Signaling Quality via Demand Lockout

Andreas Kraft[1]

Raghunath Singh Rao

This version: August 09, 2024

[1] Kraft (Andreas.Kraft@chicagobooth.edu) is an Assistant Professor in the University of Chicago Booth School of Business and Rao (Raghunath.Rao@mccombs.utexas.edu) is a Professor the McCombs School of Business, The University of Texas at Austin.

Signaling Quality via Demand Lockout

Abstract

Consumers face uncertainty about the quality of products and services in many consumption contexts. Firms often try to resolve quality uncertainty via price signaling, where a higher price implies higher quality. However, a host of consumption contexts increasingly involve a uniform price across differentiated offerings (e.g., streaming platforms), and hence, prices as signals become unavailable. In this paper, we propose and empirically test a novel mode of quality signal: firms' active exclusion of a profitable segment of consumers—a phenomenon we call *demand lockout*. Using a theoretical model, we demonstrate that the opportunity cost of locking out a profitable segment can serve as a credible signal of quality when two conditions are met: First, the non-excluded segment is large enough, and second, a significant fraction of consumers only consume if word of mouth has reduced the quality uncertainty. The value of the lockout signal increases as advertising becomes more expensive and decreases as third-party information becomes more accurate. We provide empirical observations consistent with our model in the context of the motion picture industry, hypothesizing that studios might use R ratings to credibly signal quality by excluding a non-trivial segment from consuming its product. Our empirical analysis involves the use of a large corpus of text data from thousands of movie subtitles in conjunction with machine learning methods to control for "age-inappropriate" content of movies non-parametrically. Consistent with the proposed theory, movies are more likely to actively try to get an R rating when the value of the signal is more significant. Furthermore, box office revenue numbers are consistent with our prediction that R ratings could serve as a credible signal, and the value of this signal depends on the availability and noisiness of external information, such as film reviews.

Keywords: Signaling, Machine Learning, Game Theory, Information Asymmetry, Movies, Reviews

This version: 08/09/2024

Introduction

Identifying the quality of goods and services before a purchase can be challenging for consumers in many contexts. This information asymmetry can be substantial in industries with few repeat purchases and high search costs, such as experience goods. Many intermediaries have appeared in the marketplace to provide expert-based or "wisdom-of-crowd"-based information to resolve this uncertainty. Examples of such intermediaries include Yelp® (for restaurants), IMDb® (for movies), Consumer Reports® (electronic and other high-ticket items), and Carfax® (for automobiles).

Because resolving this uncertainty is important for consumers, firms also take steps to resolve it via active interventions- actions typically referred to as signaling (Spence 1973). Among the most widely studied marketing signals are branding (Wernerfelt 1988; Erdem and Swait 1998), pricing (Shin 2005; Rao and Monroe 1989), and advertising (Nelson 1974). However, an increasing number of firms today operate in contexts where such traditional signals are less useful or often unavailable. Consider a consumer choosing a product on a video streaming platform where the product space is vast and the marginal price for any given type is zero.[1] Advertising might also not be feasible for some firms, given the high cost of a nationwide campaign and the absence of significant repeat purchases. Joshi and Musalem (2020) show that the cost of advertising typically increases in the presence of word-of-mouth. More generally, we see many instances of uniform pricing for otherwise differentiated products. For example, in grocery stores, one often sees the brand variants (e.g., different flavors of an ice cream or soft drink brand) offered at a uniform price (Chen and Cui 2013; McMillan 2007). Similarly, online music vendors often price songs—which clearly are differentiated products—at uniform prices (Shiller and Waldfogel 2011). Apple has sold billions of songs on iTunes for $0.99—the most common price. Rental car companies uniformly price their vehicles (within a category), irrespective of their age and odometer readings (Cho and Rust 2010). Dollar stores represent canonical instances of uniform pricing. Finally, the movie theater industry is perhaps the most

[1] Streaming platforms often use recommendation systems to communicate quality of the downstream content. There is some concern that their recommendation systems might not be incentive compatible and are not sufficiently informative (Bourreau and Gaudin 2018; Calvano and Jullien 2019).

prominent example: Movies usually carry the same ticket price regardless of their budgets, actors, or critical reception (Orbach and Einav 2007; Ho et al. 2018). The conundrum of uniform pricing in these disparate markets has been rationalized by appealing to factors like fairness, managerial inertia, and legacy practices, among others.

In this paper, we take as given that in many markets, pricing is uniform, and advertising is costly. Nevertheless, firms need to convey the quality of their (differentiated) offerings to potential consumers. To do so, we posit that firms might be able to use a novel signaling mechanism that we refer to as *demand lockout,* which involves intentional action by a firm to make itself inaccessible to a profitable segment of potential consumers.[2] We propose that the opportunity cost of excluding a profitable segment can serve as a credible signal that allows consumers to make meaningful inferences about the quality of a firm's offerings.

Variants of our proposed lockout are seen in many market contexts, albeit in different forms. For example, "invite-only services," such as the ones introduced by Spotify when it first entered the U.S. market, or apps like "Clubhouse," where users can access certain parts of the app only via an invitation (Sisario, 2011). This practice "locks out" a segment of consumers that would have potentially demanded the product if available. In another example, some retailers use a "pop-up" type of store format that serves certain areas only temporarily, despite sufficient demand to support a year-round store (Schneider 2018).[3] Another somewhat indirect illustration is the recent rise of corporate activism (Chatterji and Toffel 2018), where CEOs publicly express their viewpoints on controversial topics, such as climate change, gay and transgender rights, gun violence, and immigration rights. Such moves could implicitly represent examples of demand lockout. By taking a stand on a hot-button issue, a firm might effectively alienate a significant portion of its potential customer base, at least temporarily, thus locking them out. Many other instances of demand lockout can similarly be shown to have some relevance to the mechanism we lay out here. However, we hasten to add that other potential

---

[2] This strategy is different from classic versioning or segmentation strategies, where consumers self-select not to be served based on firms' menu offerings. In other words, in contrast to demand lockout, under versioning all the options are available to all the segments albeit at different prices.

[3] In the labor market, we see some people using visible tattoos to lock out demand from potential employers, such as banks (French et al. 2018), and to signal attributes, such as creativity, to the remaining employers.

explanations must be carefully considered based on their specific context. For example, firm activism could be driven by the personal convictions of a CEO, divorced from any demand-side mechanisms.

In this paper, we use data from the motion picture industry to present empirical patterns consistent with our proposed signaling mechanism. The movie context has a specific demand lockout signal that movie producers can use: getting an R rating and thus locking out a large segment of non-adult moviegoers. Our empirical strategy involves generating a list of movies that are "close" to being rated R that ultimately end up receiving either an R or a PG-13 rating. Intuitively, we rely on the fact that studios have some control over what MPAA rating they want to target, even after controlling for the core content of a movie.[4] Such demand lockout via R rating is credible because the action is profitable only if consumers generate positive word of mouth. Thus, as per our theory, only high-quality movies have an incentive to lock out demand, all else being equal.

The obvious difficulty in the empirical application is that we observe only a discrete outcome: whether a movie is R-rated or not. The empirical challenge lies in figuring out which movies are "close to being rated R." To overcome the inherent difference between R and PG-13 movies in terms of "content inappropriateness," we use a machine learning approach that leverages text data from the movie subtitles to help us identify titles where such signaling might be feasible[5]. Finally, we match R-rated movies to PG-13 movies similar in "appropriateness score" and then run our analyses with a host of controls. Our empirical results are broadly consistent with the proposed theory and show that demand lockout still has a significant managerial impact above and beyond other relevant factors, such as advertising and third-party information.

[4] One company, Film Rating Advisors, Inc., advises movie studios on how to handle the MPAA rating process (Bernstein 2014).

[5] Many movies are unambiguously going to be rated R (e.g., a film with a sexual theme) or PG-13 (a movie for kids), based on the core content. We propose that our purported signaling mechanism will be meaningfully observed within a sample of movies that, based on the script and storyline, could have received either rating-i.e., the movies on the margin of a latent rating boundary.

Related Literature

Since Akerlof (1970), it is well understood that incomplete information between transacting parties can lead to a market failure, and sometimes an informed party has an incentive to undertake costly actions that could credibly convey information to the uninformed party. An extensive theoretical literature starting with Spence (1973) has spawned signaling as a way to understand how firms and consumers communicate payoff-relevant private information. Generally, signals of quality are considered effective if they are costly enough so that only a firm offering high quality will use them. For example, “uninformative advertising” can be a credible signal of unobserved quality (Milgrom and Roberts 1986) as long as such an investment is profitable for the high-quality firm (because of repeat purchases) but not for the low-quality firm. Similarly, prices can signal quality because they distort demand profitably only for the high-quality firm (e.g., Schmalensee 1978). In addition to advertising, researchers have identified several other marketing-related signaling mechanisms such as branding (Wernerfelt 1988; Erdem and Swait 1998), money-back guarantees (Moorthy and Srinivasan 1995), specialization (Kalra and Li 2008), and the reputation of the retailer (Chu and Chu 1994). Regarding the specific signaling mechanism, our paper shares similarities with the work positing scarcity (e.g., Sapra et al. 2010) as a quality signal. Our paper complements this strand of studies by analyzing a related yet novel mechanism of demand lockout. The research closest in spirit to our paper is Stock and Balachander's (2005) model of scarcity which assumes the informational expertise of some of the early buyers wherein high-quality firms could use the scarcity as a signaling tool to separate from the low-quality firms. In contrast to their model, we assume that quality uncertainty exists initially for all the buyers and could potentially get resolved through word-of-mouth (Ashoori et al. 2020) and via expert reviews. Additionally, we endogenize the choice of consuming in the first or second period and allow for uncertainty about quality for both the firm and consumers. The mechanism we propose also differs because it does not rely on a fixed quantity being available to both innovators and imitators, but rather a specific segment of consumers not being served. Thus we feel that our approach is more appropriate for experience goods and digital goods where the notion of a fixed quantity available either does not exist or cannot be observed by consumers. On the other hand, Stock and Balachander's model is more suitable for settings where the quantity of physical goods is limited, demand is publicly observed, and thus scarcity could be a helpful signaling device. The products like automobiles and cellular phones are example product categories where scarcity could serve as a credible signal, as has been demonstrated via later empirical work (Balachander et al. 2009). Furthermore, similar to Joshi and Musalem (2020), our model is inspired by situations where firms cannot use price as a signal. Finally, our paper is also related to Miklos-Thal and Zhang's (2013) work, which shows that "demarketing" - defined as the use of a lower level of marketing relative to a higher level that expands consumer consideration- could serve as a mechanism to convey quality. Their model results in the "pooling" of high and low-quality firms while we show that high-quality firms can credibly separate from low-quality firms via demand lockout[6].

Within marketing, our study is situated within the broad domain of scholarship that has developed endogenous signaling mechanisms. Allen (1984) investigates the use of high prices (along with quantities) to convey a reputation for high quality, wherein the deviation from high price results in consumers inferring the product is of lower quality. The paper does not explicitly model word-of-mouth and assumes that once a set of consumers use a product, its quality becomes precisely known in the marketplace. Similarly, Jiang and Yang (2017) study how consumer information sharing affects a firm's quality and pricing choices when the price can signal quality. A key feature of our model is the explicit incorporation of word-of-mouth, which relates to the earlier work on observational learning.

Vettas (1997) was one of the first papers to incorporate word-of-mouth within the durable goods setting. Similar to our paper, the number of adopters in this framework affects the information diffusion of quality. However, the article mainly focuses on the informational role of quantities (for quality inferences) and, as such, does not consider other signaling mechanisms (advertising or demand lockout). Similar to our paper, when advertising is

[6] Miklos-Thal and Zhang's results rely upon product inspection in their model. In contrast, we bring in consumer heterogeneity in quality valuation as well as other firm communication mechanisms like advertising and word-of-mouth. Finally, their demarketing does not actively turn away a subset of consumers; instead, it reduces the consideration set, unlike the discrete lockout examined in our paper.

assumed to be costly, Mayzlin and Shin (2011) study the signaling role of message content and show that a high-quality firm could produce messages devoid of product attribute information to nudge consumers toward costly searches. Relatedly, Chakraborty and Harbough (2014) show that cheap talk can be credible when the seller has to decide which attribute to emphasize, and such “puffery” can improve demand. Our paper also includes uninformative advertising and shows that when it is costly, other mechanisms might have particular relevance. Furthermore, our signaling mechanism might have special significance in many online platform settings where prices are often fixed and thus cannot be employed for signaling. However, Xu and Dukes (2022) demonstrate that sometimes in such environments where a firm might have an informational advantage over consumers about their valuations (via data aggregation), a combination of list and personalized pricing could serve as a signal and allow firms to price discriminate.

Our work also relates to the literature on conspicuous consumption wherein the exclusivity itself is valuable to consumers, and thus firms have an incentive to reduce availability for consumers’ desire for uniqueness (Pesendorfer 1995; Rao and Schaefer 2013). Within this domain, researchers have studied issues like pricing (Amaldoss and Jain 2005) and word-of-mouth. In contrast to this literature, the consumers in our setting have no value for exclusivity per se. and yet, exclusivity could be used by consumers to make (rational) inferences about the product quality.

While both marketing and economics have a rich theoretical apparatus and a variety of contexts where signaling has been used to study the various marketing policies of interest, the empirical studies in the area are somewhat sparse. Not surprisingly, the empirical tests for quality signaling can be challenging as signals, by definition, try to communicate unobserved quality. Thus, our study fits within a small but emerging literature that attempts to provide data support for firms’ signaling activities (see Sahni and Nair 2020 for a recent example).

In terms of empirical methodology, our work builds on a recent stream of papers that uses text data and machine learning in the context of causal inference and theory testing. Hoberg and Phillips (2016) use text data from 10-K filings to measure product similarity between firms and to test the endogenous product differentiation theory. In marketing, Netzer et al. (2012) obtain insights about market structures using consumer reviews data, while Timoshenko and Hauser (2019) employ user-generated content to identify customer needs. Gentzkow, Kelly, and Taddy (2019) and Hartmann et al. (2019) provide comprehensive introductions to using text data in economics and marketing, respectively. Finally, our work also responds to the recent calls by academics to bring a “soul” to machine learning by synergizing the predictive power of these tools with theory (Proserpio et al. 2020).

The rest of the paper proceeds as follows. First, we describe our context and present a number of analyses that provide us with a series of empirical patterns . Then, we present a stylized theoretical model that formalizes the use of demand locking as a credible signal and is consistent with the empirical observations. We follow this by analyzing how this signal compares with an obvious alternative—namely, advertising—and how it is moderated by third-party information. We then provide a discussion of some alternative rationales for our results, wherein we present tests addressing these explanations. We conclude by discussing the limitations and managerial implications of our findings.

