, Gilbert E (2014) VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proc. Eighth Internat. AAAI Conf. Weblogs Social Media*, 8(1):216–225.

Hyun SS (2010) Predictors of relationship quality and loyalty in the chain restaurant industry. *Cornell Hospitality Quart.* 51(2):251–267.

Isola P, Parikh D, Torralba A, Oliva A (2011) Understanding the intrinsic memorability of images. *Adv. Neural Inform. Processing Systems* 24:2429–2437.

Kalleberg AL, Leicht KT (1991) Gender and organizational performance: Determinants of small business survival and success. *Acad. Management J.* 34(1):136–161.

Kalnins A, Lafontaine F (2013) Too far away? The effect of distance to headquarters on business establishment performance. *Amer. Econom. J. Microeconomics* 5(3):157–179.

Kalnins A, Mayer KJ (2004) Franchising, ownership, and experience: A study of pizza restaurant survival. *Management Sci.* 50(12): 1716–1728.

Kamshad KM (1994) Firm growth and survival: Does ownership structure matter? *J. Econom. Management Strategy* 3(3):521–543.

Ko EE, Bowman D (2020) The effect of sentiment and complexity on consumer engagement with brand-themed user-generated content. *Proc. 49th Eur. Marketing Acad*. (63560).

Kotsiantis S, Kanellopoulos D, Pintelas P (2006) Handling imbalanced datasets: A review. *GESTS Internat. Trans. Comput. Sci. Engrg.* 30(1):25–36.

Krages B (2005) *Photography: The Art of Composition* (Allworth Press, New York).

Kreitler H, Kreitler S (1972) *Psychology of the Arts* (Duke University Press, Durham, NC).

Kullback S, Leibler RA (1951) On information and sufficiency. *Ann. Math. Statist.* 22(1):79–86.

Lafontaine F, Zapletal M, Zhang X (2018) Brighter prospects? Assessing the franchise advantage using census data. *J. Econom. Management Strategy* 28(2):175–197.

Larsen V, Luna D, Peracchio LA (2004) Points of view and pieces of time: A taxonomy of image attributes. *J. Consumer Res.* 31(1): 102–111.

Lee D, Manzoor EA, Cheng Z (2019) Focused concept miner (FCM): An interpretable deep learning for text exploration. Working paper, Boston University, Questrom School of Business, Boston, MA.

Li Y, Xie Y (2020) Is a picture worth a thousand words? An empirical study of imagery content and social media engagement. *J. Marketing Res*. 57(1):1–19.

Li X, Shi M, Wang XS (2019) Video mining: Measuring visual information using automatic methods. *Internat. J. Res. Marketing* 36(2):216–231.

Liu Y (2006) Word-of-mouth for movies: Its dynamics and impact on box office revenue. *J. Marketing* 70(3):74–89.

Liu L, Dzyabura D, Mizik N (2020) Visual listening in: Extracting brand image portrayed on social media. *Marketing Sci.* 39(4): 669–848.

Liu X, Lee D, Srinivasan K (2019) Large scale cross category analysis of consumer review content on sales conversion leveraging deep learning. *J. Marketing Res.* 56(6):918–943.

Liu Y, Shankar V, Yun W (2017) Crisis management strategies and the long-term effects of product recalls on firm value. *J. Marketing* 81(5):30–48.

Lundberg S, Lee SI (2017) A unified approach to interpreting model predictions. *31st Conf. Neural Inform. Processing Systems*, Long Beach, CA.

Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, Katz R, Himmelfarb J, Bansal N, Lee S-I (2020) From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence* 2(1):56–67.

Malik N, Singh PV, Srinivasan K (2020) A dynamic analysis of beauty premium. Working paper, University of Southern California, Marshall School of Business, Los Angeles.

McGrath T (2017) *Social Listening Meets Image Sharing: A Picture Is Worth 1,000 Words* (Marketing Science Institute).

Mela CF, Gupta S, Lehmann DR (1997) The long-term impact of promotion and advertising on consumer brand choice. *J. Marketing Res.* 34(2):248–261.

Miller GA (1995) Wordnet: A lexical database for English. *Comm. ACM* 38(11):39–41.

Molnar C (2021) Interpretable machine learning: A guide for making black box models explainable. https://christophm.github.io/interpretable-ml-book/.

Montabone S, Soto A (2010) Human detection using a mobile platform and novel features derived from a visual saliency mechanism. *Image Vision Comput.* 28(3):391–402.

Montoya R, Netzer O, Jedidi K (2010) Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Sci.* 29(5):909–924.

Morgan SL, Winship C (2015) *Counterfactuals and Causal Inference* (Cambridge University Press, New York).

Mudambi SM, Schuff D (2010) What makes a helpful online review? A study of customer reviews on Amazon.com. *Management Inform. Systems Quart.* 34(1):185–200.

Narang U, Shankar V, Narayanan S (2019) The impact of mobile app failures on purchases in online and offline channels. Working paper, University of Illinois at Urbana–Champaign, Champaign, IL.

Netzer O, Lemaire A, Herzenstein M (2019) When words sweat: Identifying signals for loan default in the text of loan applications. *J. Marketing Res.* 56(6):960–980.

Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Sci.* 31(3):521–543.

