Bundling information goods to reduce search costs

R. Scott Hiller* Jason Walter†

October 31, 2018

Abstract

We model the potential for a company bundling high search cost information goods to obtain an advantage over conventional individual sales. Using the streaming video industry as an example, we model the search costs associated with films. We find that bundling services can increase a user's utility from products with higher search costs, allowing a bundling service to charge a price premium in a fully served market when the reduction in search costs exceeds the consumer's disutility from the less preferred source. This implies the consumer's benefits are greatest when streaming services focus on products with high search costs. Using theatrical data and critical review data from film releases as a proxy for demand, the number of critical reviews as a proxy for search costs, and data from the Netflix streaming library, we find empirical evidence that films with higher search costs are more likely to be included in a streaming bundle.

JEL Codes: L2; L8; O33

*Department of Economics, Fairfield University, 1073 N. Benson Rd., Fairfield, CT 06824. Tel: (203)254-4000, ext. 2795, Email: rhiller@fairfield.edu

†Social Science Department, University of Wisconsin-Stout, 721 3rd St. E. Menomonie, WI 54751. Tel: (715)232-1225, Email: walterja@uwstout.edu

1 Introduction

As digital technology and distribution have improved, the costs of producing information goods have fallen, leading to an explosion in the number of these goods. This has created an environment where the consumer can find it difficult to choose what to consume. While the ability to search for products on the internet may work to reduce obstacles, the substantial increase in the number of these goods means matching consumers to their preferred choice still entails non-trivial search costs. Reducing these search costs can then become valuable for a firm selling information goods.[1] In this paper, we show that bundling information goods through digital distribution allows a firm to minimize search costs by taking advantage of asymmetric knowledge of preferences, providing a bundle that minimizes the distance from the target preference by consumer type, and using recommendations from similar consumers to direct consumption. The firm providing the bundle then has an advantage over firms providing conventional distribution of these goods.

Bundling can be used as a form of price discrimination (Stigler, 1963; Adams and Yellen, 1976), an advantage from varied willingnesses-to-pay from mixed bundling (McAfee et al., 1989), a barrier to entry (Nalebuff, 2004), or as a method of profiting off of negatively dependent products (Chen and Riordan, 2013). We consider the conditions under which bundling can be leveraged as a means to reduce search costs. When choosing the bundle, the firm may select products that have high search costs if purchased individually. The bundler then uses knowledge of preferences, as well as

[1] Search costs can be thought of in two ways which we consider equivalent in result. Either the difficulty in finding an optimal product reducing utility when consumed or some obstacle to finding the product resulting in a potentially sub-optimal purchase. Both reduce the final utility of the consumer.

careful composition to reduce search costs for the consumer. The reduction in search costs acts as an increase in utility of individual products, all else equal. The firm can then charge a premium for providing these goods over their sale price with full search costs, and profit from their ability to decrease search costs.

This bundling is particularly effective with information goods delivered digitally, where the firm receives feedback on a large scale. Information goods can be delivered with little to no marginal cost, the bundle can be quickly changed, and usage statistics analyzed. The firm can then alter the composition and make recommendations based on revealed or stated preferences.[2] The reduction in search costs allows consumers to find products they may have missed. In a large bundle this can be done with products that may have been in the long tail of preferences, and therefore ignored by traditional retailers. The bundle removes the necessity of extensive searching, effectively increasing utility.

The market we evaluate and create a stylized model for is bundled film streaming. Films are an experience good and consumers face search costs associated with finding ideal choices. If a consumer makes a decision to view a film, she searches for a film with the potential to provide a high level of utility. Despite widespread availability of data online, navigating the data, finding reviews to trust, and choosing a film is not a simple task. Reviewing all options, if she decides to watch she can choose between a single purchase and a streaming video bundle. We model both a partially and fully served market, to examine markets with and without the ability to expand depending on source. The model shows that higher search costs of films lead

[2]Revealed preferences are captured by analyzing selections and viewing information. Stated preferences can be captured by consumer reviews/rating of films.

to an increase in streaming subscribers in both market types, all else equal. This effect is strengthened as the provider of the bundle improves the scale effect on reducing search costs. We also derive the conditions under which streaming services charge a price premium for bundled products, a practice that can only occur when the reduction in search costs exceeds the disutility of source preference.

We analyze the streaming video industry to model this market because of its fit for distribution of information goods, as well as the availability of data. We use a uniquely collected dataset to test the predictions of our model, and although this paper focuses on the streaming video industry our model is easily generalized and can be modified to any firm offering a bundle of information goods. This can include news services, music delivery, and even booksellers, all industries that have also been significantly disrupted by the expansion of digital production and delivery provided by the internet.

Our theoretical model focuses on the ability of a streaming service to reduce search costs relative to traditional retail, while the empirical results demonstrate how the service can use the scale of the bundle to include these high search cost films that are potentially too far in the long tail of preferences to be successful via retail channels. Empirically, a preference for high search cost films in the streaming bundle would suggest the theory model is successfully explaining some of the advantage of these video streaming services. While the exact search cost of a good cannot easily be measured, we can proxy for search costs with consumer exposure from the initial theatrical run. Considering theatrical release may not be a perfect measure of search costs, we also include the number of critical reviews of a film. As the streaming video market may be considered a secondary market after theater

releases, consumers could be exposed to the films by the time they reach the streaming market.

We first account for the demand of a film through variables measuring the level of success of the initial theatrical run. For search costs we use data on theater distribution as a first proxy, where films distributed to more theaters will be better known to consumers, and therefore have lower search costs. Because distribution may at times be pulled from demand for the film rather than promotion, we use the number of professional critical reviews at the time of release as a second proxy. The more extensive the distribution of the movie, the more professional critics will review it for their audience, establishing a measure of the search cost at the time of initial release unrelated to demand and unchanged by the selection into a streaming bundle.

The data for the streaming bundle comes from the service Netflix, the largest streaming service in the United States in this period. The data includes all films in the bundle from September 2012 through September 2014 that were released after 2000, with matching characteristics from the theatrical run of the film. This set is combined with most films not included in the Netflix library, with some very obscure films included in Netflix that we were unable to match to any external characteristics that had to be excluded. This creates an attenuation bias in our results, and the effects in this paper should be seen as a minimum measure of the ability of the bundle to reduce search costs.

The success of Netflix may be tied with the ability to identify high search cost films of substantial utility to consumers. Results from the Netflix selection model show that once demand has been controlled for by gross receipts of the theatrical run, higher search costs (represented by a smaller theatri-

cal release or fewer professional critical reviews) increases the likelihood of selection into the streaming library. User reviews, which change well after the theatrical run, do not have the same effect on inclusion. The number of user reviews increases for films selected by Netflix, indicating a reduction in search costs for those films.

2 Background

The seminal work on information as it relates to search was first discussed by Stigler (1961), who initially discussed the consumer's quest to identify lower prices. Subsequently, the theory of search cost and product attributes has expanded tremendously. Nelson (1970) separates products into two types: search and experience, while Darby and Karni (1973) expands the discussion of product attributes by including credence goods.[3] Search goods are known for the ability of consumers to differentiate based on quality and price. Experience goods are those that require consumption to determine their value (e.g. CDs, canned tomatoes), which then allows consumer to identify specific products or brands with the highest value for future purchase. Credence goods contain attributes that must be identified by alternative source (e.g. certified organic, sustainably produced certifications).

By this definition, many digital information goods (including films) would most closely resemble experience goods as the consumer's utility cannot be determined until after the product is used. However, films create an additional challenge for consumers; the utility received from one film is not necessarily representative of a studio's other films, making it difficult for consumers to identify the ideal film to watch based on producer characteristics.

