

A Taste for Obscurity?
An Individual-Level Examination of “Long Tail” Consumption

Anita Elberse
Assistant Professor
Harvard Business School
Boston, MA 02163, USA
Phone: +1 617 495 6080
Fax: +1 617 496 5853
Email: aelberse@hbs.edu

September 14, 2007

Acknowledgements:

I thank Matt Graves and Ben Rotholtz (Rhapsody) and Stephen Langsford (Quickflix) for generously providing the data used in this study, and Chris Anderson, Thomas Eisenmann, John Gourville, Sunil Gupta, Alvin Silk, and participants of the MSI Trustees Meeting and seminars at Harvard Business School and the Yale School of Management for helpful comments and suggestions.

A Taste for Obscurity?

An Individual-Level Examination of “Long Tail” Consumption

Abstract

The idea that online channels facilitate the distribution of a vast assortment of products is undisputed, but what consequence the increased supply will have on consumer demand is heavily debated. Proponents of the “long tail” principle recently argued that lower transaction and search costs will lead to a shift away from hit content and cause more fragmentation in consumers’ choices. This perspective is in sharp contrast with the more established “superstars” theory, which predicts that those forces will in fact homogenize consumption patterns, and a few products will emerge as winners in the market place. In this study, using two large customer transactions data sets obtained from American online music service Rhapsody and Australian online DVD rental business Quickflix, which together cover over a million products and over 20 million individual transactions, I examine consumption patterns for obscure and hit products. I find that a large share of consumers, particularly those who consume with a higher frequency and concentrate on a narrow selection of genres, regularly opt for obscure products in the tail of the distribution that are likely not available in bricks-and-mortar stores. Casting doubt on the “democratizing” nature of online channels, however, I also show that even for consumers who regularly choose the most unpopular products, hit products typically constitute the lion’s share of their choices. Moreover, reminiscent of the “double jeopardy” concept, consumers of obscure products generally appreciate those products less than the more popular products. I discuss managerial implications.

Keywords: online distribution, long tail, superstars, winner-take-all markets, hit versus niche products, double jeopardy, media and entertainment industry.

A Taste for Obscurity?

An Individual-Level Examination of “Long Tail” Consumption

Online channels facilitate the distribution of a vast assortment of products. Some observers believe that the increasing variety of products offered through online retailers such as Amazon, iTunes, and Netflix will fuel a shift in demand away from a relatively small number of "hit" products in the head of the distribution toward a much larger number of niche products in the tail. Recently dubbed the “long tail” principle (Anderson 2004, 2006), it mostly follows from the idea that online retailers are able to provide a much larger number of products than bricks-and-mortar stores in a cost-efficient manner, and that consumers are subject to lower transaction and search costs to discover obscure products, for instance using search and collaborative filtering tools (e.g., Brynjolfsson, Hu and Smith 2003). In his best-selling book, Anderson (2006) argues that we are leaving the “watercooler era” in which people choose from the same, relatively small pool of hit products and are entering the “micro-culture era” when people are all into different products. This perspective sharply contrasts with the more established theory of superstars, which predicts that lower transaction and search costs will homogenize, not fragment, patterns of consumption (Rosen 1981). Because consumers prefer more talented performers, and technology allows greater access to these performers, a few superstars and their products come to dominate the marketplace, creating winner-take-all (or “winner-take-most”) markets (Frank and Cook 1995).

Extending research that addresses these contrasting views at an aggregate level (e.g., Brynjolfsson et al 2003, Elberse and Oberholzer-Gee 2007), I study relevant demand-side phenomena at the level of individual consumers. Specifically, I examine whether only a small group of “fanatics” is interested in obscure products or a large group of consumers regularly ventures into the long tail, what characterizes those who consume obscure products, and how large their appetite for those products is. I also investigate these consumers’ appreciation for obscure versus hit products. I do so using two novel large customer transactions data sets obtained from Rhapsody, an American online music service, and Quickflix, an

Australian online DVD rental business. Together, the data cover over 1 million products, nearly 80,000 subscribers, and over 23 million transactions.

I find that a large share of consumers, particularly those who consume with a higher frequency and concentrate on a narrow selection of genres, regularly choose obscure products likely not available in bricks-and-mortar stores. Casting doubt on the “democratizing” nature of online channels, however, I also show that even for consumers who regularly opt for the most obscure products, hit products typically constitute the lion’s share of their choices. Additionally, reminiscent of the “double jeopardy” concept introduced by sociologist William McPhee (1963), consumers of obscure products appreciate those products *less* than the more popular products (also see Ehrenberg, Goodhardt and Barwise 1990). To the best of my knowledge, this study is the first to demonstrate the existence of these phenomena, which I trace back to early theories of mass behavior, in the context of the massive assortments mostly found in online environments.