Empirical Application

We use the motion picture industry as our empirical context to study the implication of our theory. This context is ideal for our proposed theory for several reasons: (1) Prices do not vary across movies of different qualities; (2) quality is revealed ex-post the consumption—the definition of a classic experience good; and (3) repeat purchases for a given movie are relatively low. We can also observe both some quality proxies and revenue numbers over time. Most importantly, we also have a specific demand lockout signal that movie producers can use: getting an R rating and thus locking out a large segment of non-adult moviegoers. Obviously, movie producers do not give themselves an R rating, as these ratings are assigned by the Motion Picture Association of America (MPAA). However, movie producers have full control over their movie’s content, which greatly influences whether they get an R rating.

The main empirical challenge for applying the insights from our model into this context comes from the fact that firms, or movie studios in our case, design different products targeted for different segments. Thus, even if we could observe quality directly, providing correlational evidence that R-rated movies have higher quality would not be sufficient to identify strategic signaling behavior.[7] The reason is that R-rated movies differ in content from non-R-rated movies, regardless of any intended strategic signaling. For example, consider a war movie. Viewers interested in such a movie are likely to prefer a more realistic depiction of war, and given that such a depiction necessarily involves violence, the R-rated content is simply necessary for a high-quality product. To identify signaling, we need to disentangle from the signaling effect the inherent quality differences resulting from differences in content. We outline the stylized simultaneity problem in Fig. 1a. We assume that movie quality has two dimensions: The first dimension is realism, inappropriateness, nudity, and other factors that lead to an R rating. The second dimension comprises all the other elements that cause consumers to enjoy a movie more, including the movie soundtrack, originality, acting, pace, and script quality. To solve the simultaneity problem, we use movie subtitles and apply machine learning models to recover the latent distribution of "inappropriateness." We refer to this retrieved index as the "R-level," a continuous variable between zero and one. A higher R-level corresponds to a more "inappropriate" movie because it is more likely to receive an R rating. This latent score allows us to compare movies similar in "appropriateness" and capture the signaling effect on the other dimension. As depicted in Fig. 1b, we leverage the fact that an R-rating does not directly affect quality when comparing two movies with an identical R-level. Returning to the example of the war movie, we can now compare two movies that depict similar levels of violence and realism but differ in their MPAA designation.

Institutional Details on Movie Ratings: Movie ratings in the United States are provided by the Classification and Ratings Administration (CARA) – an independent arm of the Motion Picture Association (MPA). A panel of 10 anonymous raters rates the movies into five designations: G, PG, PG-13, R, and NC-17. The raters are full or part-time MPA employees and must have a child aged between 5 and 15 years (Whitten 2022). The raters come from diverse backgrounds, serve for about 7 years (or until their youngest child reaches age 21), and primarily look for sex, nudity, and language in the movies. The panel reviews about 700 movies a year, and the main

[7] A simple comparison of means in our dataset shows that the average critics' rating of R-rated movies is 8 percentage points higher than the average rating for PG-13 movies.

aim of these ratings is to guide parents if the movies are suitable for children. The ratings are based on a simple majority voting, and at least 5-panel members must have seen a movie for it to be rated.

The criteria for constructing ratings are subjective (e.g., how much nudity should lead to an R-rating) and algorithmic (e.g., if F-word is used as an expletive more than once, the movie automatically gets an R-rating). While CARA indeed assigns the rating, producers ultimately have control over the decisions regarding a movie's visual and linguistic content. Thus, movie producers understand how ratings work as they know numerous past decisions and CARA advisories.

If movie producers are unhappy with a rating, there is an appeal process for changing the rating, or a movie can be resubmitted after an appropriate re-editing to get the desired rating (Whitten 2022). Many high-profile cases show that the producers in the past have reacted to rating decisions and, in many cases, had successfully changed the content to get a different rating. Producers of the extreme horror movie *Infinity Pool* were not happy with the NC-17 designation and did extensive re-editing (with the help of an external consultant) and ultimately scored an R-rating from CARA. The movie's producer, Brandon Cronenberg, put it aptly, "It's always fixable because you can always cut things." (Jacobs 2023). There are instances wherein an appeal (without any editing) could result in a change in rating. For example, in 2013, the producers of the movie *Philomena* successfully appealed to MPPA to change the rating from "R" to "PG-13" (Pulver 2013). These and numerous other examples suggest that ratings are malleable (to a certain extent) from the producers' perspective.

But could movies with somewhat similar content obtain different ratings? This, perhaps, is the most pertinent assumption within our empirical setting. Again, anecdotal observation suggests many instances of movies with similar content receiving different ratings, presumably due to minor adjustments by the producers. For example, as discussed by Clyde (2014), the movie *The Fall* (2006) and the movie *Big Fish* (2003) were very similar in thematic, visual, language, and storytelling perspectives. Yet, the former received an R-rating while the latter got a PG-13. Clyde (2014) provides numerous other such examples.

Again, it is worth reiterating, as we readily accept throughout, that the R-rating of a movie often allows producers to incorporate elements essential to the movie's storyline. However, there are also instances where such a rating could play a signaling role wherein a film could receive either designation (R vs. PG-13) without significant change in the content or quality.

Data: Our dataset is built from a variety of sources. First, we collected a large corpus of English subtitles data from *OPUS* (Tiedemann 2012). This dataset contains subtitles in over 60 languages for thousands of TV shows and movies. We used the subset of movies with subtitles in English (26,075 documents). We also used a dataset from *IMDb*, an online movie database comprising metadata on 5,045 movies. Of these movies, 4,089 titles are in English and have an MPAA rating of PG, PG-13, or R. We also collected data on critic ratings from *Metacritic*, a movie rating aggregator. We collected 238,106 reviews for 13,218 movies. We obtained user reviews of movies from *MovieLens* (Harper and Konstan 2016), a movie recommendation service that contains hundreds of thousands of user reviews and more than 25 million reviews on a 1- 5 scale. The dataset contains over 60,000 movies released between 1995 and 2019. We collected box office revenues from *The Numbers*, an aggregator of movie data. We restricted our sample to the first five weeks of each movie because the information asymmetry issues will likely be most important during the initial weeks.[8] Finally, we use detailed syndicated advertising data from *Kantar Media* for all movies in the dataset. Our final dataset of 1,502 English movies that have an MPAA rating of PG-13, PG, or R that were screened in the United States between 1998 to 2013; where had complete information available on their box-office revenues, critic reviews, consumer reviews, advertising data, and subtitles data.[9] See Table 1 for the summary statistics for all relevant variables.

*Intuition***:** Before proceeding to the primary analysis, we outline the intuition of our empirical strategy. In a nutshell, we are proposing that by making a movie unavailable to a segment of consumers under the age of eighteen, the adult consumers could make the (rational) inference that the movie must be of high quality. Finding empirical evidence for this would be easy if we

[8] 1,428 movies (95% of the sample) were in movie theaters for at least the first 5 weeks.
[9] The sample size of 1,502 movies covers the majority of all releases during the sample period. For example, Brown et al. (2012) used a total of 1,414 movies widely released in in the US between 2000-2009.

could observe unobserved quality and the exact level of "inappropriateness" of a movie. However, such exact measures are obviously unavailable. However, we will consider reasonable proxies for unobserved quality, and carefully construct empirical estimate of the level of inappropriateness. This allows us to present empirical patterns largely consistent with our signaling framework. If we only observe that R-rated movies have a higher level of quality, we cannot convincingly conclude that the rating serves as a quality signal. It might simply be that movies with an R rating are of higher (unobserved) quality because consumers prefer more realistic movies. Fig. 2 depicts a stylized intuition of the equilibrium if a sharp cutoff in "inappropriate content" exists between movies rated R and rated PG-13. Only movies sufficiently close to the cutoff (i.e., the dotted box) can credibly use the signal. The identification argument and empirical analysis formalize the intuition of Fig. 2 rigorously.

To identify the quality dimension unrelated to the R-level, we need to condition on the level of "inappropriateness." To solve the problem, we introduce a latent variable that captures the probability of a movie being rated R. Using text data from subtitles, we estimate the R-level as:

$$e(W_i) = Pr(R_i = 1|W_i) \tag{9}$$

The propensity score $e(W_i)$ is a function of the high-dimensional vector of words (W_i) from the subtitle file. Given the propensity score, treatment assignment and the observed covariates are conditionally independent (Rosenbaum and Rubin 1983). That is: $W_i \perp R \,|e(W_i)$ (10)

Using the subtitles data, we match each R-rated movie based on its estimated propensity score to a movie that is statistically indistinguishable from this score but is rated PG-13. Note that many R-rated movies are not close to any PG-13 movie on this dimension; thus, any strategic signaling motive for the R rating is unlikely.[10] Upon identifying films that lie outside the common support, we can categorize movies into one of three categories: *Always R*, *Never R*, and *Maybe R.* Movies in the *Always R* or *Never R* category are sufficiently high or low in terms of inappropriate content, such that their MPAA rating is determined based on the movie's

[10] Consider, for example, the 2013 movie, *The Wolf of Wall Street*. which used the f-word 569 times. Typically, movies with more than two uses of the word receive an R rating. This movie clearly does not have a proxy PG-13. In our analysis, we use both PG-13 movies and PG movies. We refer to them collectively as PG-13 movies throughout the paper. Movie Studios often try to get their MPAA ratings changed in either direction (Hicks, 2013)

artistic needs or storyline. We thus exclude all movies that lie outside the common support from the analysis and only include the *Maybe R* movies, for which a signaling mechanism is feasible. For these movies, we make the following identification assumption:

Assumption 1*: Let* $S\big(X_i, q_i | e(W_i)\big) \in \{0,1\}$ *be the choice of signal for a movie with observable characteristics* X_i *and quality* q_i*. Then* $1 > \Pr(R = 1 | e(W_i), S(X_i, q_i | e(W_i)) = 1) > Pr(R = 1 \mid e(W_i), S(X_i, q_i | e(W_i)) = 0) > 0$*.*

The above assumption formalizes the notion that movie studios have some "wiggle room" and potentially can influence the rating process to use an R rating as a signaling mechanism, while there remains some randomness. The main goal of the ensuing econometric analysis is to test whether the choice of the signal is consistent with a signaling theory. (e.g., is the R rating more likely for higher-quality movies $\frac{\partial\, \boldsymbol{S}\left(\boldsymbol{X_i}, \boldsymbol{q_i} | \boldsymbol{e}(\boldsymbol{W_i})\right)}{\partial q_i} > 0$). Next, we describe in detail how we recover the latent distribution of the R-level and build a sample based on propensity score-matched data.

Estimation of the R rating: To construct a continuous measure of a movie's "inappropriateness," we leverage a large set of subtitle text data, which we map onto MPAA ratings. Consider the outcome variable to be the MPAA rating, represented as a binary outcome $R_i = \{0,1\}$, where 1 corresponds to movies rated R and 0 to movies rated PG-13 or PG. We start with a set of i subtitle files $\{\mathfrak{D}_i\}$ consisting of raw text. Let W_i be a numerical representation of the words used in the subtitles docs and outline the mapping $\{\mathfrak{D}_i\} \rightarrow W_i$.

To do so, we take the following steps.

1. Cleaning and standardizing the words data
2. Create n-gram
3. Create term frequency-inverse document frequency matrix

Then, using the numerical representation W_i, we use different machine learning models (lasso logistic regression, the elastic net, the random forest, and support vector regression) to predict R_i for each movie[11]. We further create an ensemble benchmark, combining the models. Past

[11] Online Appendix T6 provides detailed descriptions of the models and the data cleaning process

work has shown that even simple ensemble methods are more accurate than individual models (Mullainathan and Spiess 2017; Grimmer et al. 2017). Beyond its accuracy, we believe the ensemble also is more robust in recovering the rank order of the subtitle content (Athey and Imbens 2019).

To do so, we create an average of the predictions from the k models, weighted by the inverse root mean square error: $E_k = \sqrt{\sum_{i=1}^{n} \frac{(\widehat{y_{ik}} - y_i)^2}{n}}$. (15)

Using RMSE inverse as weighting, the average estimate for movie i is given by:

$$\overline{y_i} = \frac{\sum_{k=1}^{4}(RMSE^{-1}\,\widehat{y_{k_i}})}{\sum_{k=1}^{4} RMSE^{-1}}. \tag{16}$$

We now compare the estimates from the four models and the weighted average of all models. The average accuracies, as measured by the hit rate of the predictions, of the lasso logistic regression, the elastic net, the random forest, and the SVR are 84.9%, 83.6%, 87.6%, and 76.8%, respectively. The ensemble prediction has an accuracy of 87.46%. The receiver operating characteristics curves (ROC) with the areas under the curve (AUC) for different models are displayed in Fig. 3. The models perform exceptionally well in predicting R ratings. AUC is over 0.9 for the random forest model, as well as for the logistic regression models and the ensemble method. We compare the average prediction to each model and find that no model performs strictly better in terms of the average ensemble prediction (87.46%) and the AUC (0.946). We thus use the weighted average of the estimates for the analysis in the next step. Finally, Fig. T4 in the online appendix T4 displays the distributions of the estimated values for R-rated and PG-13 movies. The distributions have intuitive appeal. The mass of movies rated PG-13 have estimated values around 0.3–0.4 in all models, while it is significantly above 0.5 for R-rated films. Additionally, words that have a high predictive value (see Web Appendix T6) are capturing both inappropriate language via curse words, as well as audio signs of inappropriate visual content, such as the words “bed” or “murder”.

After calculating the probability of each movie being rated R, we use the estimated value as the propensity score from the text data W_i to match each R-rated movie to the most similar PG-13 movies (Dehejia and Wahba 2002). To do so, we first exclude movies outside the common support by excluding R-rated movies for which the probability of being rated R is so high that no PG-13 movie has an equivalent probability. For example, movies such as *8 Mile* or *The Wolf of Wall Street* lie outside the common support and are not included in the propensity score-matched dataset. Of the 663 R-rated observations, we exclude 145 that lie outside the common support. After restricting the sample to observations within the common support, we follow the nearest neighbor matching approach with three matches per observation. For every R-rated movie, we pick the three PG-13 observations with the most similar estimated R-value per the ensemble method described. We set a caliper of 0.05, excluding the closest matches with a difference in the propensity score of more than 0.05. After matching, we assign weights to each matched PG-13 observation to create a balanced sample, where $\#matches$ is three whenever we can find three observations within the caliper and is less than three if only two movies or one movie rated PG-13 exists within the caliper. We estimate the propensity score with replacement. Reassuringly, the mean of the estimated propensity score in the matched sample is 0.69 and 0.7 for the R-rated and PG-13–rated groups, respectively, indicating that we have created a reasonably balanced sample[12].