[3]For a discussion of all three product attributes see Wessells (2003).

This creates the common problem for producers associated with experience goods, they are faced with trying to attract consumers who cannot easily identify the value of their product.

Digitization can exacerbate these challenges. Information goods can be shared and rented, resulting in decreased sales and increased prices (Varian, 2000). Furthermore, the digitization allows for this exchange to occur electronically. Online exchange has yielded several benefits for consumers, as noted by Bakos (1997) "electronic marketplaces reduce the inefficiencies caused by buyer search costs, in the process reducing the ability of sellers to extract monopolistic profits...."[4]

Ideas about the potential for the reduction in search costs in online markets are not new. Research began when online markets started spreading broadly, as seen in papers such as (Bakos, 1998; Lynch and Ariely, 2000). While the ability to search may lower the cost of commodity goods, determining which information good is preferred can actually be made more difficult by an increased supply. This may be seen as consumers now exploring the "long tail" of goods, as in Brynjolfsson et al. (2011). In markets where consumption decisions consider the long tail, reducing search costs becomes more valuable.

The streaming of information goods has further changed the conditions and distribution in these markets. Streaming creates a nondurable option between purchase and the time-restricted renting of the durable version of the product. In addition, the "economics of aggregation" created by bundling provide advantages in upstream competition for content, and as a deterrent to entry (Bakos and Brynjolfsson, 2000). In addition, bundling

[4]For a discussion of reducing search costs for illegal downloads see Chang and Walter (2015).

information goods has the benefit of an extremely low marginal cost, and can be used to group a large number of unrelated products (Bakos and Brynjolfsson, 1999).

After films have exhausted demand in theaters, producers are left with three main options to generate additional revenue: provide a retail (digital or physical) version of their product (i.e. unbundling), sell the film rights to a bundling service (i.e. pure bundling), or both (i.e. mixed bundling).[5] As noted by Chuang and Sirbu (1999) pure bundling does not necessarily dominate unbundling in the context of certain information goods, but a mixed bundling strategy is dominant.

Our study expands the work of Harris and Blair (2006), who test the theory that bundling can reduce search costs for stereo equipment in an experimental setting. They find that the bundle is appreciated for its convenience over purchasing component pieces and no informational leverage is obtained in this experiment. We expand the discussion to information goods which have unique characteristics (non-physical and non-traditional experience goods) and we are able to model the effects of bundling services on search cost. This also differs from Brynjolfsson et al. (2011) in that our reduction in search costs comes from bundling advantages rather than strictly search. In addition, our analysis treats reductions in search cost as a result of a service's ability to leverage user information.

The evidence shows that reduced search costs can benefit producers when combined with the ability to bundle. Unlike Chuang and Sirbu (1999), the advantage is not absolute, but relative to retailers without a similar ability to use the direct and indirect advantages of the bundle. Few papers

[5] Other options exist, such as licensing rights to TV networks.

specifically examine a streaming bundle. In the music industry Hiller and Walter (2015) comment on how production methods may change due to the rise of the streaming music industry and Aguiar and Waldfogel (2015) examine the effects of Spotify on conventional sales. With regards to streaming video Hiller (2017) uses Netflix data to explore the characteristics important to operating a profitable mixed bundle of information goods, with a focus on the negotiation between streamer and rights holder. The empirical focus of this paper is not concerned with the negotiation of the two, but the potential effect of search costs on the streaming format when compared to conventional video sales.

Netflix operates a DVD-by-mail and parallel streaming video option for viewing films. Our focus is the streaming library or the videos on demand (VOD). The streaming option is seen as the future of business for Netflix.[6] These services are in some ways substitutes, although the DVD library has more choices at the expense of availability, where the number of films available at one time are limited by prior consumer selection and the United States postal service. In contrast to the DVD service, the entire VOD library is available for consumption by subscribers at any time.

Netflix can reduce the search costs of consumers by using the streaming bundle as a source of revealed preferences, and directing consumers to films carefully selected to viewer type they may have otherwise missed. This is done by using data on what is viewed, how often it is viewed, what is watched completely, and by matching based on preferences. Recommendation engines use the ratings and habits of consumers to suggest other goods

[6]Many articles speculate on this, and Netflix executives acknowledge its importance. For examples see: http://www.bloomberg.com/bw/articles/2013-10-21/netflix and http://www.wired.com/2011/07/netflix-fees-increase-dvd/. Additionally, Netflix has expanded the VOD library to 190 countries.

a consumer would like in news (Thorson, 2008), e-commerce, and streaming music.[7][8] The combination of the recommendation engine with such a wide array of films in the bundle allows the Netflix programmer to reduce search costs within their own streaming platform. This is not something a physical retailer has the same ability to do. While they can monitor purchase or rental behavior, they cannot monitor viewing. Even if a retailer could obtain the same amount of data, limited retail space likely means less options and more focus on lower search cost films.[9]

Streaming video is increasingly important in the film market with millions of subscribers, where the rest of the industry has experienced declining physical sales and increased competition between streaming platforms. Any questions about the significance of the streaming market can quickly be answered by an examination of revenue statistics. The relevant period for this paper is 2012 through 2014, and 2014 was the first year in which American consumers spent more on digital than physical films.[10] The growth has only accelerated since then. Netflix was the largest streaming provider in this period, dwarfing both Amazon and Hulu, primary competitors to the Netflix bundle.[11]

[7]For an example of e-commerce see: http://www.wired.com/2015/04/now-anyone-can-tap-ai-behind-amazons-recommendations/, Accessed: 7/30/2016.

[8]For an example of streaming music see: http://qz.com/571007/the-magic-that-makes-spotifys-discover-weekly-playlists-so-damn-good/, Accessed: 7/30/2016.

[9]This is likely one of several reasons for the collapse of the rental market in the United States, see http://variety.com/2016/digital/features/redbox-business-model-doomed-1201706612/. Accessed: 3/4/2017.

[10]http://www.ibtimes.com/home-entertainment-2014-us-dvd-sales-rentals-crater Accessed: 07/26/2016.

[11]http://variety.com/2013/digital/news/netflix-video-usage-more-than-10-times-amazon-and-hulu-combined-1200812004/, Accessed: 7/26/2016.

3 A model of search costs

Throughout this model, we use terminology for the film market, but as stated before this can apply generally to any information good that may be bundled and contains a range of search costs. We model the consumer decision on whether or not to watch a film, and then the option between streaming and purchasing a video. When a representative consumer, x, chooses a format to watch a film, we assume she watches it once and depletes the information good of its utility.[12] Single purchases in this model reference both digital and physical options concerned with film rentals rather than ownership, where one option is assumed for convenience. This fits best with the single viewing model.[13] As noted by Bran and Matula (2014), purchasing and streaming each provide different benefits to users.[14] Resolution, convenience, selection, and an aversion to subscription models can all drive a consumer's decisions. This means that consumers may have a strong preference for one format over another. We acknowledge this preference by distributing users on a unit interval, and denote location by x, where $x \in [0, 1]$. Users with a stronger preference for retail (streaming) are located closer to zero (one).

We model this decision in both a fully and partially served market for the good. The fully served option represents a market where every consumer watches a film, but must choose between formats. The partially served market can represent either a market where the consumer has only one format available, or there exists some unserved population with both options available. After a consumer decides to watch a film, there are two potential

[12] For children's films this may not be a good assumption (Collins et al., 2009).

[13] Consumers that desire multiple viewings over an extended period of time are likely to purchase a durable option.