The findings have important managerial implications, particularly for online retailers or aggregators who attempt to profit from long tail demand. One key take-away is that their best customers—those that are more loyal and have been customers longer—likely are more interested in wide assortments. However, online retailers should not expect their “long tail” customers to predominantly opt for obscure products and like those more than the hits—in fact, the opposite is more likely—and take that into account when managing customer expectations and satisfaction. Even though obscure products often are more attractive for retailers from a margin perspective, building a business solely revolving around a wide collection of niche products may, the findings suggest, be difficult. Even for online retailers that tout their large assortments and their ability to tap into the long tail, a select group of hit products likely remains critical to their bottom line.

The manuscript is organized as follows. I start by discussing the central hypotheses in more detail, drawing on literature in economics and marketing. I then describe the data and the modeling approach, which entails estimating regular, log-log and negative binomial regression models. I end by discussing the findings, conclusions, and managerial implications.

ONLINE CHANNELS AND THE DEMAND FOR OBSCURE PRODUCTS

The End of the “Watercooler Era”? Most online stores offer a much greater variety of products than their bricks-and-mortar counterparts. For example, Apple’s iTunes offers millions of albums and songs and online retailer Amazon over 250,000 albums, while even the largest offline music retailer typically makes only 15,000 albums available. Similarly, while bricks-and-mortar stores usually stock up to 1,500 DVDs, online retailers may offer up to 80,000 unique titles (e.g., Brynjolfsson et al 2003, 2006). This phenomenon is generally attributed to supply and demand side forces (also see Elberse and Oberholzer 2007). On the supply side, sellers’ transaction costs, the costs incurred during the process of selling, are generally lower online. Whereas adding additional titles requires more physical space in traditional retail environments, increasing variety and effectively communicating the additional choices to customers, for example using recommendation engines, is much less costly for web-based retailers (Alba et al 1997, Brynjolfsson et al 2003, 2006, Brynjolfsson, Hu and Simester 2007, Oestreicher-Singer and Sundararajan 2006, The Economist 2005). The advantages are particularly strong for information goods that can be digitized and distributed online almost without cost (Bakos and Brynjolfsson 2000). On the demand side, search costs, all costs incurred by a buyer to purchase a product, are also lower in online channels—think, for example, of the costs to locate a seller and obtain information on prices and other product attributes (Alba et al 1997, Bakos 1997, 2001, also see Lynch and Ariely 2000). Given the lower transaction and search costs, consumers’ appetite for niche products emerges as the critical factor in the possible emergence of a lucrative long tail in online channels.

Proponents of the long tail phenomenon expect the internet to have a “democratizing” influence—they think it will shift the attention of a substantial group of consumers away from hit products and toward niche products that are more to their liking. Anderson (2006, page 2) for example, argues that hits “are not quite the economic force they once were,” and that “fickle customers” will “scatter to the winds as markets fragment into countless niches.” He predicts that we are “leaving the watercooler era, when most of us listened, watched, and read from the same, relatively small pool of

mostly hit content. And we're entering the micro-culture era, when we're all into different things" (Anderson 2006, page 185).

In general terms, such visions of fragmented markets or "marketplaces of ideas" are not new—they frequently surface in discussions about the effect of the growing number of media options, particularly in the context of the emergence of additional windows for distributing media content and their larger capacity for carrying content that appeals to narrow audience segments, as well as the increasing audience autonomy that new technologies provide, (e.g., Napoli 2003, Neuman 1991, Webster 2005). The rise of the internet and the truly massive assortments of online stores have moved these issues to the forefront again. The long tail idea appears to have caught on with many practitioners in the media and entertainment industry, and increasingly exerts an influence on the development and appraisal of business models in this sector (see, for example, two recent Bear Stearns (2006, 2007) equity research reports on the long tail). However, the long tail ideas have also drawn sharp criticism, and have even been dismissed as "web utopian fantasies" by one reporter (The Wall Street Journal 2006a, 2006b).

There are reasons to believe that consumption patterns will further homogenize—not fragment—in an online world. Economists have argued that people tend to converge on the same hit content regardless of the breadth and depth of niche content available. Rosen (1981) explained this "superstar" phenomenon as the result of two factors. First, lesser talent is a poor substitute for greater talent: why would people, say, listen a recording of the world's second-best opera singer if the best is also available? Second, the marginal cost of reproducing and distributing books, records, and videos is low, creating a significant cost advantage for hit products over content of limited appeal. The latter is particularly relevant to the online distribution of digitized information goods, where reproduction costs approach zero. Other researchers in economics, sociology and communications have pointed out that because of their social nature, people have a keen interest in reading the same books and watching the same movies that others consume, thereby also creating winner-take-all trends (e.g., Frank and Cook 1995, Napoli 1999, 2003, Webster 2005, Salganik, Dodds, and Watts 2006).