Main Analysis

We will formalize the signaling model later, but to facilitate the empirical analysis, consider the following intuition. If there exists a signaling mechanism in which only high-quality movies are rated R, we should observe distinct behavior in three qualitatively different stages. In the initial stage, a movie studio decides on the signal (i.e. R versus PG-13). In the subsequent stage, an initial cohort of consumers consumes the product based on the observed signals and the noisy expert reviews. In the final stage, true quality is revealed, and the subsequent cohorts of consumers buy based on the revealed quality and the intensity of word of mouth. We now analyze each of these stages separately, starting with the decision to signal. It is important to

[12] The reasonable estimation of the R-level is critical for our identification strategy and could potentially introduce bias in our analysis, it thus deserves additional scrutiny and robustness checks. We check the ratings from the different models for internal consistency and use expert reviews (from *Common Sense Media*) to validate the predictions' external validity. To save space, we present these analyses in the Web Appendix T3. Overall, our estimates have reasonable internal and external consistencies.

emphasize that these results are presented as empirical observations that are consistent with the theoretical model we subsequently describe, as opposed to causal effects.

Decision to signal: First, we consider the strategic decision by the firm to choose the R rating as a signal based on information available to the movie studio before the movie's release[13]. We assume that the decision to signal $\mathrm{S}\big(\mathrm{X_i}, \mathrm{q_i} | \mathrm{e}(\mathrm{W_i})\big) \in \{0,1\}$ to be a function of quality ($\mathrm{q_i}$) and movie characteristics ($\mathrm{X_i}$). In our empirical context, we can consider the effect of quality on the decision to signal, the effect of other signals (i.e., advertising), and the effect of third-party information. We estimate a model of the propensity to use the signal: the propensity of a movie being rated R. We use the sample of matched movies and use the weights assigned from the propensity score matching algorithm in the weighted logistic regression below:

$$\boldsymbol{P(R_i = 1)} = \frac{\boldsymbol{exp}\{x_i'\beta + genre_i'\ \beta + year_i'\ \beta\}}{\mathbf{1} + \boldsymbol{exp}\{x_i'\beta + genre_i'\ \beta + year_i'\ \beta\}}, \tag{17}$$

where $x_i{}'$ includes the mean, standard deviation, and count of the critics' reviews, the movie budget, the amount spent on advertising before release, and indicator variables for the movie being from a major studio or foreign. Also, $genre_i$ and $year_i$ are dummy variables for the genre and year of release, respectively. The results of this model are presented in Table 2. It is worth recalling that we have used a sample of matched movies, comparing movies similar in content (in terms of inappropriateness) rated R or PG-13.

If the R-rating is a signal of quality, we would expect movies with higher quality to be more likely to be rated R. Although the actual quality being signaled is unobservable, we can reasonably assume it positively correlates with the critics' review score. The coefficient of critics' reviews is positive ($p < 0.05$), indicating that movies with a higher critic score are more likely to be rated R.

[13] Note that consumer reviews are not included in this specification because these typically arrive after the decision about R-rating has been made. Critic reviews are an ex-post measure of quality but studios generally have access to the expected review quality at the time of choosing a signal (Brown et al. 2012). By using critic reviews as a proxy for quality, we assume that critic reviews affect demand only through the informational effect of (partially) revealing quality.

Critics' scores are available to consumers when making their consumption decision, serving as third-party information. Because alternative information might crowd out the returns to signaling quality, the value of the signal decreases as the third-party information accuracy increases[14]. Ideally, we could estimate the accuracy of critics' reviews using a measure of the true quality. However, since the true quality is latent, we instead assume that critics' reviews are noisy but unbiased draws from the actual quality distribution. Therefore, additional reviews for a movie will increase the accuracy of third-party information. Similarly, a lower standard deviation of the critics' reviews increases the accuracy of these reviews, as it reduces the noise of the measure and allows for a more precise inference about the expected quality. The coefficient on the critics' standard deviation is positive ($p < 0.01$), and the coefficient on the critics count is negative (but not significant). Note that a higher standard deviation implies more uncertainty leading to greater usefulness of the signaling. Similarly, a higher count suggests more information obviating the need for signaling. Similarly to critics reviews, advertising can serve as an alternative signal and can provide information that crowds out returns to signaling. The coefficient on advertising is negative but not significant. The effect of the budget, another possible signal, is negative ($p < 0.01$), implying that movies with a larger budget are less likely to seek out an R rating. Consumers observe the movie studio associated with a particular movie, which can also signal quality. How being produced by major, foreign, or domestic minor studios signals quality relative to the other categories is not obvious a priori, and the results are somewhat ambiguous.

Opening Week Revenue: Next, we consider the heterogeneous returns to signaling during the opening week. Absent any signaling motive, R rating only contracts demand, which implies lower revenues. Leveraging the remaining noise in the rating process, we estimate the heterogeneous difference in revenue between movies rated R and rated PG-13. We estimate a linear model using the propensity score-matched data to test the signaling. Similar to the model for the decision to signal, we use the sample of matched movies and use the weights assigned from the propensity score matching algorithm in the weighted regression:

[14] We derive this, as well as other propositions formally in the latter part of the paper.

$$\log(revenue)_i = x_i'\beta + genre_i'\,\beta + year_i'\,\beta + \epsilon_i \tag{18}$$

where $x_i{}'$ includes the R rating, mean, standard deviation, and count of the critics' reviews, as well as the log of the budget and the log of advertising before the movie release. In addition, we include the genre and year dummies. The results are in Table 3.

Signaling implies that only high-quality firms can profitably lockout demand. In the empirical setting, this implies that returns for quality are higher for movies rated R compared to PG-13 movies. The interaction term $R \times Critics\ Mean$ is positive ($p < 0.01$), supporting this prediction[15]. Recall that the signal is most useful when the quality uncertainty is highest ($\frac{\partial \Delta \Pi_{\mathrm{h}}}{\partial \gamma} \leq 0$). Critics' information reduces this uncertainty and decreases the returns to signaling. To test this prediction, consider the heterogeneous effects of critics' standard deviations (more uncertain information) and critics' review count (more information), with and without a signal. The coefficient on the interaction $R \times Critics'\ standard\ deviation$ is positive ($p < 0.01$), and the coefficient on $R \times Critics'\ count$ is negative ($p < 0.05$), giving support to the prediction. The results for advertising spending and the budget are somewhat ambiguous. The interacted effect of the $R \times Budget$ is marginally significant and positive ($p < 0.1$), which seems consistent with the signaling theory if the budget is correlated with unobserved quality. Finally, return to the signal are highest for domestic movies, for movies not produced by a major studio, and significantly lower for major studio productions ($p < 0.01$) and foreign productions ($p < 0.01$), consistent with the signaling theory if uncertainty is highest with domestic, non-major studio movies.

Revenue after the opening week: We next fit a model for weeks 2–5. We assume that viewers observe (close to) the movie's actual quality in week one and generate word of mouth. Controlling for weekly advertising spending and budget size, the intensity of word of mouth increases in the number of people who watched the movie in the previous periods. Note that the revenue in week 1 is a function of all observables at the time $\Pi_{i1}(x_i)$. If the quality is precise as predicted from prior information x_i, the subsequent revenue will be a function only

[15] We find that the overall effect of critic reviews is negative, which is consistent with some previous empirical evidence (Moon et al. 2010)

of previous revenue. That is, $\Pi_{it} = f(\Pi_{i,t-1})\ \forall\ t > 1$. However, true quality affects the valence and intensity of word of mouth. We thus define $\Pi_{it} = f(\pi_{it-1}, \bar{q}, R)$, where $\bar{q}$ is defined as $\bar{q_i} = q_i - E[q|\pi_{it-1}]$ to capture the quality shock experienced by consumers in the previous period. Because quality is not directly observable, we define $\bar{q_i} = PR_i^U - PR_i^C$, where PR_i^U is the percentile rank of the consumer score (from the *MovieLens* dataset) for movie *i* and PR_i^C is the percentile rank of the critics' reviews (from the *Metacritic* dataset) for movie i. The advantage of this measure is that it uses only the rank order of movies and does not rely on the shape of the distributions of critics and consumer scores. A movie that received a high ranking from critics but a low score from consumers thus has a negative value of $\bar{q_i}$. In other words, the revealed quality was lower than the expected quality. Finally, we include the count of consumer reviews to measure the intensity of WOM consumers observe.

We estimate the following revenue model for the weeks after opening week[16]:

$$\Pi_{it} = \Pi_{it-1} \times (x_{it}'\beta + genre_i'\ \beta + year_i'\ \beta) + \epsilon_{it}, \tag{19}$$

where x_i includes the R rating, weekly advertising spending, the quality shock $\bar{q}$, consumer ratings, the number of consumer reviews, and the other dummies used in the previous regressions.

The results are in Table 4. The signaling model implies that revenue is higher for R-rated movies if, and only if, they are of high quality[17]. The coefficient on $R \times Consumer\ review\ mean$ is positive (p<0.1) and implies that movies with a consumer review mean above 3.1 had higher revenues when they were rated R vis-à-vis their PG-13 counterparts, evaluated at the median number of reviews. Note that the median rating is 3.2, supporting the theory that only high-quality movies benefit from the lockout mechanism. Finally, if the signal communicates information not contained by the third-party information, then errors in the critics' scores should already be anticipated by consumers. Testing this prediction, we find that quality shocks have a positive coefficient for movies not rated R. Still,

[16] We experimented with week-specific parameters and found that, as expected, effects are strongest for week 2. The estimated coefficients can be interpreted as the average across the weeks.
[17] The unconditional effect of the R-rating is positive, but not significant.

the effect for R-rated movies is negative ($\beta =$ -0.23) and significant (p<0.01), providing support for the prediction that the signal reduces information asymmetry. We also find support for the hypothesis that profit in subsequent weeks depends on the intensity of WOM, and the interacted coefficient $R\ \times\ \#\ of\ Consumer\ Reviews$ is positive ($\beta = 0.038$) significant (p<0.05).

Additional Analyses: We present two sets of additional analyses in the Web Appendix (T4) that test the robustness of these observations. In the first analysis, we consider different ways to construct the sample. Recall that in the analysis presented so far, we have used the three closest PG-13 movies to an R-rated movie using the nearest neighbor matching approach using their propensity scores. We re-estimate the three empirical models (i.e., eq. 17,18 and 19) using nearest neighbor matching with different potential neighbors (1,5 and 10). Next, we consider estimation where instead of picking up neighbors, all the observations are used in the analysis using Inverse Propensity Score (IPS) weighting (Hirano et al. 2003), with the PG-13 movies closer to an R-rated movie getting higher weight. We construct the weight for each movie as the inverse of the propensity to have received that rating. This estimator is attractive because being rated R or PG-13 is independent of the propensity we derived from the subtitle text. However, while the estimator does give R-rated movies with a high R-level a low weight and PG-13 movies with a low R-level a low weight, it does not exclude movies outside the common support, which could potentially affect results. Reassuringly, our results are not particularly sensitive to the choice of the matching algorithm, and we find the key results generally hold in all specifications. Secondly, while we do not have any ex-ante theory about the effect of genres, the proposed signaling mechanism could presumably be more likely in specific genres. For example, in genres with a higher overall propensity of R-rated movies (e.g., Horror), an R-rating might presumably play a different role than in genres with more PG-13 movies (e.g., Comedy). We investigate this by considering the analysis for each genre separately. To do so, we chose the ten most popular genre tags and re-estimated the full model separately for each tag. The sample size is relatively small, and there is insufficient power to make strong inferences. However, we find that there is a positive relationship between R-rating and quality for Drama, Thriller, Comedy (n.s), Crime, Romance, and Mystery (n.s). Interestingly, there seems to be little evidence of our purported signaling mechanism for movies in the horror genre, as having an R-rating as perhaps horror movies are generally expected to be rated R. See the full results in tables (T3, T4, and T5)[18].

So far, we have presented a number of empirical observations. Because they do not rely on exogenous shocks, it is difficult to treat them as causal effects and great care should be taken in evaluating the coefficients and not treating them as the effect of a counterfactual change in R-rating for a given movie. We now present a formal signaling model that aims to rationalize the observed patterns.

Model

Before presenting the formal model, consider the following scenario: A firm could be endowed with a high-quality product (say, a movie) or a low-quality product. A high-quality (low-quality) product has a greater (lower) likelihood of success in the marketplace. We assume that firms differ in their probability of realizing a high-quality (low-quality) product and term a firm with a higher probability of a high-quality (low-quality) product as a high-type (low-type) firm. We assume that nature randomly draws the firm type, and after such a draw, a firm knows the expected quality of its offering and, thus, its own type. For example, a movie studio, after the completion of a movie, might discover- say, via extensive internal marketing research with potential viewers- about the likelihood of its success. However, even when a product enters the marketplace, it could still turn out to be of low quality since the tools like marketing research and other information-gathering devices are imperfect. However, a product of *high* expected quality is *less likely* to be of low quality and vice-versa. Later, we relax the assumption that the firm is randomly endowed with a product – we allow the firm's investments to affect the quality of its offering.

Consider a firm that faces a market consisting of two consumer segments $i \in \{1,2\}$ across two time periods. We denote the sizes of segment 1 and segment 2 as λ^1 and λ^2, respectively. The consumer segments vary in their opportunity costs of consumption, as described in more detail

[18] Unfortunately we lack power to precisely estimate coefficients for most genres.

shortly. We normalize the total market size to 1, so the parameter λ^1 captures the proportion of consumers in segment 1[19]. As is standard in signaling models, we assume that nature exogenously selects the firm's type j, which, for tractability, is assumed to be discrete and defined via the set $J = \{l, h\}$, where l and h denote low- and high-type firms, respectively. The product can be of low ($\underline{q}$) or high ($\overline{q}$) quality. With probability α_j, the product is of high quality. The high-type firm has a higher probability of producing a high-quality product. We define $E[q_j] = \alpha_j\,\overline{q} + (1-\alpha_j)\underline{q}$. To simplify the analysis, we let $\alpha_h = (1-\alpha_l) = \alpha > \frac{1}{2}$. At the beginning of the game, nature reveals the firm quality as being high or low, and thus the probability α_j. The prior probability of the firm being high type is given by $\mu_0 \in (0,1)$. The size of customer segments and the prior distribution over consumer types are assumed to be common knowledge.

Consumers within each segment (rationally) decide to consume in the first or second period based on their opportunity cost. We denote consumers who endogenously choose to consume in the first period as *influentials* and consumers who consume in the second period as *followers*. Upon consumption, influentials observe the actual quality of a product and generate word-of-mouth within their respective segments. A fraction of the followers observes and update their beliefs based on word-of-mouth (WOM). Table 5 summarizes the key notations used in the model, and Fig. 4 shows the timeline.