[14] Bran and Matula (2014) discusses this is in the context of music, but the benefits apply to most information goods as well.

options, either they already know what they want to watch (in which case search costs are zero), or they will have to search for their selection (which we model). Generally, a consumer has the option of viewing from either the available bundle on the streaming service or the product available for sale.[15] From the choices available, she must decide whether to watch any of the available films.

The consumer attempts to select a film that will maximize utility given the available information. The consumer selects film, r, which she believes is her best possible choice. The realized utility from the film is equal to the value of the ideal choice minus the distance of her selection from the ideal film. This distance is due to search costs attributed to imperfect information, so if there were zero search costs the consumer would always choose the ideal film. The expected utility and search costs can lead consumers to view a film that is suboptimal. Consumers that choose to watch a film must make a single purchase or buy (or renew) a streaming subscription.[16] Therefore, the expected utility that a representative consumer, x, receives from each choice of a random movie, V_r is

$$\mathrm{U}(x, V_o) = \begin{bmatrix} \left(V_o - E\left(\frac{1}{(\beta_{r,A})}\right)^2\right) - x\tau - P_A & \text{if purchasing video} \\ \left(V_o - E\left(\frac{1}{\beta_{r,F}}\right)^2\right) - (1-x)\,\tau - \frac{P_F}{N} & \text{purchasing streaming subscription} \\ 0 & \text{if not purchasing either format} \end{bmatrix}$$

Consumer utility derives from the film chosen, where V_0 is the level of utility that would be provided from the ideal choice in the set. The expected distance from the ideal film is denoted by $E(\frac{1}{\beta_{r,*}})$, which decreases

[15] All films modeled are assumed available on either format, but many films are not available on both formats. The partially served model, with one format option, can be used to model that decision.

[16] The streaming subscription decision will also depend on how many other films the consumer intends to buy.

the consumer utility at a quadratic rate. $1/\beta_{r,*}$ represents the distance of movie r (which the consumer believes is the best option from those films available) from that ideal choice, which differs by format: "A" denotes retail and "F" denotes streaming. The expression $\beta_{r,*}$ is a random variable with a chi-squared distribution, implying that within a set of movies, the consumer will not always choose her ideal film.[17]

The number of films (N) a subscriber of a streaming service watches within an interval is fixed and greater than or equal to one. When a consumer signs up for a subscription she can watch as many films as time permits, however, a forward looking consumer will still evaluate the utility of all the movies she expects to watch within the subscription length and compare it to the number of videos she plans to purchase. This is equivalent to multiplying the utility of each format by the number of films watched (N) within a given period, without loss of generality, our analysis focuses on a single film within her choice set. P_A and P_F are the prices of a video purchase and a streaming video subscription, respectively.[18] The intensity of a user's source preference is represented by τ; we assume that the price of watching a film is greater than any user's source preference, thus $\tau < P_A$. We acknowledge that in many cases the bundled library may be limited relative to the retail market, however, as long as the streaming service can provide a film (or good) that more closely matches the user's preferences our assumptions are appropriate.

The average distance of a film from the ideal choice depends on the

[17] A chi-squared distribution is assumed due to consumers familiarity with her own movie preferences, however, other distributions could be used.

[18] Note that $\frac{P_F}{N}$ is the price of the film when we incorporate subscription use. We would expect that the utility of the films watched using the subscription would be relatively similar. Modeling ad based streaming services (like Crackle), would require making the following substitution: $\frac{P_F}{N} = P_{AD}$, where P_{AD} is the "cost" of watching an advertisement.

quality of information, implying better information will reduce search costs and reduce the distance between selected films and the ideal. Therefore, average utility declines as search costs move the selections further from V_0. All information available to help retail consumers make a selection is also available for streaming consumers. In addition, streaming services have additional personalized user information, which allows them to better identify films that match a user's preferences. Streaming services also obtain information for a variety of other consumers, allowing them to better identify users with similar tastes, and further leverage user information to improve recommendations.

Let distance from the expected movie be $E(\frac{1}{\beta_{r,A}}) = \gamma_r$ when purchasing, and $E(\frac{1}{\beta_{r,F}}) = \gamma_r \sqrt{1 - N\alpha\,(1 - x^S)}$ when streaming, where the search costs for a film are represented by γ_r. Note that $(1 - x^S)$ represents the total quantity of users of the streaming service, and α scales the number of users and views into a reduction in search costs.[19]. When we replace the expected distance of a film with the search costs for each format a decrease in search costs (γ_r) causes the expected utility of the selected movie $(V_o - \gamma_r^2)$ to increase through better choices.[20]

Traditional retailers use aggregate consumer information to select their offerings, while streaming services can make additional reductions in search costs due to the service's ability to obtain more detailed viewing information. By using this data, the streaming service will always be able to make better recommendations, but the reduction in search costs will depend on

[19] If a service knew exactly how to optimally identify the videos for users to watch, then search costs would be zero. This means that the reduction from suboptimal search (α), must satisfy $\alpha < \frac{1}{N(1-x)}$

[20] The expected value of an inverse chi-squared distributed random variable is $\frac{1}{v-2}$, where v denotes the mean of the (chi-squared distributed) random variable. For simplicity, we let $\gamma_r = \frac{1}{v-2}$ and require $\gamma_r > 0$.

information availability. Therefore, an increase in the number of movies a user watches (N) provides additional user information, thereby allowing better matching by the service. The combination of this enhanced data is combined with the scale of the bundle that directs consumers to films that may not generate demand in a traditional retail environment.

With the search costs accounted for in these parameters, the utility of any selected movie for the representative consumer x becomes[21]

$$U(x, V_o) = \begin{bmatrix} V_o - \gamma_r^2 - x\tau - P_A & \text{if purchasing video} \\ V_o - \gamma_r^2 - (1-x)\left(\tau - N\alpha\gamma_r^2\right) - \frac{P_F}{N} & \text{purchasing streaming subscription} \\ 0 & \text{if not purchasing either format} \end{bmatrix}$$

For streaming users, $(1-x)\left(\tau - N\alpha\gamma_r^2\right)$ represents the change in utility from source preference and reduced search cost. It is important to note that x represents the location of the user on the unit line, but if x is the marginal consumer, it can also be used to represent an interval of users. For example, $(1-x)(\tau)$ represents the magnitude of consumer x's preference for the streaming format, whereas $(1-x)\left(-N\alpha\gamma_r^2\right)$ represents the reduction in search costs associated with $(1-x)$ consumers using the service.

Let K represent the source and search cost differential, K is

$$K = \left(\tau - N\alpha\gamma_r^2\right) \tag{1}$$

Source and search cost differential is an important consideration for the marginal consumer as well as the streaming provider. The value of the differential shows whether users are more attracted to streaming due to the

[21] We assume regardless of search costs and source preferences at least one individual will purchase videos, thus $V_o - \gamma_r^2 - P_A > 0$, and at least one individual will purchase a subscription, thus $V_o - \gamma_r^2 - \frac{P_F}{N} > 0$.

format's features or service it can provide. The model accounts for the fact that the reduction in search costs ($N\alpha\gamma_r^2$) comes from improved matching by the service. This reduction will be compared to the source preference to see which is greater: disutility from using an inconvenient format or the reduction in search costs. When $K > 0$, the source preference is greater than the increased utility derived from the reduction in search costs, and the opposite is true when $K < 0$.