Newman D, Lau JH, Grieser K, Baldwin T (2010) Automatic evaluation of topic coherence. *Human Language Tech Annual Conf. North Amer. Chapter Assoc. Comput. Linguistics* (Los Angeles), 100–108.

Nishiyama M, Okabe T, Sato I, Sato Y (2011) Aesthetic quality classification of photographs based on color harmony. *Comput. Vision Pattern Recognition*, 33–40.

Parsa HG, Gregory A, Terry MD (2010) Why do restaurants fail? Part III: An analysis of macro and micro factors. *Emerging Aspects Redefining Tourism Hospitality* 1(1):16–25.

Parsa HG, Self JT, Njite D, King T (2005) Why restaurants fail. *Cornell Hotel Restaurant. Admin. Quart.* 46(3):304–322.

Parsa HG, Self J, Sydnor-Busso S, Yoon HJ (2011) Why restaurants fail? Part II—The impact of affiliation, location, and size on restaurant failures: Results from a survival analysis. *J. Foodservice. Bus. Res.* 14(4):360–379.

Parsa HG, Van Der Rest JPI, Smith SR, Parsa RA, Bujisic M (2015) Why restaurants fail? Part IV: The relationship between restaurant failures and demographic factors. *Cornell Hospitality Quart.* 56(1):80–90.

Pauwels K, Hanssens DM, Siddarth S (2002) The long-term effects of price promotions on category incidence, brand choice, and purchase quantity. *J. Marketing Res.* 39(4):421–439.

Pearl J (2014) The deductive approach to causal inference. *J. Causal Inference* 2(2):115–129.

Puranam D, Narayan V, Kadiyali V (2017) The effect of calorie posting regulation on consumer opinion: A flexible latent Dirichlet allocation model with informative priors. *Marketing Sci.* 36(5):726–746.

Rafieian O, Yoganarasimhan H (2021) Targeting and privacy in mobile advertising. *Marketing Sci*. 40(2):193–218.

Rajaram P, Puneet M, Schwartz E (2021) Finding the sweet spot: Ad targeting on streaming media. Preprint, submitted August 6, http://dx.doi.org/10.2139/ssrn.3496039.

Röder M, Both A, Hinneburg A (2015) Exploring the space of topic coherence measures. *Proc. Eighth ACM Internat. Conf. Web Search Data Mining*, 399–408.

Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. *Biometrika* 70(1):41–55.

Rubin DB (1973) Matching to remove bias in observational studies. *Biometrics* 29(1):159–183.

Ryu K, Lee H-R, Kim WG (2012) The influence of the quality of the physical environment, food, and service on restaurant image, customer perceived value, customer satisfaction, and behavioral intentions. *Internat. J. Contemporary Hospitality Management* 24(2):200–223.

Seiffert C, Khoshgoftaar TM, Van Hulse J, Napolitano A (2008) Resampling or reweighting: A comparison of boosting implementations. *20th IEEE Internat. Conf. Tools Artificial Intelligence*, vol. 1, 445–51.

Sulek JM, Hensley RL (2004) The relative importance of food, atmosphere, and fairness of wait: The case of a full-service restaurant. *Cornell Hotel Restaurant Admin. Quart.* 45(3):235–247.

Takahashi K, Doman K, Kawanishi Y, Hirayama T, Ide I, Deguchi D, Murase H (2016) A study on estimating the attractiveness of food photography. *2016 IEEE Second Internat. Conf. Multimedia Big Data*, 444–449.

Timoshenko A, Hauser JR (2019) Identifying customer needs from user-generated content. *Marketing Sci.* 38(1):1–20.

Tirunillai S, Tellis GJ (2012) Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Sci.* 31(2):198–215.

Tirunillai S, Tellis G (2014) Extracting dimensions of consumer satisfaction with quality from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *J. Marketing Res.* 51(4):463–479.

Toubia O, Netzer O (2016) Idea generation, creativity, and prototypicality. *Marketing Sci.* 36(1):1–20.

Troncoso I, Luo L (2020) Look the part? The role of profile pictures in online labor markets. Working paper, University of Southern California, Marshall School of Business, Los Angeles.

Tveterås R, Eide GE (2000) Survival of new plants in different industry environments in Norwegian manufacturing: A semi-proportional Cox model approach. *Small Bus. Econom.* 14(1):65–82.

Vogel DR, Dickson GW, Lehman JA (1986) *Persuasion and the Role of Visual Presentation Support: The UM/3M Study* (Management Information Systems Research Center, School of Management, University of Minnesota, Minneapolis).

Wadhera D, Capaldi-Phillips ED (2014) A review of visual cues associated with food on food acceptance and consumption. *Eating Behav.* 15(1):132–143.

Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* 113(523):1228–1242.

Wagner J (1994) The post-entry performance of new small firms in German manufacturing industries. *J. Indust. Econom.* 42(2): 141–154.

Wooldridge J (2009) *Introductory Econometrics* (South-Western Cengage Learning, Mason, OH).

Xiao L, Ding M (2014) Just the faces: Exploring the effects of facial features in print advertising. *Marketing Sci.* 33(3):338–352.

Yoganarasimhan H (2019) Search personalization using machine learning. *Management Sci.* 66(3):1045–1070.

Zhang S, Lee D, Singh P, Srinivasan K (2018) How much is an image worth? Airbnb property demand analytics leveraging a scalable image classification algorithm. Working paper, Harvard University Business School, Boston, MA.

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