Demand: Each consumer can buy, at most, one unit of the product, which can be either high or low quality. Consumer utility from the consumption of product of quality realization $q \in \{\underline{q}, \overline{q}\}$ is given by $u = q - p$, where q indexes the quality, and *p* denotes the product price. Consumers are heterogeneous in their opportunity cost of consumption, represented by k. We assume that this cost follows PDFs distributions $f^1(k)$ and $f^2(k)$, for segment 1 and segment 2, respectively. Consumers in segment 1 have a higher opportunity cost, in the sense of first-order stochastic dominance (FOSD), such that the respective CDFs satisfy $F^1(k) \leq F^2(k)\ \forall k$[20].

---

[19] Throughout the paper we use λ to denote λ^1 whenever there is no ambiguity. A high type firm refers to a firm in possession of a product of high (expected) quality.

[20] FOSD is not necessary for the equilibrium, but the equilibrium does not hold under symmetric segments.

Consumers have non-negative opportunity costs, bounded at the low level of quality ($F^i(\underline{k}) = 0 \ \forall \mathrm{i} \in \{1,2\}$ for $\underline{k} = \underline{q} - p$), and are willing to consume the high-quality product ($F^i(\bar{q} - p) = 1 \ \forall \mathrm{i} \in \{1,2\}$). To simplify the exposition, we normalize this upper bound such that $\bar{q} - p = 1$. We further assume that $F^1(k)$ and $F^2(k)$ are strictly increasing and continuous.

Demand Lockout: The firm operates in a market wherein it has access to a tool that it could deploy to make its product unavailable to one of the segments. As will become clear that within our context, this tool could be used as a signal, and we denote R as an instance of the firm's use of demand lockout and N as serving all consumers. We later relax this assumption and consider the situation where the firm can only imperfectly exclude a proportion of the segment.

Consumer's beliefs about the type of a firm are a function of the observed signal s and the given prior μ_0. We denote the consumer's posterior probability of a firm being high type

by: $\mu_C(\ s)$. The expected quality, conditional on the observed signal and the prior, is denoted by

$E[q|\mu_C(\ s)] = \mu_C(\ s)E[q_h] + \left(1 - \mu_C(\ s)\right)E[q_l]$, and expected utility is denoted by $E[u|\ \mu_C(s)] = E[q|\mu_C(s)] - p$.

The actual quality gets revealed post-consumption, and WOM is generated within each segment. We assume that each follower is connected to an exogenously given positive mass of influentials and observes WOM conditional on one of their connections having consumed the product. We capture these connections parsimoniously by assuming that a proportion of followers $\omega \times D_1^i$ in each segment observes WOM, which is a function of the proportion of consumption by influentials in segment i the first period: $D_1^i = \int_{\underline{k}}^{E[u|\ \mu_C(s)]} f^i(k)dk$. We assume $0 < \omega < 1$ and is common across the two segments and bounded to ensure that WOM is never fully observed within any segment. A higher ω implies more efficient WOM.

Consumers rationally choose to consume in the first or second period. We assume that the consumer discount rate is sufficiently steep such that they will only delay the purchase if their expected utility in the first period is negative. The firm's discount rate is fixed at 1[21].

Let $\sigma = (\sigma_F, \sigma_C)$ denote the strategies of the firm and consumers, respectively. $\sigma_F(s|j) \in [0,1]$ denotes the probability that the firm sends signal s, conditional on its type. The consumer strategy for a consumer with opportunity cost k in segment i is given by: $\sigma_C(k, i, s) \in c_1 \times c_2(w)$, where $c_1 \in \{0,1\}$, $c_2 \in \{0,1\}$ and $c_1 + c_2 \leq 1$. Where, c_1 (c_2) denotes the consumers' consumption decision in period 1 (period 2). The firm's action space is given by: $s \in \{R, N\}$, where R denotes demand lockout, and N denotes serving all consumers.

A consumer's expected utility from consumption in the first stage is given by: $\mathrm{E}[u_1] = E[u|\ \mu_C(\ s)]$. The expected utility for a consumer from segment i with opportunity cost k from consuming a product of type j during the second time period, conditional on receiving WOM, is given by:

$$\mathrm{E}[u_2|w = 1] = \beta \times \begin{cases} \bar{q} - p, for\ q = \bar{q} \\ \max\left\{\underline{q} - p, 0\right\}, for\ \ q = \underline{q} \end{cases},$$

Where $\beta \in [0,1]$ is defined as the discount rate. The consumer in segment i consumes in the first period if the utility is higher than in the second period and (weakly) higher than the opportunity cost. The first condition always holds trivially because of the assumption of sufficiently steep discounting, and the second condition holds whenever: $E[u|\ \mu_C(\ s)]\ \geq k$.

Thus, demand in the first period, conditional on quality beliefs induced by firm action s is given by: $D_{j1}^i = \int_{\underline{k}}^{E[u|\ \mu_C(\ s)]} f^i(k)dk$. In the second period, the consumers who did not buy in the first period remain in the market. These followers consume only if they observe a WOM and

---

[21] Formally, we let the consumer discount rate $\beta \to 0$. This assumption is made to facilitate tractability but generally seems applicable in contexts where consumers have high preference for early consumption like watching a movie on opening weekend, or buying a newly released version of a popular video game. Setting the consumers discount rate to be (infinitesimally) small allows us to parsimoniously capture the tradeoff between waiting for more information and consuming now. The intuition of the model holds, as long as consumers' discount rate is less than one. Note that we could easily allow for the firm's discount rate to be less than one. A decrease in the firm's discount rate is mathematically equivalent to a decrease in the rate of WOM ω. Intuitively, a movie goer is impatient to watch a movie while a studio is patient in ticket sales accruing across a span of few weeks.

their utility is positive: $(q_j - p) \geq k$[22]. Because consumers observe quality if they receive WOM, the demand in the second period from segment i is given by:

$$D_{j2}^{i} = \omega\, \alpha_j \int_{\underline{k}}^{E[u|\, \mu_C(s)]} f^{i}(k)dk \times \left(1 - \int_{\underline{k}}^{E[u|\, \mu_C(s)]} f^{i}(k)dk\right),$$

where the first three terms together imply the probability of observing WOM, and the final term measures the size of the consumers who have not yet consumed.

Analysis: We first consider the case without asymmetric information. Consumers observe the firm type and form (correct) expectations about product quality and consume accordingly in the first period. In the second period, WOM reveals the actual quality, and some consumers who waited consume the product if it is revealed to be of high quality. Profits for the two (firm) types are given by $\Pi_h = p \sum_{i \in \{1,2\}} \lambda^{i} \left(\int_{\underline{k}}^{E[u_h]} f^{i}(k)dk \left(1 + \alpha\, \omega \int_{E[u_h]}^{1} f^{i}(k)dk\right)\right)$ and $\Pi_l = p \sum_{i \in \{1,2\}} \lambda^{i} \left(\int_{\underline{k}}^{E[u_l]} f^{i}(k)dk \left(1 + (1-\alpha)\, \omega \int_{E[u_l]}^{1} f^{i}(k)dk\right)\right)$, respectively, where $E[u_h] = \alpha\, \bar{q} + (1-\alpha)\underline{q} - p$ and $E[u_l] = (1-\alpha)\bar{q} + \alpha\underline{q} - p$ denote the expected utility for the high and low types, respectively.

If the firm cannot communicate its type, consumers use the prior belief μ_0 about the firm type to form expectations about quality. Let $E[u_{\mu_0}] = \mu_0 E[u_h] + (1-\mu_0)E[u_l]$. Profits under uncertain firm type are given by:

$$\Pi_h = p \sum_{i \in \{1,2\}} \lambda^{i} \left(\int_{\underline{k}}^{E[u_{\mu_0}]} f^{i}(k)dk \left(1 + \alpha\, \omega \int_{E[u_{\mu_0}]}^{1} f^{i}(k)dk\right)\right), \text{ and}$$

$$\Pi_l = p \sum_{i \in \{1,2\}} \lambda^{i} \left(\int_{\underline{k}}^{E[u_{\mu_0}]} f^{i}(k)dk \left(1 + (1-\alpha)\, \omega \int_{E[u_{\mu_0}]}^{1} f^{i}(k)dk\right)\right).$$

Demand Lockout Signal: We now provide the details of our proposed signaling mechanism that could potentially allow firms to communicate their type. Suppose that firms can credibly

[22] Information provided by WOM can only be valuable for consumers if it can affect their actions. Because consumers with a low opportunity cost will consume regardless of the quality of the product, they will never wait to consume. At the same time, consumers with a higher opportunity cost want to avoid low quality products and thus have incentive to wait for WOM.

exclude the segment with lower opportunity cost (segment 2) from being served and make the product accessible only to segment 1. In the initial stage (period 0), after observing its type j, the firm decides whether to exclude segment 2 or to serve both segments. In the first stage of the game, when the product is released, all consumers observe whether the firm uses this exclusion strategy and update their quality beliefs. Consumers with sufficiently low opportunity cost (influentials) consume the product and generate WOM. In the final stage (period 2), upon the receipt of WOM, the followers make consumption decisions.

Consider a separating equilibrium in which the high-type firm's action is *lockout (R),* and the low-type firm's action is *no lockout (N).* Suppose this equilibrium fully reveals the asymmetric information, such that consumers' posterior beliefs are: $\mu_C(R) = 1$ and $\mu_C(N) = 0$. Suppose the high-type firm excludes segment 2 and restricts demand to segment 1. In that case, all consumers in segment 1 believe the firm is of the high type and consume the product as long as the expected product quality is higher than their opportunity cost. Profit for the high-type firm is given by:

$$\Pi_h|R \ = p\lambda^1 \left(\int_{\underline{k}}^{E[u_h]} f^1(k)dk \ \left(1 + \alpha \ \omega \ \int_{E[u_h]}^{1} f^1(k)dk\right)\right) \tag{1}$$

Note that the first term in (1) is the consumption by influencers (with relatively lower opportunity costs) within segment 1, and the second term is the consumption by followers within the same segment who consume if they receive positive WOM (with probability $\alpha\omega$).

If no lockout is used, then under the (assumed) separating equilibrium, consumers update their beliefs and consider the firm to be of the low type, and only consumers with a sufficiently low opportunity cost will consume. Profit for the low-type firm under no lockout is given by:

$$\Pi_l|N \ = p \sum_{i\in\{1,2\}} \lambda^i \left(\int_{\underline{k}}^{E[u_l]} f^i(k)dk \left(1 + (1-\alpha) \ \omega \int_{E[u_l]}^{1} f^i(k)dk\right)\right) \tag{2}$$

Note that the first term in (2) is the consumption by influencers (with relatively lower opportunity costs) within both segments, and the second term is the consumption by followers within these two segments who consume if they receive a positive WOM (with probability $(1 - \alpha)$).

For this equilibrium to exist, neither type should have an incentive to deviate. If the low-type deviates and mimics the high-type (via demand lockout), consumers' posterior belief is that the firm is the high type and its profit is given by $\Pi_l|R \ = \lambda^1 p\left(\int_{\underline{k}}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\ \omega\ \int_{E[u_h]}^{1} f^1(k)dk\right)\right)$. Segment 1 consumers with a sufficiently low opportunity cost ($k \leq E[u_h]$) consume in the first period because they (mistakenly) believe the firm to be high type. In the second period, there is no consumption if the product is low quality, and if the product is high quality, some consumers with high opportunity cost ($k \geq E[u_h]$) receive WOM and consume the product.

If the high-type firm deviates and mimics the low type (by not locking out the demand), consumers' posterior belief is that the product is low quality and profits are $\Pi_h|N \ = p\sum_{i\in\{1,2\}}\lambda^i\left(\int_{\underline{k}}^{E[u_l]} f^i(k)dk\left(1+\alpha\ \omega\ \int_{E[u_l]}^{1} f^i(k)dk\right)\right)$. Here, influentials believe the product to be of low expected quality in the first period and consume accordingly. In the second period, WOM reveals the quality to some followers in each segment, and they consume the product if it is high quality. The remaining followers do not observe WOM and do not consume it.

For an equilibrium that results in the correct beliefs about the quality, the high-type firm prefers using lockout, and the low-type firm prefers not using lockout. The low type can follow the equilibrium path and not lockout demand, thereby revealing its low expected quality. Alternatively, the low type can mimic the high type by locking out, upon which consumers believe the product to be of high expected quality. The cost of mimicking comes directly from locking out demand, eliminating the profits from consumers in segment 2. Simultaneously, it gains a number of influentials in segment 1 because they assume the product to be high quality. In the second period, WOM reveals the quality, and followers only consume if the product turns out to be high quality. The low-type firm thus has no incentive to deviate as long as segment 2 is sufficiently large.

The high-type firm needs to prefer lockout over not lockout. The cost of not locking out demand comes from reduced consumption in segment 1 due to the belief that the expected quality is low. In the first period, only consumers with a low opportunity cost (i.e., $k \leq E[u_l]$) consume the product, believing it to be low quality. WOM allows some followers to observe the quality, but there is less consumption compared to the case where the signal reveals the high type. The gain from serving both segments comes from consumers with a low opportunity cost in segment 2, as well as followers in segment 2 that observe WOM. The high type has no incentive to deviate as long as the cost of being perceived as low quality exceeds the additional segment served. The cost of being perceived as low type is higher if WOM is relatively ineffective at informing followers of the actual quality (i.e., low ω). Similarly, if there are more consumers with relatively high opportunity costs in segment 1, the cost to being perceived to be low quality is higher. Proposition 1 formalizes the separating equilibrium.

Proposition 1: *In a separating equilibrium, the high-type firm excludes the low opportunity cost segment, and the low-type firm serves both segments (*$\sigma_F(R|h) = 1$ and $\sigma_F(R|l\ ) = 0$*). Consumers' beliefs are that firms using the lockout signal are of the high type and firms not using the lockout signal are the low type (*$\mu_C(R) = 1$ and $\mu_C(N) = 0$*). Furthermore, for a sufficiently low prior probability of the firm being high type (*$\mu_0 < \mu_0^*$*), the separating equilibrium is unique.*

In the appendix, we provide detailed proof for this proposition and describe the necessary conditions for its existence[23]. The uniqueness is shown in the Web Appendix (Proofs of some of the auxiliary Lemmas and Propositions are also in the Web Appendix).

To understand the intuition behind the equilibrium, we compare the payoff to increased consumption in the first period in segment 1 (see Fig. 5). Both high and low-type firms prefer to increase consumption in the first period. Increased consumption in the first period also has a spillover on demand in the second period due to a higher WOM. Intuitively, the single crossing property in the model follows from the fact that WOM plays a more significant role when consumers have a higher opportunity cost because more consumers delay consumption until

[23] There exists a region in which pooling equilibria exist, but a separating equilibrium with the low type excluding consumers is never an equilibrium.

the second period. Thus, reducing uncertainty in the high opportunity cost segment 1 is relatively more valuable to the high type because a larger mass of consumers only consume conditional on positive WOM.