3.1 Partially served market

We begin by modeling a market where either some consumers choose not to watch a film using either option, which represents a potential growth market, or only one source is available to consumers. In this partially-served market, the marginal consumer is indifferent toward purchasing a film (since x^P<1), which implies that $V_o - \gamma_r^2 - x\tau - P_A = 0$, and the number of consumers that purchase the video is represented by the interval [0,x^P], where

$$x^P = \frac{V_o - \gamma_r^2 - P_A}{\tau} \tag{2}$$

This shows that the number of purchasers falls as search costs (γ_r), price of purchasing (P_A), or the source preference (τ) increase.[22] Purchasers increase as utility from the ideal option, V_o increases. As a result, when the utility derived from films increases, so does the number of consumers. These results are expected for the film industry. Using a similar approach, we can derive the marginal streaming subscriber in a partially-served market (this requires that x^S>0) as

[22]Relevant derivatives are available in Appendix 1

$$(1 - x^S) = \frac{V_o - \gamma_r^2 - \frac{P_F}{N}}{\tau - N\alpha\gamma_r^2} \tag{3}$$

The number of streamers rises as search costs (γ_r), the firm's ability to reduce search costs (α), and number of streams per consumer (N) rise. Streaming use falls as the price of streaming (P_F) or the source preference (τ) increase. The number of streamers also increases as utility from the target option, V_o increases. As before, we can also determine the overall quantity of streaming subscriptions, which is represented by the interval [x^S,1]. If we assume that there exists a market for streaming, it generates the necessary condition that $\tau - N\alpha\gamma_r^2 > V_o - \gamma_r^2 - \frac{P_F}{N}$. This implies that $K > 0$ for the marginal consumer in the partially served market, leading to our first proposition

Proposition 1 *In a partially served market, streaming users' source preferences strictly dominate gains from reduced search costs provided by the streaming service.*

To fully understand the impact of this proposition, we need to evaluate the effects on the marginal consumer. When a new consumer signs up for the streaming service, two things are driving the decision. A new consumer will have a larger source preference than the existing consumers, and she derives lower utility from the subscription because of this. At the same time, as she uses the streaming service the service's ability to reduce search cost increases, thereby raising the utility of all streaming users. Comparing these marginal effects we can identify the mechanism affecting demand in the partially served market. If it is always the case that the additional utility from reduced search costs derived from an additional user's information

are greater than the next consumer's (disutility from) source preference, it implies that if one consumer signs up for the streaming service then all consumers will sign up for the service. It also implies that the market will no longer be partially served.

This is similar to the results discussed by Shy (2001), who identifies when product differentiation has a smaller influence than network effects, all consumers will purchase the product. Product differentiation and consumer heterogeneity have similar effects on utility. In addition, the reduction in search costs resulting from an additional consumer's information is essentially a positive network externality, because as more consumers stream the service's ability to identify the ideal film in a set improves for all consumers using the service, which increases streaming users' utility.

The source preference is the difference between the utility from streaming for the most fervent streamer and the disutility of the format provided to the biggest streaming critic. Our result shows that the utility from source preference must be larger than the utility derived from the service's ability to reduce search costs in a partially served market. It also means that no matter how good a service is at reducing search costs, there will always be someone who refuses to stream content. Furthermore, in a partially served market the preference for the streaming format is the main driver of video consumption as opposed to the benefits the service provides of reduced search costs. The demand in the market is from the introduction of a new format which appeals to a previously unserved population. It should be noted that reducing the direct search costs (γ_r), ceteris paribus, will still increase a streaming user's utility.[23]

[23]Comparing streaming from higher search cost films (holding all else equal) requires comparing the utility function with different (direct) search costs. Let γ_H and γ_L represent

Streaming services gain from the knowledge of consumer preferences to reduce search costs. By using these direct and indirect methods, a streaming service can employ the user's revealed preferences to improve recommendations. This simplifies the search for films within their offerings, and allows the service to scale down search costs for its users. For this reason, the number of streaming subscribers, (1-x^S) will increase at the expense of other options as films' search costs increase. Leading to the second proposition

Proposition 2 *In the presence of higher search costs, the number of streaming subscribers and overall utility of consumers increases relative to other options due to the service's ability to utilize revealed preferences and direct selections within the bundle to reduce search costs for consumers.*

This result seems counterintuitive at first glance, and requires distinguishing between the two effects occurring. It is true that the direct effects of higher search costs are to lower consumer utility. However, as the search costs increase, the reduction in search costs provided by the service becomes more valuable. If it becomes more challenging for users to identify the ideal film, the reduction in search cost becomes more important. As a result, the streaming service's ability to reduce search costs indirectly through user information improves, and therefore the number of subscriptions is increased. The service's ability to reduce search costs is greater in the presence of high search costs relative to low search costs.

As expected, there is a limit to how challenging the ideal film can be for

high and low search costs. Comparing the utility in both scenarios (and holding the number of users constant), yields: $[V_o - \gamma_L^2 - (1-x)(\tau - N\alpha\gamma_L^2) - \frac{P_F}{N}] - [V_o - \gamma_H^2 - (1-x)(\tau - N\alpha\gamma_H^2) - \frac{P_F}{N}]$, which can be reduced to $[(\gamma_H^2 - \gamma_L^2)(1-x)N\alpha]$. Assuming positive search costs or $\alpha < \frac{1}{N(1-X)}$, implies that $(\gamma_H^2 - \gamma_L^2)(1-x)N\alpha > 0$. We will show in the next section that the indirect effects of high search costs may change this effect.

the services to effectively reduce search cost. As search costs rise, the utility from streaming the video reaches a threshold where the search cost exceeds either: 1) the film's utility after paying for the service, or 2) users' source preferences. At this point, the utility becomes negative and no consumer will stream the video.[24]

The remainder of the results from marginal changes in variables are expected. As the average user views more films (N), the price per film decreases, thereby increasing subscriptions. As the price of the subscription (P_F) and source preferences (τ) increase, streaming subscriptions decrease. Finally, an increase in utility from the ideal film (V_o) causes subscriptions to increase.

3.2 Fully served market

In the fully served market, all consumers watch the film, but must still choose which format to employ. Using the same utility function as before, we identify the marginal consumer (denoted x^*), who is indifferent between purchasing and streaming the video. Setting the utility of each source equal yields the marginal consumer in the fully-served market as

$$x^* = \frac{\left(\frac{P_F}{N} + \tau - P_A - N\alpha\gamma_r^2\right)}{2\tau - N\alpha\gamma_r^2} \tag{4}$$

In this equilibrium, we can test the marginal effects of changing the relevant variables.[25] Since users closer to zero (one) on the unit interval prefer purchasing (streaming) films, the location of the marginal consumer

[24]The threshold is reached when either: $\gamma^2 \geq \frac{\tau}{N\alpha}$ or $\gamma^2 \geq V_o - \frac{P_F}{N}$ occurs.
[25]Relevant derivatives are available in Appendix 1

can be used to identify changes to the quantity demanded of each source. Therefore, we can conclude that increasing search costs (γ_r), the number of streams per user (N), improving the reduction in search costs (α), and higher purchase price (P_A) all decrease retail purchases of films (x). At the same time, a higher subscription price (P_F) increases retail purchases (x).

Since this is a fully served market, the opposite holds for streaming subscriptions in every case, and the demand for streaming in a fully served market is simply (1-x^*), or

$$(1 - x^*) = \frac{\left(P_A + \tau - \frac{P_F}{N}\right)}{2\tau - N\alpha\gamma_r^2} \tag{5}$$

The effects of users' source preference on the demand of each format is not as clear. User source preference is directly related to the features of each format. Examples may include streaming becoming more convenient on more devices or improving resolution, or retailers selling films with alternate editing, special access, and additional features. With this differentiation we would expect source preferences to intensify, affecting demand.

To identify the effects on demand we must determine which, if any, format attracts users when source preferences intensify. If $\alpha N\gamma^2 + 2P_A - \frac{2P_F}{N} > 0$ it implies that a larger source preference (τ) will increase demand for the retail format (x^*).[26] This means that $2\tau > \frac{P_F}{N} + \tau - P_A > N\alpha\gamma_r^2$, or that

$$K > P_A - \frac{P_F}{N} > -\tau \tag{6}$$

[26]Note, that in order for both formats to be available in the market, it must be that 1 $> x^* > 0$.

Let $\frac{P_F}{N} = P_A + \varepsilon$, where ε is a streaming premium if greater than zero and a discount if less than zero. This substitution allows us to focus on the role of pricing each format, which is a variable managers of both formats can directly control in the marketplace. In addition, we can rewrite $\alpha N\gamma^2 + 2P_A - \frac{2P_F}{N}$ as $\alpha N\gamma^2 - 2\varepsilon$, which implies

Corollary 3 *If the streaming price premium is less than half the reduction in search costs, then demand for the retail format will increase as consumers' source preference increases.*

From these results, we can also derive the potential benefit to streaming services from search costs. Substituting the streaming price premium into equation 6 provides $\tau > \varepsilon > -K$, or $\tau > \varepsilon > N\alpha\gamma_r^2 - \tau$. Since $\varepsilon > 0$ implies that a streaming service is charging a premium for the average stream of a film relative to the retail purchase price, this leads to the following result

Proposition 4 *In order for a streaming service to charge a premium for the average stream over the purchase price of a film, the reduction in search costs of the film must exceed the user's source preference.*

This result implies that streaming services can raise the premium on their service only if it is able to identify films with high search costs (such that K < 0) and match them to the appropriate consumer. More formally, substituting $\frac{P_F}{N} = P_A + \varepsilon$ into equation 5 yields

$$1 - x^* = \frac{(P_A + \tau - P_A - \epsilon)}{2\tau - N\alpha\gamma_r^2} = \frac{\tau - \varepsilon}{\tau + K} \tag{7}$$

For any given film, $\frac{\tau-\varepsilon}{\tau+K}$ represents the number of streaming consumers,

so in order to increase the streaming price premium and maintain market share, the search costs of the film must increase in excess of the premium increase. This reinforces the idea that streaming services have an advantage by bundling films. Search costs obscure the true value of a film, and further decrease demand for the retail format. The bundling by the streaming service allows for convenient distributions of films, but also more subtly matches films to users, thereby decreasing search costs and allowing the film to reach its true value by having it be viewed by the appropriate audience.

Our theoretical framework has provided several important results. For streaming services who have to construct a bundle to offer consumers and optimize their recommendations, we've shown that the service's recommendation are paramount to attracting consumers. However as a profit maximizer, the firm's costs for obtaining films rights is principal to their profit. If we assume that the retail price of film (P_A) is exogenous, then films with low search costs will be more profitable for studios in the retail market.[27] As a result, we expect that obtaining the rights to high search cost films would be significantly cheaper for a streaming service. Furthermore, we expect that a streaming service's desire to decrease search costs for consumers, while minimizing their own costs, would make high search cost films a practical and profitable option. Therefore, we expect a higher proportion of high search cost movies within a streaming service's bundle.

[27]Generally, new films (for sale or rent) are homogeneously priced at most retailers (digital or physical). The price of "classics" are generally homogeneous as well.

4 Empirical model

Given the predictions of our theory model, we look for evidence of search costs influencing film selection by streaming services. Statistical analysis suggesting regular bundling of films that struggle in traditional distribution would suggest an advantage for the streaming bundle with these films. For this, we use information from Netflix, the largest streaming service in the United States. In addition, we include a very large number of films that were not in the Netflix library that could have potentially been selected during this period.[28] This data is included in order to model the selection decision, with box office and rating data for each.

The empirical methods build on results found in Hiller (2017), which uses a negotiation model to examine the attributes of movies important to Netflix and the rights holders of films. That paper is devoted to finding the conditions that allow for a profitable agreement between the two parties considering the necessary demand effect for Netflix, and the potential for displacement for the rights holder. The demand characteristics of films used in our paper match what was found to be important in that paper, allowing for the effect of search costs of selection to the Netflix library to be measured after accounting for a film's popularity and the difficulty of the licensing negotiation.

Our primary concern is evidence of the effects of search costs, requiring measures for any obfuscation to consumers. First, we use number of theaters as a measure of the depth of the market from the theatrical run. Films that are distributed more broadly will also have greater promotion and lower

[28] 3,148 films, or approximately 56% of our sample were never included in the Netflix library.

search costs. A wider distribution network means more theaters, likely more promotion, and lower search costs. We recognize that there is a potential endogeneity concern of the number of theaters, and Netflix may use the measure in their demand so we also use the number of professional critical reviews of a film.

Professional critical reviews are released with the initial theatrical run. We assume that films with a larger promotional budget and more name recognition will be reviewed more broadly, a function of their lower search costs. More reviews should also make it more likely for a consumer to learn about a film, further reducing search costs. The fewer the number of reviews, the greater the expected search cost. This can be used as a proxy because films are reviewed by professional critics on their initial theatrical release.[29] From the data there is a clear relationship between the number of professional reviews and theatrical success, but the number of reviews and the rating of quality of the film do not show the same relationship. This leads us to the conclusion that the number of professional ratings serves as a proxy for search costs if quality is also appropriately controlled for. In contrast, the number of amateur user ratings will increase beyond the initial theatrical run, and should increase more quickly as search costs are lowered.

It is difficult to imagine a scenario where the number of theaters misstates search costs. Widespread distribution would usually follow considerable promotion. If a film is an unexpected success, films could be pulled into many theaters, this is presumably done by word of mouth if not promotion, both a sign of low search costs. Still, controlling for revenue as well as using the number of theaters should help to alleviate concerns of a bust failing to

[29]Professional critics release their reviews before success or failure is obvious, and the number of reviews does not change dramatically after that success is revealed.

reach the public or a film in a small number of theaters finding an enthusiastic repeat audience. The number of critical reviews could be subject to more concern about understating search costs. Critics may be more inclined to review films they enjoy, even if that doesn't find a widespread audience. Still, the inclusion of theatrical revenue with the number of critical reviews should help to alleviate concern about over or understating search costs.

We must also account for expected demand with several other measures from the theatrical run, including genre, studio, and year of release. We use the assumption that more success in the theatrical run implies a higher demand for viewing in the Netflix library. Box office data acts as an important measure of the established demand for each film before potential inclusion in the Netflix library.

Additionally, we employ two measures of professional critical ratings, released with the initial theatrical run of the film, and amateur user ratings, rated as users find the film as additional measures of demand. The number of ratings is used for search cost while the reviews themselves act as a measure of the quality of a film that is not necessarily causal when discussing the theatrical success of the film. This can be thought of as vertical differentiation, which provides another variable on which Netflix programmers may choose films. With this data, we model the selection decision for inclusion into the Netflix library. We pool the dataset, dividing films into those selected at any point in this period for the Netflix library, and those never included in the library. This choice ($Selection$) is determined by

$$Selection_{ij} = \gamma_0 + \beta' x_{ij} + \delta' z_{ij} + \gamma_1 SearchCost_{ij} + \alpha_j + \epsilon_{ij} \qquad (8)$$

Where x_{ij} is a vector of characteristics related to the release of film i by studio j, including genre and year released, and the average critical rating of the film from our review aggregators. The studio indicator (α_j) picks up any increased likelihood of a film being selected due to the rights holder licensing many films at once or not at all, and a variable for the year of release establishes any increased likelihood of choice from newer films. The commercial success of the film is captured by vector z_{ij}, which contains information on the release and gross data of the film. Search cost is measured in two ways, by the number of theaters the film is released into and by the number of professional critical reviews the film receives. Equation 8 is estimated as a pooled logit with fixed effects for studio and genre.