Consumers can accurately infer the firm type from the observed signaling action in the separating equilibrium. The high-type firm benefits from signaling its quality in two ways. First, consumers with higher opportunity costs are now willing to consume the product and generate positive WOM that leads the followers to purchase the product in the subsequent period. Interestingly, the high type eliminates the segment that, in the absence of a signal, would have consumed the product at a higher proportion. Similar to other signaling equilibria (e.g., Milgrom and Roberts 1986), the net benefit of signaling is bounded above by the incentive compatibility of the low type. As the high opportunity cost segment becomes larger, the opportunity cost of not using the lockout signal to signal quality to this segment increases for the low type. Specifically, as $\lambda^1 \to 1$, the signaling equilibrium's incentive compatibility constraint on the low-quality type breaks down. For $\frac{\int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+(1-\alpha)\ \omega\ \int_{E[u_l]}^{1} f^2(k)dk\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\ \omega\left(\int_{E[u_h]}^{1} f^1(k)dk-\int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)} < \frac{\lambda^1}{(1-\lambda^1)}$, the low type has the incentive to mimic the high type and to exclude segment 2 to "mislead" consumers in the first period into believing they are of the high type. This breaks the equilibrium, and both firms revert to not excluding any segment when segment 1 is sufficiently large. As is true in most signaling games, uncertainty plays a vital role in the existence of the outlined separating equilibrium. We formally illustrate the role of uncertainty (via Lemma 1) (proof is in the Web Appendix).

Lemma 1: *There exists an $\alpha^* < 1$, such that no separating equilibrium survives if firm uncertainty is too low ($\alpha > \alpha^*$). When μ_0 is sufficiently high, high-type profits under full certainty are lower than high-type profits in a separating equilibrium under uncertainty ($\Pi_h|(R, \alpha = \alpha^*) > \Pi_h(N, \alpha \to 1)$.*

This result on the role of uncertainty is noteworthy because it highlights an interesting tension in signaling games with endogenous word-of-mouth. If the firm observes quality perfectly (i.e., $\alpha \to 1$), an action (such as locking out demand) that reveals high quality will induce all consumers to consume in the first period and render WOM useless. However, whenever WOM does not affect the outcome, the low type can mimic the high type without losing any revenue, thus rendering this equilibrium infeasible. As a result, this uncertainty can benefit the high-quality type because it allows for a credible signaling mechanism.

Advertising versus demand lockout : Demand lockout is one of the many actions available to firms to signal their product type, absent pricing. An obvious alternative to this strategy is advertising – a signaling mechanism well-studied in the literature. We derive an equilibrium with advertising in the Web Appendix.

Both advertising and demand lockout can be credible signals of quality. Demand lockout is a credible signal because it excludes a segment of consumers willing to consume the product. It is effective when the low opportunity cost segment is relatively more important for a low-quality offering. Within our framework, both these signaling mechanisms (i.e., advertising and demand lockout) could induce the same beliefs about quality. However, it is unclear how firms should pick a specific marketing action as a signaling mechanism.

We now consider the case where both signaling mechanisms (lockout and advertising) are available. The high-type firm can choose either signal based on profit maximization. The action space is given by $s =$ {R, N} × {A, NA}, where A and NA refer to advertising or not advertising. While several pooling equilibria are possible under some off-path beliefs, we focus on separating equilibria. The only feasible low-type action in separating equilibria is to play $(N \times NA)$. Because advertising and lockout are costly and reveal the same information in isolation, the high type always prefers a separating PBE with only advertising or demand lockout over both signaling actions [24]. We assume that the high type uses the profit-maximizing signal. As long as $\Pi_h|(R \times NA) > \Pi_h|(N \times A)$, or, equivalently, $\lambda^2 \left(\int_{\underline{k}}^{E[u_h]} f^2(k)dk \, \left(1 + \alpha\, \omega\, \int_{E[u_h]}^{1} f^2(k)dk \right)\right) < \frac{a}{p}$, the demand lockout dominates the advertising signal. In other words, if the advertising costs are high, the opportunity cost from excluding a segment makes

[24] In our theoretical model, the two signaling mechanisms reveal the same information. In an empirical setting, we would expect the signals to be imperfect substitutes. For example, if consumers only imperfectly observe advertising or demand lockout, both signals could be imperfect substitutes.

lockout a more attractive signal. Note that the distribution of segment 1 does not enter the equation because both signals have the same (positive) effect on segment 1 demand but differ in their cost. We can now characterize the equilibrium behavior as follows:

Proposition 2*: Whenever advertising is sufficiently expensive* $(a > a^*)$*, there exists a range for which advertising is a PBE; but is dominated by lockout signaling (*$\Pi_h|(R \times NA) > \Pi_h|(N \times A)$*).*

Proposition 2 shows that the demand lockout signaling mechanism does not replace traditional advertising as a signal if the firm has complete control over the advertising cost a**.** However, in many contexts, the cost of conducting a visible advertising campaign is relatively high and does not scale with firm size. This situation might be particularly evident for niche firms offering high-quality products. In this situation, demand lockout could serve as an alternative. When advertising is relatively costly or when the segment of consumers with high opportunity cost is relatively large (but not too large), the use of advertising as a signal is dominated by the demand lockout.

If the size of the segment with high opportunity cost is too small, a high type has no incentive to signal. As the size of this segment increases, the high type prefers to spend on advertising to capture both segments. When the size of this segment is sufficiently large, the opportunity cost from excluding the low opportunity cost segment gets sufficiently small, and the high-type firm prefers to use demand lockout as a signal. Finally, as the high opportunity cost segment approaches 1, the equilibrium breaks down; it is no longer incentive-compatible for the low-quality type not to use demand lockout. Instead, the high-type firm uses advertising as a signal to separate.

Third-party information: So far, we have assumed (somewhat unrealistically) that the products are released without any available information. In reality, expert reviews or other information sources are often available when consumers are planning to consume experience goods. We now explicitly incorporate third-party information within our model as another source of (noisy) information consumers could use. We modify the timing of the game slightly. First, nature decides the firm type. Second, upon observing its type, a firm can choose to advertise, use demand lockout, or not use any signal. Third, expert reviews arrive with certain valence (i.e., whether the reviews are positive or negative), and consumers purchase one unit at most after observing the firm action and expert reviews.

In the absence of any information, consumers use their prior (maybe from their past consumption experiences) to infer the quality $E[q|\mu_0] = \mu_0 E[q_h] + (1-\mu_0)E[q_l]$. Now assume that consumers receive an unbiased but noisy third-party signal. This signal is a proxy for reviews, given by $r(q_j) \in \{\underline{q}, \bar{q}\}$. The accuracy of the information is given by $P(r(q_j) = q_j) = \gamma$, where $\gamma \in (\frac{1}{2}, 1)$, i.e., the higher quality product is more likely to receive a positive valence than a lower quality product and vice versa. Because even the firm only knows the *expected* quality of their product, third-party information affects consumers' posterior belief even when the signaling equilibrium fully reveals the firm type.

Using Bayes rule, consumers form their beliefs about the quality of product j. To simplify notation, let $q(S) = P\left(q_j = \bar{q} \middle| \mu_C(S)\right) = \mu_C(S)\alpha + (1-\mu_C(S))(1-\alpha)$ denote the probability of a product being high quality, conditional on the belief about firm type.

$$P(q_j = \bar{q}|r(q_j) = \bar{q}) = \frac{P(r(q)=\bar{q}|q=\bar{q})P(q=\bar{q})}{P(r(q)=\bar{q}|q=\bar{q})P(q=\bar{q}) + P(r(q)=\bar{q}|q=q_l)P(q=\underline{q})} = \frac{\gamma q(S)}{\gamma q(S) + (1-\gamma)(1-q(S))} \quad (5)$$

$$P(q_j = \bar{q}|r(q_j) = \underline{q}) = \frac{P(r(q)=\underline{q}|q=q_h)P(q=q_h)}{P(r(q)=\underline{q}|q=q_h)P(q=q_h) + P(r(q)=q_l|q=q_l)P(q=\underline{q})} = \frac{(1-\gamma)q(S)}{(1-\gamma)q(S) + \gamma(1-q(S))} \quad (6)$$

To reduce notational clutter, let $b_h|q(S) = \frac{\gamma q(S)}{\gamma q(S)+(1-\gamma)(1-q(S))}$ and $b_l|q(S) = \frac{(1-\gamma)q(S)}{(1-\gamma)q(S)+\gamma(1-q(S))}$.

The reviews induce the following beliefs about the quality of the product, conditional on receiving a positive review, $E[q|(S, \gamma, r(q_j) = \bar{q})] = b_h\bar{q} + (1-b_h)\underline{q}$, or a negative review, $E\left[q|\left(S, \gamma, r(q_j) = \underline{q}\right)\right] = b_l\bar{q} + (1-b_l)\underline{q}$.

First, consider the benchmark without any signal. Consumers make their inferences about the quality of the product based solely on the observed third-party information and prior belief.

Lemma 3*: As the informativeness of reviews increases ($\gamma \to 1$), consumer beliefs approach the correct quality level $E[q|(S, \gamma, r(q_j) = q_j)] \to q_j$). For all non-perfect review accuracies ($\gamma < 1$), a positive review cannot induce full consumption.*

This intuitive lemma formally establishes that, unless the reviews are perfectly accurate, there remains information asymmetry and, thus, a potential for signaling. The reviews dampen the asymmetric information problem, and some consumers participate as influentials in the market whenever they observe a positive review but would not consume absent third-party information. However, because the reviews are noisy, a high-quality type could sometimes mistakenly receive an (overall) negative review, and some marginal consumers will not consume the product. Unless the reviews are perfectly accurate, there remains an incentive for the high-quality product to signal its quality and to induce a separating equilibrium where high opportunity cost consumers learn the expected quality of the product.

Next, we analyze the profit difference for the high type between the signaling-induced separating equilibrium and the pooling equilibrium for different levels of third-party accuracy. To do so, we calculate the difference between profits with the lockout signal and profits in the pooling equilibrium ($\Delta\Pi_j = \Pi_h|(R, \mu_C(R) = 1) - \Pi_h|(N, \mu_C(N) = \mu_0)$).

Proposition 3: *A lockout signaling equilibrium (Proposition 1) continues to exist in the presence of expert reviews for sufficiently low accuracy of reviews. For sufficiently low* μ_0*: If third-party information is inaccurate, high-type profits are higher in a separating equilibrium. If the accuracy of third-party information is high, high-type profits are higher in the pooling equilibrium* $\Delta\Pi_j < 0$ *for* $\gamma > \bar{\gamma}$ *and* $\Delta\Pi_j \geq 0$ *for* $\gamma \leq \underline{\gamma}$*).*

The proposition establishes that in our setting, third-party reviews are a substitute for the signaling mechanism. If the content of third-party reviews is sufficiently informative, the high type has no incentive to take a costly action to communicate quality to consumers. For a sufficiently low accuracy of third-party reviews, the lockout equilibrium continues to exist, and profits from the separation are higher than those from the pooling. If reviews are perfectly accurate, the only possible equilibrium is the pooling equilibrium, where consumers are perfectly informed about product quality from the third-party review.

To summarize, in its parsimony, our model is intended to provide intuition for the demand lockout mechanism transparently. We add three important extensions in the accompanying Web Appendix (Part T3). First, we introduce the notion of *endogenous costs* - modeling that a product's (expected) quality is a function of a firm's endogenous investment cost. Next, we remove our assumption of no communication across the two consumer segments. In the second extension, we consider the case where a limited *spillover of word of mouth* happens across the segments. Thirdly, we allow the lockout to be imperfect such that a proportion of consumers in the locked-out segment can consume. Our analysis and results suggest that the demand lockout continues to be a useful signaling strategy with the inclusion of these additional layers into the model.

Alternative Explanations: We have presented many empirical facts, consistent with the theoretical model above. Clearly, there are other potential theoretical models that are consistent with the empirical observations. We now present some obvious alternative explanations that we recognize as limitation. One concern might be that the MPAA systematically discriminates against specific movies and is more likely to give more restrictive classifications to movies produced by smaller and less powerful studios. A second concern is that the MPAA might use unobservable movie characteristics, such as high quality, in its decision, and we would not be able to parse this effect from signaling. Third, we have primarily considered quality on a vertical scale. However, under asymmetric information, consumers also need to make an inference about fit. R-rated movies are generally more realistic, violent, and graphic, and a consumer who prefers these attributes will correctly use the fact that R-rated movies are more consistent with her taste. In other words, it could be hypothesized that the differences in beliefs about horizontal characteristics drive the results. Fourth, we assume that there is random noise in the ratings process. If the rating choice is correlated with some unobservable, we might be capturing the effect of that unobservable as well. Although these claims are plausible, we discuss (in Web Appendix T4) why these explanations are potentially not fully consistent with the empirical facts presented.

Concluding Remarks

In this paper, we have proposed a novel signaling mechanism based on the exclusion of a profitable segment of potential consumers. We show that the opportunity cost of excluding a segment of consumers can serve as a credible signal of quality to a segment with higher opportunity cost. In contexts where differentiated pricing is not feasible, we find that the signal coming from locking out demand can substitute traditional signals like advertising and that expert reviews do not eliminate the value of such a signal. We find that the proposed theory is largely consistent with the data from the motion picture industry, hypothesizing that movie studios use R ratings to credibly exclude a segment (i.e., moviegoers under age 17) from their market demand to signal quality. Using subtitle data to control for differences in content, we find evidence consistent with multiple comparative statics of the signaling equilibrium. We find that movies with more uncertain quality use R ratings more extensively, and the revenues during the opening and subsequent weekends are consistent with the R ratings' role in mitigating information asymmetry. These results are managerially relevant beyond other factors such as advertising and third-party information. For example, a high-quality movie produced by a non-major studio with relatively few reviews (i.e., a low Critic's count, a high critics' mean) benefits particularly much from this signaling. Even evaluated at the median of all continuous variables, log-revenue for domestic, non-major movies is significantly higher for R-rated movies (difference of 1.06).

Our paper has several limitations. The first limitation is inherent to all empirical studies of signaling. Because we cannot observe the actual quality of the movies, we cannot fully identify a causal effect of the R rating. However, accounting for quality differences after controlling for the "inappropriateness" of a movie, we observe the empirical patterns consistent with the proposed theory. We consider multiple alternative hypotheses but cannot completely rule out that some other theoretical mechanism could induce the same data-generating process. Second, how well our findings will generalize to different settings is not clear. The MPAA's R rating provides a convenient demand lockout device in the movie industry. In our discussion, we alluded to other lockout devices (e.g., "by invitation only" programs, geographic lockout). Still, future research should consider how firms can implicitly exclude segments and communicate that exclusion to the remaining segments. For example, one possible way is by advertising in channels visited only by one segment (e.g., advertisements in niche magazines); another is by using strong political stances to "quasi-lockout" segments that disagree with the message.

The results of our study have several managerial implications. We show that locking out potential demand can convince consumers of a product's quality. Managers need to consider the signaling effects of serving specific segments or refusing to serve certain segments. In the context of the movie industry, we show that locking out a segment of potential consumers can improve profits. These results are significant for firms in asymmetric information settings, where marketing has little control over pricing and promotion.

References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3), 488-500.

Allen, F. (1984). Reputation and product quality. *The RAND Journal of Economics*, 311-327.

Amaldoss, W., & Jain, S. (2005). Pricing of conspicuous goods: A competitive analysis of social effects. *Journal of Marketing Research*, *42*(1), 30-42.

Ashoori, M., Schmidbauer, E., & Stock, A. (2020). Exclusivity as a Signal of Quality in a Market with Word-of-Mouth Communication. Review of Marketing Science, 18(1), 99-115.

Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11.

Balachander, S., Liu, Y. & Stock, A. (2009). An Empirical Analysis of Scarcity Strategies in the Automobile Industry. *Management Science*, 55(10), pp. 1623 – 1637

Bernstein, P. (2014). *How The MPAA Really Works And How to Get The Rating You Want*. Retrieved from Indiewire: https://www.indiewire.com/2014/07/how-the-mpaa-really-works-and-how-to-get-the-rating-you-want-24671/

Bourreau, M., & Gaudin, G. (2018). Streaming Platform and Strategic Recommendation Bias. *Available at SSRN*.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Brown, A. L., Camerer, C. F., & Lovallo, D. (2012). To review or not to review? Limited strategic thinking at the movie box office. *American Economic Journal: Microeconomics*, *4*(2), 1-26.

Calvano, E., & Jullien, B. (2019). Can we trust the algorithms that recommend products online? A theory of biased advice with no pecuniary incentives and lab evidence. *Working Paper*.

Chakraborty, A., & Harbaugh, R. (2014). Persuasive puffery. *Marketing Science*, *33*(3), 382-400.

Chatterji, A. K., & Toffel, M. W. (2018). The new CEO activists. *HBR'S 10 MUST*, 47.

Chen, Y., & Cui, T. H. (2013). The benefit of uniform price for branded variants. *Marketing Science*, *32*(1), 36-50.

Cho, I. K., & Kreps, D. M. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, *102*(2), 179-221.

Cho, S., & Rust, J. (2010). The flat rental puzzle. *The Review of Economic Studies*, *77*(2), 560-594.

Chu, W., & Chu, W. (1994). Signaling quality by selling through a reputable. *Marketing Science*, 13(2), 177-189.

Clyde, J. (2014), Movies with the same content but different MPAA ratings". KSL.com.

Dehejia, R. H., & Wahba, S. (2002). Propensity score matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1), 151-161.

Erdem, T., & Swait, J. (1998). Brand equity as a signaling phenomenon. *Journal of Consumer Psychology*, 7(2), 131-157.

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3), 535-74.

Grimmer, J., Messing, S., & Westwood, S. J. (2017). Estimating heterogeneous treatment effects and the effects of heterogeneous treatments with ensemble methods. *Political Analysis*, 25(4), 413-434.

Harper, F. M., & Konstan, J. A. (2016). The *movielens* datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4), 19.

Hartmann, J., Huppertz, J., Schamp, C., & Heitmann, M. (2019). Comparing automated text classification methods. *International Journal of Research in Marketing*, 36(1), 20-38.

Hicks, C. (2013). Has Hollywood Lost Its Mind?: A Parent's Guide to Movie Ratings. Familius.

Hirano, K., Imbens, G. W., & Ridder, G. (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica*, *71*(4), 1161-1189.

Ho, J. Y., Liang, Y., Weinberg, C. B., & Yan, J. (2018). Uniform and Differential Pricing in the Movie Industry. J*ournal of Marketing Research*, 55, no. 3 (2018): 414-431.

Hoberg, G., & Phillips, G. (2016). Text-based network industries and endogenous product differentiation. *Journal of Political Economy*, 124(5), 1423-1465.

Jacobs, Julia (2023), Infinity Pool and the Battle for an R-rating. *New York Times*

Jiang, B., & Yang, B. (2019). Quality and pricing decisions in a market with consumer information sharing. Management Science, 65(1), 272-285.

Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *In European conference on machine learning (pp. 137-142). Springer, Berlin, Heidelberg*.

Kalra, A., & Li, S. (2008). Signaling quality through specialization. *Marketing Science*, 27(2), 168-184.

Lewis, D. D., Yang, Y., Rose, T. G., & Li, F. (2004). Rcv1: A new benchmark collection for text categorization research. *Journal of machine learning research*, 5(Apr), 361-397.

Mayzlin, D., & Shin, J. (2011). Uninformative advertising as an invitation to search. *Marketing Science*, *30*(4), 666-685.

McMillan, R. S. (2007). Different flavor, same price: The puzzle of uniform pricing for differentiated products. *Working Paper*

Miklós-Thal, J., & Zhang, J. (2013). (De) marketing to manage consumer quality inferences. *Journal of Marketing Research*, *50*(1), 55-69.

Milgrom, P. & Robers, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy*, 94(4), 796-821.

Moon, S., Bergey, P. K., & Iacobucci, D. (2010). Dynamic effects among movie ratings, movie revenues, and viewer satisfaction. *Journal of marketing*, *74*(1), 108-121.

Moorthy, S., & Srinivasan, K. (1995). Signaling quality with a money-back guarantee: The role of transaction costs. *Marketing Science*, 14(4), 442-466.

Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31.2 (2017): 87-106.

Nelson, P. (1974). Advertising as information. *Journal of Political Economy*, 82(4), 729-754.

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

Orbach, B. Y., & Einav, L. (2007). Uniform prices for differentiated goods: The case of the movie-theater industry. *International Review of Law and Economics*, 27(2), 129-153.

Pesendorfer, W. (1995). Design innovation and fashion cycles. *The American Economic Review*, 771-792.

Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130-137.

Proserpio, D., Hauser, J. R., Liu, X., Amano, T., Burnap, A., Guo, T., ... & Yoganarasimhan, H. (2020). Soul and machine (learning). *Marketing Letters*, *31*(4), 393-404.

Pulver, A. (2013). Philomena: Weinsteins win MPAA appeal against R rating. *The Guardian*

Rao, A. R., & Monroe, K. B. (1989). The effect of price, brand name, and store name on buyers' perceptions of product quality: An integrative review. *Journal of Marketing Research*, 26(3), 351-357.

Rao, R. S., & Schaefer, R. (2013). Conspicuous consumption and dynamic pricing. *Marketing Science*, *32*(5), 786-804.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41-55.

Sahni, N. S., & Nair, H. S. (2020). Does Advertising Serve as a Signal? Evidence from a Field Experiment in Mobile Search. *The Review of Economic Studies*, *87*(3), 1529-1564.

Sapra, A., Truong, V. A., & Zhang, R. Q. (2010). How much demand should be fulfilled? *Operations Research,* 58(3), 719-733.

Schmalensee, R. (1978). A model of advertising and product quality. *Journal of Political Economy*, 86(3), 485-503.

Schneider, M. (2018). Glossier Will See You Now. *New York Times*.

Shiller, B., & Waldfogel, J. (2011). Music for a song: an empirical look at uniform pricing and its alternatives. *The Journal of Industrial Economics*, *59*(4), 630-660.

Shin, J. (2005). The role of selling costs in signaling price image. *Journal of Marketing Research*, *42*(3), 302-312.

Simester, D. (1995). Signaling price image using advertised prices. *Marketing Science*, 14(2), 166-188.

Sisario, B. (2011). *New Service offers music in quantity, not by Song*. The New York Times.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.

Spotify. (2018, 10 10). *Celebrating a Decade of Discovery on Spotify*. Retrieved from https://newsroom.spotify.com/2018-10-10/celebrating-a-decade-of-discovery-on-spotify/

Stock, A. & Balachander S. (2005). The Making of a Hot Product: A Signaling Explanation of Marketers' Scarcity Strategy. *Management Science*, 51(8), pp. 1181-1192

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267-288.

Tiedemann, J. (2012). Parallel Data, Tools and Interfaces in OPUS. *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, Vol. 2012, pp. 2214-2218.

Timoshenko, A., & Hauser, J. R. (2019). Identifying customer needs from user-generated content. *Marketing Science*, 38(1), 1-20.

Vettas, N. (1997). On the informational role of quantities: Durable goods and consumers' word-of-mouth communication. *International Economic Review*, 915-944.

Wernerfelt, B. (1988). Umbrella branding as a signal of new product quality: An example of signaling by posting a bond. *The RAND Journal of Economics*, 458-466.

Whitten, S.(2022). How many F-Bombs Trigger an R-rating? An Obscure Movie Industry Panel Decides. *CNBC.com*

Xu, Z., & Dukes, A. (2022). Personalization from customer data aggregation using list price. *Management Science*, *68*(2), 960-980.

Fig. 1a: Stylized simultaneity problem of movie ratings and quality

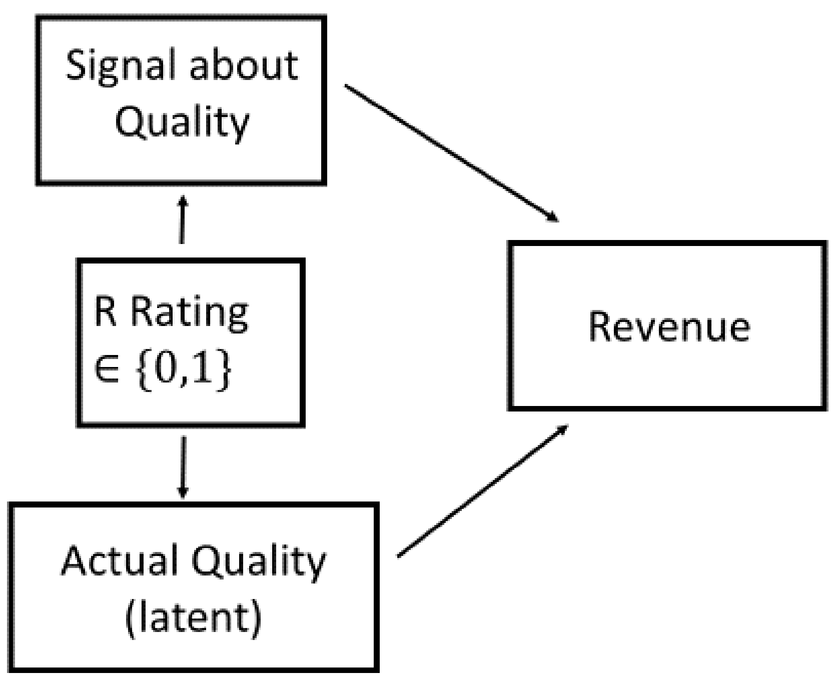


Fig. 1b: Stylized illustration of the potential solution to the simultaneity problem

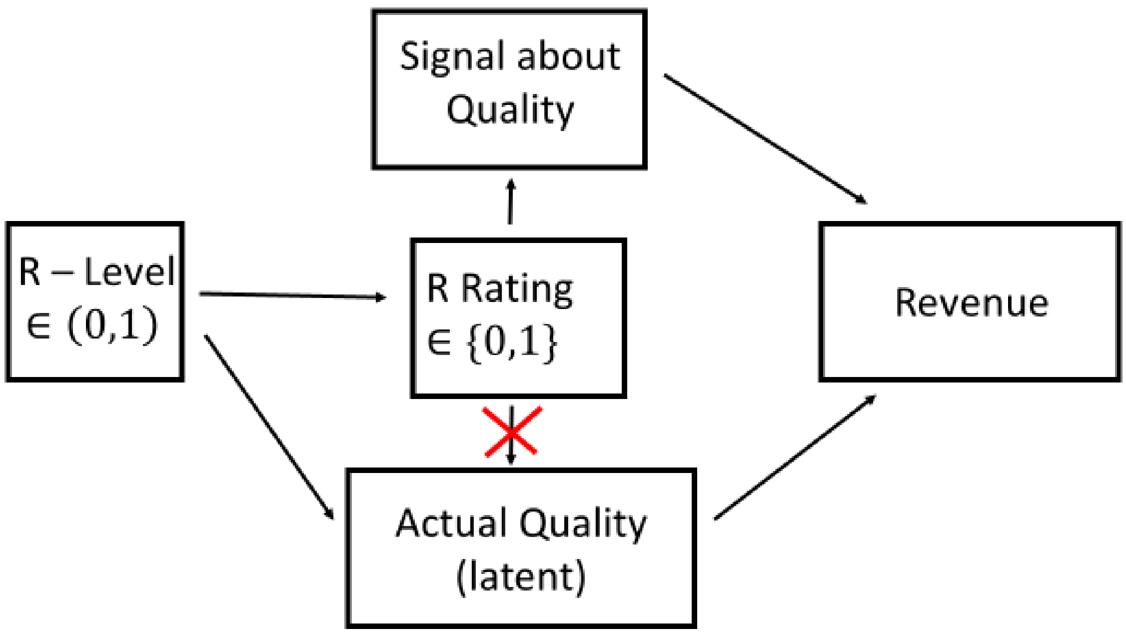
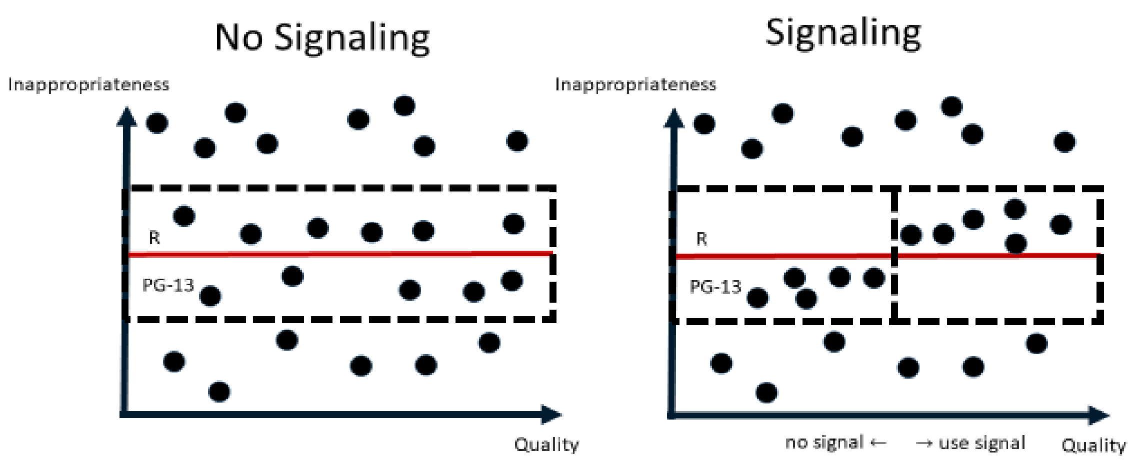
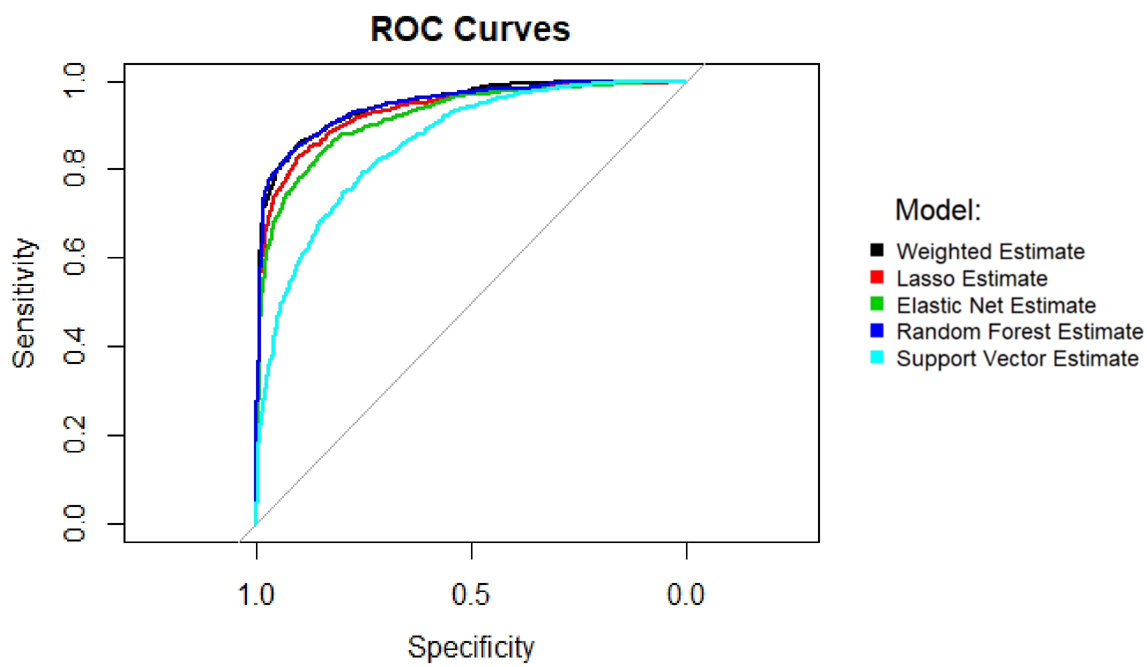