4.1 Data

We used several sources to compile the necessary film data. We collected all of the offerings of Netflix contemporaneously as no historical database of the library seems to exist beyond the company, using the website Instantwatcher.com to scrape the entire listing of movies and television shows on a weekly basis. The title of the offering, year, Motion Picture Association of America (MPAA) rating, status as a movie or television show, and average Netflix consumer rating on a 1 to 5 scale were collected for each listing. Although television shows, comedy and music specials were collected, they are excluded from this dataset as they are not the focus of this paper. The dataset includes 104 weeks of the Netflix VOD library from September 6, 2012 through September 4, 2014. Films range in release from the 1930s through 2014, but in order to appropriately match each to box office revenue and film ratings any title released before the year 2000 is dropped.

Through the traditional windowing process of films, rights holders restrict the distribution to one format at a time, beginning with theaters (Nelson et al., 2007; Waterman and Weiss, 2010). Films are then released through additional channels according to expected profitability in a method of temporal price discrimination. For that reason, films released in 2014 are dropped from our sample for potential selection as the windowing process makes their inclusion in Netflix unlikely for reasons unrelated to demand or search costs. These films have likely been released too recently to have made it through the windowing process.

Two databases were used to match characteristics to the recognized movie releases in the library: Box Office Mojo and The Numbers.[30] From these we included information on the studio, genre, gross earnings in theaters, and the number of theaters in which the title was released. All dollar values in this paper are adjusted for inflation using the CPI. The dataset excludes Netflix original productions as inclusion of this material was always intended, and remains a given. The theatrical data is used as a measure of the demand for films on Netflix.

Film review data comes from two aggregators of critical reviews, Metacritic and Rotten Tomatoes.[31] Metacritic creates a score based on a 100 point scale for films released. The score is created with a process that "carefully curate[s] a large group of the world's most respected critics, assign scores to their reviews, and apply a weighted average to summarize the range of their opinions."[32] The Rotten Tomatoes score is similar, but the

[30] Available at http://boxofficemojo.com and http://the-numbers.com Last accessed 07/05/2016.

[31] Available at http://metacritic.com and http://rottentomatoes.com Last accessed 07/05/2016.

[32] http://www.metacritic.com/about-metascores, Last accessed 07/05/2016.

rating focuses on a binary like or dislike of the film by the critic. The score "represents the percentage of professional critic reviews that are positive for a given film or television show."[33] The Rotten Tomatoes website provides not only the rating, but the number of reviews by professional critics, the review score by amateur users of the website, and the number of amateurs that have reviewed the film. We use the RT data more extensively, while the Metacritic score is used primarily for robustness to ensure the coefficients on ratings do not change dramatically.

The dataset is restricted to films that could be matched on theatrical run data and critical ratings, limiting the pool of films to those films recognized by the extensive databases from which we collected information. This means that some of the smaller films included on Netflix are excluded from our dataset, films that are not recognized in at least one of our databases. This must lead to the expectation that search costs are greater for these titles when compared to the titles in our dataset, as too few professional and amateur critics have reviewed them to appear on either website. The necessary exclusion of these films means the empirical results likely contain an attenuation bias toward the effects of search cost on selection and our results of the relationship of search cost in the bundle represent a minimum.

Table 1 gives the summary statistics for important variables. The variable $Meta$ provides the Metacritic rating for a film, RT represents the Rotten Tomatoes score, and $AverageRating$ is used an average of the two professional critical scores, where either score may be used if no value could be found for the other service. $AmateurRating$ is the Rotten Tomatoes score created by users of the website rather than professional critics. The possible

[33] https://www.rottentomatoes.com/about/, Last accessed 07/05/2016.

rating for each measure is on a scale of zero to 100. There are many more amateur reviews (*AmateurReviews*) for most films than there are professional (*ProReviews*) and the number of reviews and score for the amateur values are collected from a single period of time, but are subject to change as additional reviews come in.[34] *Theaters* is the number of theaters a film is in, recorded in thousands, and *Gross* is the gross revenue of a film from its theatrical run, recorded in millions.

4.2 Results

Table 2 gives the marginal difference in probability of inclusion in the Netflix library associated with a one unit change in the independent variable. The two measures of professional ratings are used in columns 1 and 2, with our preferred measure of *AverageRating* used for columns 3 through 5. The demand proxy variables are included in each regression, and the measures of search costs are used separately in columns 1 through 4, and jointly in column 5.

The first proxy for search costs is number of theaters, *Theaters*, found in columns 1 through 3. None of the three critical ratings used have a significant impact on inclusion in the Netflix library with *Theaters* as the only search cost. This seems to indicate that professional critical reception of a film does not influence the selection. Established demand is important, however, as the coefficient *Gross* indicates. For every million dollar increase, a film is approximately 0.35 percent more likely to be selected. This effect is eliminated, however, for very high gross films as the square term of this variable, *GrossSquared* is negative and significant. Increased potential de-

[34] These data were collected from June 2 through June 5th, 2016.

mand explains why newer films are substantially more likely to be selected for the library, represented by the coefficient on $Year$.

The coefficient on the first search cost, $Theaters$, shows a substantial negative effect on the probability of selection. For every 1,000 additional theaters a film is in, the likelihood of inclusion falls by approximately 10 percent. Considering the gross revenue of films is accounted for in these regressions, this coefficient can be thought of as the effect of the distribution of the film rather than a demand effect. The broader the distribution of the film, the lower the search costs. Those lower search costs mean less of an opportunity for Netflix to increase value through bundling, and imply they are less likely to be included.

Still, it is possible that some films may be pulled into more theaters through demand, which would make theaters a demand variable. For that reason, columns 4 and 5 include regressions using the number of critical reviews as a proxy for search costs. In these regressions nothing changes dramatically among the coefficients on the demand variables, with the exception of the coefficient on $AverageRating$, which has become significant. When search costs are properly accounted for, the appeal of vertically differentiated products is recognized. Because the mean number of reviews change by year, the variable of interest, $DiffProReviews$ is the difference in professional reviews of film i versus the mean number of reviews for films released in a given year. The coefficient shows a negative impact on the probability of inclusion. In column 4 the coefficient indicates that each additional professional review, representing a reduction in search costs, decreases the probability by about 0.16 percent.

In column 5 both search cost variables are included in the regression,

and the coefficient of both proxies is decreased in magnitude slightly. Even when included together the coefficients on our search costs remain significant and the effects can be substantial with large increases in either. Collectively, inclusion in the Netflix library is more likely for a film that has been shown in fewer theaters and reviewed by fewer professional critics, both plausible proxies for lower search costs.

The lag between production and inclusion may provide a concern about a reduction in search costs from other services before addition to the Netflix bundle. We do not think this is a major concern, as Netflix dominated the streaming movie service in this period. However, as a robustness test we exclude any movie with a release date before 2010. The resulting sample is limited to 1,727 movies but the coefficients on each variable are remarkably similar to those in Table 2.[35] The coefficient on theaters has a more substantial negative impact, while the square term is similar in magnitude but significant. The gross variable is no longer significant with the smaller sample, while all of the ratings coefficients look very similar to the full sample regression. The one variable that has changed is Year, where a newer film from 2010 to 2013 is less likely for inclusion than the older over this short time period, perhaps due to less opportunity to be added over the two year period. The newest movies in our sample show an increased likelihood from inclusion with search costs relative to the full sample, and alleviate concerns about changed effects over time.