Fig. 2: Stylized illustration of signal identification



<u>**Note:**</u> The dotted box shows movies that can either get an R rating or a PG-13 rating without making significant changes to the content. On the left side, there is no signaling. The right side depicts the separating equilibrium, where movies close to the cutoff with high quality use the R rating as a demand lockout signal.

Fig. 3: Comparing the predictive performance of different models



Note: AUC for Weighted Estimate: 0.9644, Lasso Estimate: 0.933, Elastic Net Estimate: 0.919, Random Forest Estimate: 0.9461, Support Vector Regression: 0.8583

Fig 4: Timing of the Game

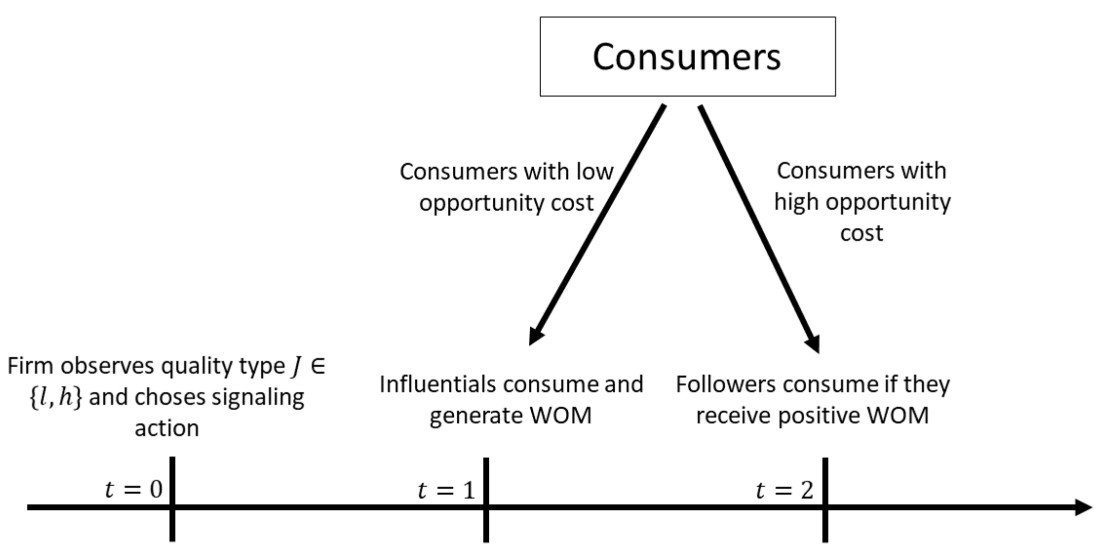


Fig. 5: Segment-level Demand under Separating Equilibrium
(q_l = 4 and q_h = 6)

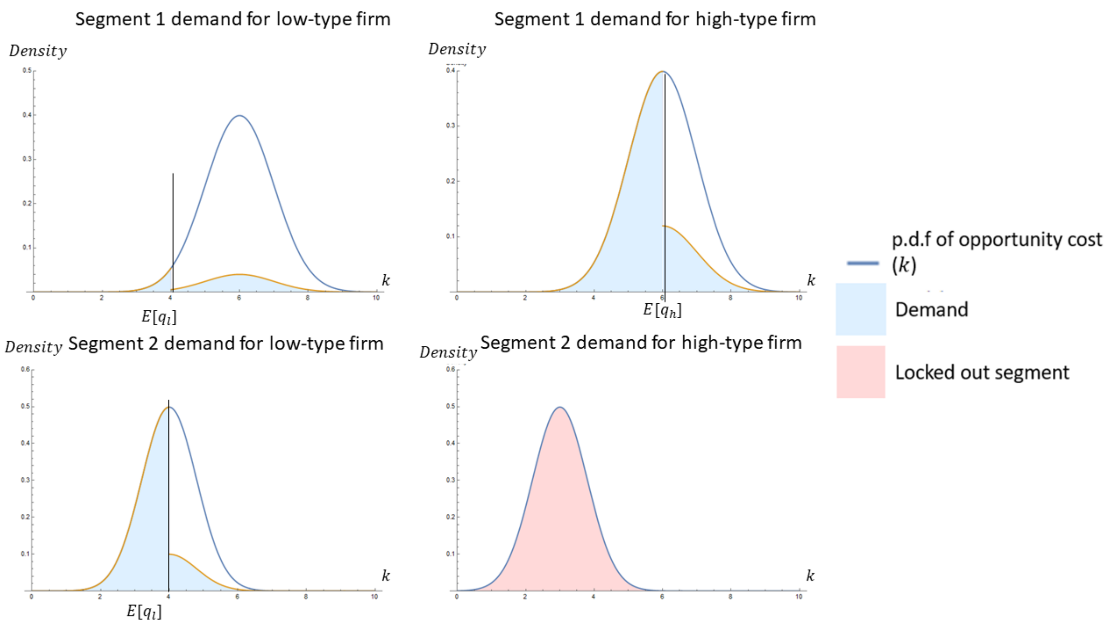


Table 1: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
R	1,502	0.441	0.497	0	1
Revenue week 1	1,502	18.085	23.764	0.001	270.019
Revenue week 2	1,499	11.889	15.464	0.004	146.53
Revenue week 3	1,461	7.1675	9.231	0.001	111.856
Revenue week 4	1,428	4.5682	6.162	0.001	69.926
Revenue week 5	1,428	2.9852	4.576	0.001	66.330
Total Advertising Spending	1,502	18.374	12.331	0	61.6949
Critics' mean	1,502	56.281	15.432	14.154	94.679
Critics' count	1,502	30.698	8.420	4	77
Critics' standard deviation	1,502	15.801	3.260	6.952	28.831
Foreign	1,502	0.160	0.367	0	1
Major studio	1,461	0.264	0.441	0.000	1.000

Note: Revenue and budget in $millions. Consumer reviews are on a 1–5 scale. Critic reviews are on a 0–100 scale.

Table 2: Logistic Regression of R-rating on movie characteristics

	Dependent variable: R rating	
	(1)	(2)
Ad spending before release	-0.011	-0.007
	(0.010)	(0.014)
Budget	-0.007^{**}	-0.014^{***}
	(0.003)	(0.005)
Critics' count	-0.047^{***}	-0.030
	(0.014)	(0.021)
Critics' standard deviation	0.281^{***}	0.252^{***}
	(0.035)	(0.044)
Major studio	-0.005	-0.173
	(0.220)	(0.292)
Foreign	0.749^{***}	0.567^{*}
	(0.267)	(0.328)
Critics' mean	0.038^{***}	0.023^{**}
	(0.008)	(0.010)
Constant	-3.969^{***}	-1.939
	(0.715)	(1.180)
Year Fixed Effects	No	Yes
Genre Fixed Effects	No	Yes
Observations	705	705
Log Likelihood	-346.143	-229.792
Akaike Inf. Crit.	708.703	527.584
Note: Budget and advertising spending $Millions	*p<0.1, **p<0.05, ***p<0.01	

Note: This analysis uses a propensity-score matched sample of observations within common support

Table 3: Regression of opening week revenue on movie characteristics

	Dependent variable: Log Revenue			
	(1)	(2)	(3)	(4)
R	-5.438***	-5.727***	-5.895***	-5.462***
	(1.322)	(1.299)	(1.381)	(1.394)
logBudget	0.825***	0.775***	0.553***	0.559***
	(0.142)	(0.140)	(0.150)	(0.152)
logAd spending before release	0.415***	0.180**	0.192**	0.146*
	(0.079)	(0.085)	(0.086)	(0.086)
Critics' count	0.019	0.048**	0.040*	0.054**
	(0.022)	(0.022)	(0.023)	(0.024)
Critics'_standard deviation	-0.214***	-0.270***	-0.250***	-0.222***
	(0.056)	(0.055)	(0.059)	(0.060)
Critics'_mean	-0.069***	-0.079***	-0.070***	-0.069***
	(0.012)	(0.012)	(0.013)	(0.013)
Foreign		0.832**	0.660*	0.798**
		(0.356)	(0.360)	(0.357)
Major studio		1.569***	1.716***	1.930***
		(0.279)	(0.297)	(0.297)
R × logBudget	0.133	0.113	0.302*	0.310*
	(0.159)	(0.157)	(0.167)	(0.169)
R× logAd spending before release	-0.134	0.090	0.086	0.102
	(0.090)	(0.096)	(0.096)	(0.095)
R × Critics' count	-0.031	-0.063**	-0.048*	-0.045*
	(0.025)	(0.025)	(0.026)	(0.026)
R × Critics' standard deviation	0.256***	0.314***	0.285***	0.246***

	(0.060)	(0.059)	(0.063)	(0.064)
R× Critics' mean	0.036***	0.047***	0.036**	0.038***
	(0.014)	(0.013)	(0.014)	(0.014)
R × Foreign		-1.584***	-1.407***	-1.498***
		(0.397)	(0.400)	(0.395)
R × Major studio		-0.826**	-1.082***	-1.327***
		(0.332)	(0.347)	(0.347)
Constant	18.962***	19.371***	20.185***	19.318***
	(1.188)	(1.170)	(1.318)	(1.342)
Year Fixed Effects	No	No	Yes	Yes
Genre Fixed Effects	No	No	No	Yes
Observations	716	701	701	701
R^2	0.525	0.564	0.584	0.608
Adjusted R^2	0.518	0.555	0.566	0.584

Note: Budget data missing for 77 movies & studio info is unavailable for 41 movies *p<0.1, **p<0.05 ***p<0.01

Note: The sample is propensity-score matched within common support.

Table 4: Regression of revenue in weeks 2-5 on movie characteristics

	Dependent variable: Revenue (Week 2-5)			
Lag Revenue ×	(1)	(2)	(3)	(4)
Intercept	0.496***	0.483***	0.459***	0.342***
	(0.030)	(0.031)	(0.100)	(0.115)
Weekly advertising spending		0.015***	0.014***	0.014***
		(0.004)	(0.004)	(0.004)
R	0.014	0.004	-0.438***	-0.303**
	(0.013)	(0.014)	(0.122)	(0.143)
Budget			-0.0003**	-0.0002
			(0.0002)	(0.0002)
Consumer review mean			0.021	0.061*
			(0.029)	(0.035)
Quality shock $\bar{q}$		0.221***	0.199***	0.166***
		(0.046)	(0.057)	(0.060)

# of Consumer Reviews				-0.035**
				(0.016)
R × Consumer review mean			0.133***	0.085*
			(0.037)	(0.045)
R × Quality shock $\bar{q}$		-0.147**	-0.260***	-0.228***
		(0.059)	(0.070)	(0.072)
R × # of Consumer Reviews				0.038**
				(0.018)
Year Fixed Effects	Yes	Yes	Yes	Yes
Genre Fixed Effects	Yes	Yes	Yes	Yes
Observations	2,972	2,972	2,806	2,806
R^2	0.818	0.822	0.824	0.825
Adjusted R^2	0.817	0.820	0.822	0.823
Residual Std. Error	4.149 (df = 2955)	4.109 (df = 2940)	4.196 (df = 2772)	4.193 (df = 2770)
F Statistic	779.832*** (df = 17; 2955)	424.710*** (df = 32; 2940)	383.005*** (df = 34; 2772)	362.278*** (df = 36; 2770)

Note: $\bar{q}_i = PR_i^U - PR_i^C$, where PR_i^U is the percentile rank of the consumer score for movie i, and R_i^C is the percentile rank of the critics. Budget data is missing for 77 movies in the raw data. # of Reviews in 10,000s

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

Note: The sample is propensity-score matched within a common support.

Table 5: Key model notations

Symbol	Description
i	Segment
j	Product type
$F^i(k)$	The cumulative density function of opportunity cost in segment i
p	Exogenous price
α	Probability of quality being high (low) for high (low) type.
c_j	Variable cost for product type j
C_j	Fixed cost for product type j
q_j	Quality of product
k^i	Opportunity cost for segment i
λ^i	Size of segment i
δ	The prior probability of a product being the high type
γ	Accuracy of third-party information
ω^i	Word-of-mouth base rate for segment i
a	Advertising cost
N/NA	Not signaling

R	Lockout signal
A	Advertising signal

Technical Appendix

This appendix presents the proofs of the four main propositions in the paper. The proofs of the general case, lemmas, and model extensions are presented in an accompanying Web Appendix (WA).

Proof of Proposition 1: First, we lay out the conditions for a potential separating equilibrium in wherein the high-quality product serves segment 1, and the low-quality product serves both segments. Recall that prices are exogenously set at $p > 0$ for all products and $\bar{q} - p = 1$.

Suppose there exists a separating equilibrium with $\sigma_F(R|h) = 1$ and $\sigma_F(R|l) = 0$, which (following Bayes rule) induces consumer beliefs: $\mu_C(R) = 1$ and $\mu_C(N) = 0$. For notational ease, let $[u_h] = \alpha\,\bar{q} + (1-\alpha)\underline{q} - p$, $E[u_l] = (1-\alpha)\bar{q} + \alpha\underline{q} - p$, and $E\left[u_{\mu_0}\right] = \mu_0 E[u_h] + (1-\mu_0)E[u_l]$. Recall that we assume that $F^1(k)$ and $F^2(k)$ are strictly increasing and continuous.

The payoffs for the two types (on the equilibrium path) are given by:

$\Pi_h|R = p\lambda^1\left(\int_{\underline{k}}^{E[u_h]} f^1(k)dk\,\left(1+\alpha\,\omega\,\int_{E[u_h]}^{1} f^1(k)dk\right)\right)$, and

$\Pi_l|N = p\sum_{i\in\{1,2\}}\lambda^i\left(\int_{\underline{k}}^{E[u_l]} f^i(k)dk\left(1+(1-\alpha)\omega\int_{E[u_l]}^{1} f^i(k)dk\right)\right)$.

The payoffs for the two types (off the equilibrium path) when mimicking the other type are given by: $\Pi_h|N = p\sum_{i\in\{1,2\}}\lambda^i\left(\int_{\underline{k}}^{E[u_l]} f^i(k)dk\left(1+\alpha\,\omega\,\int_{E[u_l]}^{1} f^i(k)dk\right)\right)$ and $\Pi_l|R = p\lambda^1\left(\int_{\underline{k}}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\,\omega\,\int_{E[u_h]}^{1} f^1(k)dk\right)\right)$. For this equilibrium to hold, we require: (IC 1) $\Pi_h|R > \Pi_h|N$ and (IC 2) $\Pi_l|N > \Pi_l|R$. After some algebra, the two constraints can be summarized by:

$$\frac{\int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+\alpha\,\omega\,\int_{E[u_l]}^{1} f^2(k)dk\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+\alpha\omega\left(\int_{E[u_h]}^{1} f^1(k)dk-\int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)} < \frac{\lambda^1}{(1-\lambda^1)} < \frac{\int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+(1-\alpha)\,\omega\,\int_{E[u_l]}^{1} f^2(k)dk\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\,\omega\left(\int_{E[u_h]}^{1} f^1(k)dk-\int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)}.$$

Next, we need to check if there exists a set of parameters for which both the above constraints hold. This equilibrium exists for some λ^1 whenever $\int_{E[u_l]}^{1} f^2(k)dk + \int_{\underline{k}}^{E[u_l]} f^1(k)dk - \int_{E[u_h]}^{1} f^1(k)dk < 0$. Noting that $\frac{\partial E[u_l]}{\partial\alpha} < 0$ and $\frac{\partial E[u_h]}{\partial\alpha} > 0$, which implies, by continuity of the probability density: $\frac{\partial}{\partial\alpha}\left(\int_{E[u_l]}^{1} f^2(k)dk + \int_{\underline{k}}^{E[u_l]} f^1(k)dk - \int_{E[u_h]}^{1} f^1(k)dk\right) > 0$, we first check if the inequality holds at the minimum of the expression (i.e. at $\alpha = \frac{1}{2}$)

Evaluated at $\alpha \to \frac{1}{2}$, $E[u_l] = E[u_h] = \frac{\underline{q}+\overline{q}}{2} - p$, the condition for the equilibrium is $\int_{\frac{\underline{q}+\overline{q}}{2}-p}^{1} f^2(k)dk < \int_{\frac{\underline{q}+\overline{q}}{2}-p}^{1} f^1(k)dk - \int_{\underline{k}}^{\frac{\underline{q}+\overline{q}}{2}-p} f^1(k)dk$, which simplifies to:

$2\int_{\underline{k}}^{\frac{\underline{q}+\overline{q}}{2}-p} f^1(k)dk < \int_{\underline{k}}^{\frac{\underline{q}+\overline{q}}{2}-p} f^2(k)dk.$

At the upper bound, $\lim_{\alpha\to 1}\int_{E[u_l]}^{1} f^2(k)dk \to 1$, $\lim_{\alpha\to 1}\int_{E[u_h]}^{1} f^1(k)dk \to 0$, and $\lim_{\alpha\to 1}\int_{\underline{k}}^{E[u_l]} f^1(k)dk \to 0$, and the equilibrium condition implies $1 < 0$, which never holds.

Thus, because $\frac{\partial}{\partial\alpha}\left(\int_{E[u_l]}^{1} f^2(k)dk + \int_{\underline{k}}^{E[u_l]} f^1(k)dk - \int_{E[u_h]}^{1} f^1(k)dk\right) > 0$, there exists, one $\alpha^* = \{\alpha | \int_{E[u_l]}^{1} f^2(k)dk + \int_{\underline{k}}^{E[u_l]} f^1(k)dk - \int_{E[u_h]}^{1} f^1(k)dk\}$, whenever $2\int_{\underline{k}}^{\frac{\underline{q}+\overline{q}}{2}-p} f^1(k)dk < \int_{\underline{k}}^{\frac{\underline{q}+\overline{q}}{2}-p} f^2(k)dk$.

Whenever $2\int_{\underline{k}}^{\frac{\underline{q}+\overline{q}}{2}-p} f^1(k)dk < \int_{\underline{k}}^{\frac{\underline{q}+\overline{q}}{2}-p} f^2(k)dk$ and $\alpha < \alpha^*$, there exists a range $\lambda^1 \in (\underline{\lambda^1}, \overline{\lambda^1})$, such that the separating equilibrium is a PBE. Because $\frac{\lambda^1}{(1-\lambda^1)}$ can take any positive value, it remains to be shown that $\frac{\int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+(1-\alpha)\,\omega\,\int_{E[u_l]}^{1} f^2(k)dk\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\,\omega\left(\int_{E[u_h]}^{1} f^1(k)dk - \int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)}$ is positive, to finish the proof of the existence of the equilibrium. After some algebra, one can show that the term is positive for the equilibrium regions described above, because $0 < \int_{E[u_l]}^{1} f^2(k)dk < \int_{E[u_h]}^{1} f^1(k)dk - \int_{\underline{k}}^{E[u_l]} f^1(k)dk$.

The proof deriving the pooling equilibria, applying the Intuitive Criterion (Cho and Kreps, 1987) and considering uniqueness of the equilibrium is provided in the WA.

Proof of Proposition 2: We aim to show that there exists a range in which both focal separating equilibria exist and the profit for the high-quality type is higher under the lockout separating equilibrium. Rearranging the profits in the lockout and advertising separating equilibrium, profit for the high type from the lockout signal is higher than advertising whenever: $\lambda^2\left(\int_{\underline{k}}^{E[u_h]} f^2(k)dk\ \left(1 + \alpha\,\omega\,\int_{E[u_h]}^{1} f^2(k)dk\right)\right) < \frac{a}{p}$.

We aim to show that there exists a range in which both separating equilibria exist, but the lockout equilibrium dominates the advertising equilibrium. Consider the parameters for which the signaling equilibrium with the high-type playing lockout is a PBE:

$\frac{\int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+\alpha\ \omega\ \int_{E[u_l]}^{1} f^2(k)dk\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+\alpha\omega\left(\int_{E[u_h]}^{1} f^1(k)dk-\int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)} < \frac{\lambda^1}{(1-\lambda^1)} < \frac{\int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+(1-\alpha)\omega\ \int_{E[u_l]}^{1} f^2(k)dk\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\omega\left(\int_{E[u_h]}^{1} f^1(k)dk-\int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)}$. A necessary condition for this PBE is that $\int_{E[u_l]}^{1} f^2(k)dk < \int_{E[u_h]}^{1} f^1(k)dk - \int_{\underline{k}}^{E[u_l]} f^1(k)dk$. Suppose $0 \le \int_{E[u_h]}^{1} f^2(k)dk - \int_{\underline{k}}^{E[u_l]} f^2(k)dk$, which implies that there exists some α such that the advertising separating equilibrium is a PBE as well: $\sum_{i\in\{1,2\}} \lambda^i \int_{E[u_l]}^{E[u_h]} f^i(k)dk\left(1+(1-\alpha)\omega\left(\int_{E[u_h]}^{1} f^i(k)dk - \int_{\underline{k}}^{Eu_l} f^i(k)dk\right)\right) < \frac{a}{p} < \sum_{i\in\{1,2\}} \lambda^i \int_{E[u_l]}^{E[u_h]} f^i(k)dk\left(1+\alpha\omega\left(\int_{E[u_h]}^{1} f^i(k)dk - \int_{\underline{k}}^{Eu_l} f^i(k)dk\right)\right)$. Suppose advertising cost is approaching the upper limit of the IC constraint: $\frac{a}{p} \to \sum_{i\in\{1,2\}} \lambda^i \int_{E[u_l]}^{E[u_h]} f^i(k)dk\left(1+\alpha\omega\left(\int_{E[u_h]}^{1} f^i(k)dk - \int_{\underline{k}}^{E[u_l]} f^i(k)dk\right)\right)$. Then, profit is higher in the separating equilibrium with the high type playing lockout if $\lambda^2\left(\int_{\underline{k}}^{E[u_h]} f^2(k)dk\ \left(1+\alpha\ \omega\ \int_{E[u_h]}^{1} f^2(k)dk\right)\right) < \sum_{i\in\{1,2\}} \lambda^i \int_{E[u_l]}^{E[u_h]} f^i(k)dk\left(1+\alpha\omega\left(\int_{E[u_h]}^{1} f^i(k)dk - \int_{\underline{k}}^{E[u_l]} f^i(k)dk\right)\right)$, which implies: $\lambda^2\left(\int_{\underline{k}}^{E[u_l]} f^2(k)dk\ \left(1+\alpha\ \omega\ \int_{E[u_l]}^{1} f^2(k)dk\right)\right) < \lambda^1 \int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+\alpha\omega\left(\int_{E[u_h]}^{1} f^1(k)dk - \int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)$. Rearranging, this gives $\frac{\int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+\alpha\ \omega\ \int_{E[u_l]}^{1} f^2(k)dk\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+\alpha\omega\left(\int_{E[u_h]}^{1} f^1(k)dk-\int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)} < \frac{\lambda^1}{(1-\lambda^1)}$, which is the incentive constraint of the lockout signaling equilibrium and always holds when the lockout signal is a PBE. Thus, there exists a region in which both separating equilibria are PBE's and if $\frac{a}{p}$ is sufficiently high, profits from the lockout equilibrium are higher.

Now, suppose the advertising cost is approaching the lower bound: $\frac{a}{p} \to \sum_{i\in\{1,2\}} \lambda^i \int_{E[u_l]}^{E[u_h]} f^i(k)dk\left(1+(1-\alpha)\omega\left(\int_{E[u_h]}^{1} f^i(k)dk - \int_{\underline{k}}^{Eu_l} f^i(k)dk\right)\right)$. Then, profit is higher in the lockout separating PBE if $\lambda^2\left(\int_{\underline{k}}^{E[u_h]} f^2(k)dk\ \left(1+\alpha\ \omega\ \int_{E[u_h]}^{1} f^2(k)dk\right)\right) < \sum_{i\in\{1,2\}} \lambda^i \int_{E[u_l]}^{E[u_h]} f^i(k)dk\left(1+(1-\alpha)\omega\left(\int_{E[u_h]}^{1} f^i(k)dk - \int_{\underline{k}}^{E[u_l]} f^i(k)dk\right)\right)$, which implies $\lambda^2\left(\omega \int_{\underline{k}}^{E[u_h]} f^2(k)dk\,(2\alpha-1)\left(1-\int_{\underline{k}}^{E[u_h]} f^2(k)dk\right) + \int_{\underline{k}}^{E[u_l]} f^2(k)dk\ \left(1+(1-\alpha)\ \omega\ \ \int_{E[u_l]}^{1} f^2(k)dk\right)\right) < \lambda^1 \int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\omega\left(\int_{E[u_h]}^{1} f^1(k)dk - \int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)$, which can be rearranged to: $\frac{\left(\omega \int_{\underline{k}}^{E[u_h]} f^2(k)dk(2\alpha-1)\left(1-\int_{\underline{k}}^{E[u_h]} f^2(k)dk\right) + \int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+(1-\alpha)\omega \int_{E[u_l]}^{1} f^2(k)dk\right)\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\omega\left(\int_{E[u_h]}^{1} f^1(k)dk-\int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)} < \frac{\lambda^1}{1-\lambda^1}$. This contradicts the IC of the lockout equilibrium $\frac{\lambda^1}{1-\lambda^1} < \frac{\int_{\underline{k}}^{E[u_l]} f^2(k)dk\left(1+(1-\alpha)\,\omega \int_{E[u_l]}^{1} f^2(k)dk\right)}{\int_{E[u_l]}^{E[u_h]} f^1(k)dk\left(1+(1-\alpha)\,\omega\left(\int_{E[u_h]}^{1} f^1(k)dk-\int_{\underline{k}}^{E[u_l]} f^1(k)dk\right)\right)}$.

Finally, noting that $\Pi_h|A$ decreases in a and $\Pi_h|R$ is not affected by a, we have shown that, for a region where both separating equilibria are PBE of the game, there exists some $a^* = \left\{a|p\sum_{i\in\{1,2\}}\lambda^i\left(\int_{\underline{k}}^{E[u_h]} f^i(k)dk\left(1+\alpha\,\omega\int_{E[u_h]}^{1} f^i(k)dk\right)\right)-a\ =\lambda^1 p\left(\int_{\underline{k}}^{E[u_h]} f^1(k)dk\ \left(1+\alpha\,\omega\ \int_{E[u_h]}^{1} f^1(k)dk\right)\right)\right\}$.

Proof of Proposition 3: The proof of existence follows similar steps as the proof of Proposition 1 and is presented in the WA. The comparisons of profits is straightforward and also presented in the WA.