In Table 3 the focus is changed to the amateur user ratings. The difference in construction from professional ratings is derived from the length of time the reviews are submitted. Professional reviews are released to coin-

[35]The table is excluded for space, but available upon request.

cide with the initial theatrical run and user reviews are added as consumers view the movie, which any point after release. The theoretical framework predicts that inclusion in Netflix will decrease search costs through the use of the bundle. The endogeneity of the search costs in amateur ratings with inclusion on Netflix means this measure cannot serve as a proxy, but may serve as a useful comparison.

In columns 1 and 2 the same inclusion regression is run using amateur review data. All of the coefficients on the demand variables provide similar results. The difference in amateur reviews (in thousands), $DiffAmateurReviews$, the difference in the number of reviews from the mean for that year, shows a miniscule positive effect for every increase of 1,000 reviews compared to the mean number for that year. This effect cannot serve as a proxy for search costs due to the fact that the number can change continuously, and reflects the number of user reviews for a film as of a single point in time. We include it as a robustness test and the absence of significance reinforces prior results.

Finally, we can consider the potential for a decrease in search costs by inclusion. The difference in the number of user reviews cannot be used as a proxy for search costs, but can be used to search for evidence of a reduction in search costs, as predicted by the model. In columns 3 and 4 the dependent variable is that difference in amateur reviews. $Year$ is excluded from these regressions as each variable is measured against the mean for that year. Least squares regressions shows that as expected, the $Gross$ and number of $Theaters$ of a film increase the number of user reviews relative to the mean, an indication that demand and search costs are important in the number of reviews. $Selected$ is an indicator variable that takes the value of one only

if the film was selected for the Netflix library during this period. The large positive effect on the number of reviews from selection shows that Netflix is exposing more consumers to the film, where at least some of that increase is coming from a reduction in search costs.[36]

5 Conclusion

Valuing the experience of information goods before consumption is one of the more difficult challenges a consumer may face. This difficulty in evaluation coupled with the explosion in the number of information goods in a digital environment creates costs for consumers in determining the selection that will maximize utility. Determining which film to watch is a prime example of this problem for consumers. Search costs can cause consumers to make a suboptimal choice by obscuring the best option in the choice set.

The arrival of a new format for viewing films has coincided with this increase in search costs, the streaming bundle. The streaming bundle can successfully use the increased information in the digital environment to reduce search costs, allowing the bundler to gain an advantage over conventional distribution formats. Individual sales do much less to reduce search costs without repeated interaction and viewing data. A bundling service can take advantage of revealed preferences and a broad selection. The reduction in search costs will be increased as the number of subscribers increases and the bundler receives more information.

We model the effect of these search costs on a decision to view films,

[36] We perform the same robustness test on this specification, excluding any film released before 2010. The results are qualitatively the same, with a slightly larger negative effect from theaters on inclusion, and a smaller effect on the difference in reviews from the same. The user score coefficient was tiny and insignificant.

as well as the decision of which format to use. This decision is important in a market where streaming is increasingly dominant in growth and total revenue. The continued existence of both markets demonstrates that this decision may still vary largely by individual film, and despite the growth of streaming both options are likely to continue into the future. We find that source preference should remain important in the multiple options of this market, and that low search cost films may not allow enough of an advantage to the streaming company for inclusion in streaming bundles.

In the film industry the bundler may not be able to provide much advantage to those films with wide distribution and recognition, but for smaller films that have higher search costs the service provides a higher utility for the consumer compared to an individual purchase that may be suboptimal. This is a unique benefit of bundling found among goods, like films, that can be bundled in large numbers and suffer from the problem of high search costs. We find evidence, using Netflix data, that higher search cost films are more likely to be included in their streaming bundle. Given the difficulty in finding proxies for search cost no single specification is sufficient to indicate the validity of our hypothesis, but taken together we find strong evidence for the reduction in search costs.

Digitization of markets will continue to increase the availability of products, leading to greater consumer search costs. In this paper, we have shown evidence that bundling information goods provides consumer benefits beyond those conferred by retail services. We have shown that a bundling service can minimize search costs by taking advantage of asymmetric knowledge of consumers in a market saturated with products. As a result, it is likely that services that are able to improve the consumption decision of the

users will become more popular and important in the market for information goods. We expect that in the future, bundling services will have a more prominent role in the market for information goods.

Appendix 1: Derivations

Partially Served Market: Retail

$$\frac{\partial x^P}{\partial \gamma} = -\frac{2}{\tau}\gamma < 0$$

$$\frac{\partial x^P}{\partial P_A} = -\frac{1}{\tau} < 0$$

$$\frac{\partial x^P}{\partial \tau} = \frac{\gamma^2+P_A-V_o}{\tau^2} < 0$$

$$\frac{\partial x^P}{\partial V_o} = \frac{1}{\tau} > 0$$

Partially Served Market: Streaming

$$\frac{\partial\left(1-x^S\right)}{\partial \gamma} = \frac{-2\gamma(\tau+aP_F-NaV_o)}{(\tau-N\alpha\gamma^2)^2} > 0$$

$$\frac{\partial\left(1-x^S\right)}{\partial P_F} = \frac{-1}{N(\tau-N\alpha\gamma^2)} < 0$$

$$\frac{\partial\left(1-x^S\right)}{\partial N} = \frac{\left(-\alpha N^2\gamma^4+\alpha V_o N^2\gamma^2-2\alpha P_F N\gamma^2+\tau P_F\right)}{N^2(\tau-N\alpha\gamma^2)^2} > 0$$

$$\frac{\partial\left(1-x^S\right)}{\partial \alpha} = \frac{N\gamma^2\left(V_o-\gamma^2-\frac{P_F}{N}\right)}{(\tau-N\alpha\gamma^2)^2} > 0$$

$$\frac{\partial\left(1-x^S\right)}{\partial V_o} = \frac{1}{\tau-N\alpha\gamma^2} > 0$$

Fully Served Market

$$\frac{\partial x^*}{\partial \gamma} = \frac{-2N\alpha\gamma\left(\tau+P_A-\frac{P_F}{N}\right)}{(2\tau-N\alpha\gamma^2)^2} < 0$$

$$\frac{\partial x^*}{\partial N} = -\frac{\left(2\tau P_F-2N\alpha\gamma^2P_F+N^2\alpha\tau\gamma^2+N^2\alpha\gamma^2P_A\right)}{N^2(2\tau-N\alpha\gamma^2)^2} < 0$$

$$\frac{\partial x^*}{\partial \tau} = \frac{\alpha N\gamma^2+2P_A-\frac{2P_F}{N}}{(2\tau-N\alpha\gamma^2)^2}$$

$$\frac{\partial x^*}{\partial \alpha} = -\frac{N\gamma^2\left(\tau+P_A-\frac{P_F}{N}\right)}{(2\tau-N\alpha\gamma^2)^2} < 0$$

$$\frac{\partial x^\#}{\partial P_F} = \frac{1}{2N\tau-N^2\alpha\gamma^2} > 0$$

$\frac{\partial x^*}{\partial P_A} = -\frac{1}{2\tau - N\alpha\gamma^2} < 0$

References

Adams, W. J. and J. L. Yellen, "Commodity Bundling and the Burden of Monopoly," *The Quarterly Journal of Economics*, 1976, *90* (3), pp. 475–498.

Aguiar, Luis and Joel Waldfogel, "Streaming Reaches Flood Stage: Does Spotify Stimulate or Depress Music Sales?," Technical Report, National Bureau of Economic Research 2015.

Bakos, J. Y., "Reducing buyer search costs: Implications for electronic marketplaces," *Management Science*, 1997, *43* (12), 1676–1692.

Bakos, Y., "The emerging role of electronic marketplaces on the Internet," *Communications of the ACM*, 1998, *41* (8), 35–42.

_ **and E. Brynjolfsson**, "Bundling information goods: Pricing, profits, and efficiency," *Management Science*, 1999, *45* (12), 1613–1630.

_ **and** _ , "Bundling and Competition on the Internet," *Marketing Science*, 2000, *19* (1), 63–82.

Bran, S. and J. Matula, "Music Talk: Streaming Vs Owning," `http://blog.ourvinyl.tv/music-talk-streaming-vs-owning/` December 2014.

Brynjolfsson, E., Y. Hu, and D. Simester, "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales," *Management Science*, 2011, *57* (8), 1373–1386.

Chang, Y. and J. Walter, "Digital Piracy: Price-Quality Competition between Legal Firms and P2P Network Hosts," *Information Economics and Policy*, 2015.

Chen, Y. and M. H. Riordan, "Profitability of product bundling," *International Economic Review*, 2013, pp. 35–57.

Chuang, J. and M. Sirbu, "Optimal bundling strategy for digital information goods: network delivery of articles and subscriptions," *Information Economics and Policy*, 1999, *11*, 147–176.

Collins, A., V. Fernandez-Blanco, and J. Prieto-Rodriguez, "Characteristics of buyers and renters of cultural goods: the case of movies," *Applied Economics*, 2009, *41* (2).

Darby, M. and E. Karni, "Free competitionand the optimal amount of fraud," *Journal of Law and Economics*, 1973, *16*, 67–88.

Harris, J. and E. A. Blair, "Consumer preference for product bundles: The role of reduced search costs," *Journal of the Academy of Marketing Science*, 2006, *34* (4), 506–513.

Hiller, R. S., "Profitably Bundling Information Goods: Evidence From the Evolving Video Library of Netflix," *Journal of Media Economics*, 2017, *30* (2), 65–81.

_ **and J. Walter**, "The rise of streaming music and implications for music production," *Working Paper: available in SSRN*, 2015.

Lynch, J. G. and D. Ariely, "Wine online: Search costs affect competition on price, quality, and distribution," *Marketing Science*, 2000, *19* (1), 83–103.

McAfee, R., J. McMillan, and M. Whinston, "Multiproduct monopoly, commodity bundling, and correlation of values," *The Quarterly Journal of Economics*, 1989, *104*, 371.

Nalebuff, B., "Bundling as an entry barrier," *The Quarterly Journal of Economics*, 2004, pp. 159–187.

Nelson, P., "Information and consumer behavior," *Journal of Political Economy*, 1970, *78*, 311–329.

Nelson, R. A., C. E. Reid, and O. Gilmore, "An analysis of the out-of-market gap for DVDs in the US," *Review of Industrial Organization*, 2007, *30* (4), 303–323.

Shy, Oz, *The economics of network industries*, Cambridge University Press, 2001.

Stigler, G., "The economics of information," *Journal of Political Economy*, 1961, *69*, 213–225.

_ , "United States v. Loew's Inc.: A note on block-booking," *Sup. Ct. Rev.*, 1963, p. 152.

Thorson, Emily, "Changing patterns of news consumption and participation: News recommendation engines," *Information, Communication & Society*, 2008, *11* (4), 473–489.

Varian, H. R., "Buying, sharing and renting information goods," *The Journal of Industrial Economics*, 2000, *48* (4), 473–488.

Waterman, D. and A. Weiss, "Time consistency and seller commitment in intertemporal movie distribution: an empirical study of the video window.," *Journal of Industrial Economics*, 2010, *58* (3), 717–717.

Wessells, C., "The economics of information: markets for seafood attributes," *Marine Resource Economics*, 2003, *17*, 153–162.

Table 1: Summary statistics

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
Meta	56.29	17.06	5	98	5140
RT	58.25	28.46	0	100	5504
AverageRating	57.26	22.49	0	99	5687
AmateurRating	61.33	19.22	0	100	5673
ProReviews	82.09	63.85	2	346	5687
AmateurReviews	112247.9	335216.9	5	4079936	5673
Year	2007.16	3.9	2000	2013	5687
Theaters	0.89	1.28	0.001	4.47	5687
Gross	10.58	24.62	0	354.55	5687

Table 2: Evidence from professional reviews

	(1)	(2)	(3)	(4)	(5)
	Inclusion	Inclusion	Inclusion	Inclusion	Inclusion
Meta	0.00023				
	(0.00041)				
RT		0.00025			
		(0.00023)			
AverageRating			0.00025	0.0018***	0.0012***
			(0.00029)	(0.00031)	(0.00033)
DiffProReviews				-0.0016***	-0.0012***
				(0.00016)	(0.00018)
Year	0.014***	0.015***	0.016***	0.018***	0.018***
	(0.0018)	(0.0017)	(0.0017)	(0.0017)	(0.0017)
Theaters	-0.090***	-0.12***	-0.14***		-0.074***
	(0.026)	(0.026)	(0.025)		(0.027)
TheatersSquared	0.0037	0.0095	0.013		0.0034
	(0.0085)	(0.0084)	(0.0083)		(0.0085)
Gross	0.0034***	0.0031**	0.0035***	0.00083	0.0046***
	(0.0012)	(0.0012)	(0.0012)	(0.00088)	(0.0012)
GrossSquared	-2.0e-05**	-2.0e-05**	-2.1e-05***	-4.1e-06	-2.3e-05***
	(7.8e-06)	(7.9e-06)	(7.9e-06)	(5.8e-06)	(7.9e-06)
N	5140	5504	5687	5687	5687

Standard errors in parentheses. Coefficients represent the marginal difference in probability of inclusion in the Netflix library associated with a one unit change in the independent variable. Estimated with a conditional fixed effects logit, fixed effects for studio and genre not reported.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Evidence from amateur reviews

	(1)	(2)	(3)	(4)
	Inclusion	Inclusion	Reviews Difference	Reviews Difference
AmateurRating	-0.00052	-0.00053		
	(0.00033)	(0.00033)		
DiffAmateurReviews		0.000040		
		(0.000025)		
Selected				17380.3**
				(7879.7)
Year	0.016***	0.016***		
	(0.0017)	(0.0017)		
Theaters	-0.14***	-0.15***	96008.8***	97585.7***
	(0.025)	(0.025)	(15096.8)	(15108.5)
TheatersSquared	0.012	0.013	-27531.6***	-27720.5***
	(0.0083)	(0.0084)	(4447.2)	(4446.5)
Gross	0.0039***	0.0035***	6886.9***	6839.9***
	(0.0012)	(0.0013)	(743.4)	(743.5)
GrossSquared	-2.3e-05***	-2.3e-05***	42.4***	42.8***
	(8.0e-06)	(7.9e-06)	(7.24)	(7.24)
N	5673	5673	5673	5673

Standard errors in parentheses. Coefficients represent the marginal difference in probability of inclusion in the Netflix library associated with a one unit change in the independent variable in Columns 1 and 2, and change in the difference in number of user reviews in 3 and 4. Fixed effects for studio and genre not reported.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$