Relevant marketing studies provide mixed evidence. On the one hand, some studies reveal that, when customers can choose from a larger selection of products, they may opt for products that better match their preferences or they may more strongly fulfill a need for variety within and across consumption occasions (Kahn 1995, Kahn and Lehmann 1991, Hoch et al. 1999, Iyengar and Lepper 2000, Chernev 2003, also see Häubl and Trifts 2000, and Lynch and Ariely 2000). On the other hand, fueling doubts about a long tail trend, some recent marketing studies suggest that overwhelmingly large assortments may create consumer confusion or frustration, and could even lead some consumers to making no choice at all (Huffman and Kahn 1998, Lehmann 1998, Gourville and Soman 2005, Dhar 1997, Iyengar and Lepper 2000, Boatwright and Nunes 2001, Chernev 2003, Broniarczyk, Hoyer and McAlister 1998).

Studying the Long Tail: A Need to Focus on Individual Consumers Early research on the long tail trend reveals there is at least some aggregate demand for obscure products not available in bricks-and-mortar stores. Brynjolfsson et al (2003) were among the first to point out that the Internet enables online retailers to catalog, recommend, and provide a large number of products for sale. In an analysis of Internet book sales, they estimate that the increased product variety of online bookstores—the long tail—enhanced consumer welfare by \$731 million to \$1.03 billion in the year 2000. In a study of the U.S. home video industry, Elberse and Oberholzer-Gee (2007) find a long-tail effect in that the number of DVD titles that sell only a few copies every week increases almost twofold in the period from 2000 to 2005. Brynjolfsson et al (2007) examine how reduced search costs affect the concentration in product sales, and empirically show that a women’s clothing retailer’s online channel exhibits a significantly less concentrated sales distribution, when compared with its traditional catalog sales channel, even after controlling for differences in individual consumers. The latter study is exceptional in its consideration of individual-level dimensions, but predominantly focuses on aggregate, market-level trends.

Although it is critical to the long-term viability of online businesses that are based on long-tail consumption, it remains largely unknown which consumers are responsible for that demand, and how they value the experience. Is only a small group of “fanatics” interested in obscure products—in which case it is difficult to maintain that a significant shift in media consumption is taking place—or do a large

group of consumers regularly venture into the long tail? What are the characteristics of those who consume obscure products, and how large is their appetite for those products? And how do these consumers value obscure products compared to the hits? I attempt to address these questions.

Question 1: Who Consumes Obscure Products—and How Much? According to McPhee (1963), whose groundbreaking “theory of exposure” lies at the root of the marketing literature on the relationship between brand shares and loyalty (e.g., Ehrenberg et al 1990), we should expect that consumers with light exposure to products in a certain category will constitute a larger proportion of the audiences of the popular titles than of the obscure titles: “the most popular product gets not only more raw numbers of people of otherwise marginal participation in the field, but a disproportionate share of its audience (a larger fraction of its already larger audience) consists of just such marginal people” (page 127). For instance, in the area of sports, McPhee predicted that those least interested in sports generally are a larger proportion of the already larger audience interested in the most popular sport. Put differently, the largest sport is more popular because it has a uniqueness, a monopoly, among those who choose few competitors. McPhee (1963) called the general phenomenon “*natural monopoly*,” and ascribed it to two sources: first, the heterogeneity of motivation among members of the population to seek out unobvious alternatives and, second, chance events even within a population absolutely homogenous in all intentions.

In the context of this study, this theory would suggest that the choice for products in the long tail is positively related to the frequency of consumption: hit products “monopolize” light consumers of the category, while heavy consumers choose a mix of hit and obscure products. This is directly at odds with the intuitively appealing notion that audiences for small titles tend to be composed of loyal, die-hard fans (Webster 2005). McPhee based his ideas on explorations in settings with typically less than a dozen alternatives. To the best of my knowledge, the phenomenon has not been shown to exist for large assortments with thousands of products. In fact, after using frequency as a control variable in their regression analysis of online and offline sales for more than 6,000 products by a women’s clothing retailer, Brynjolfsson et al (2007) recently reported that customers with a higher frequency of historical transactions were *less* likely to purchase obscure items—the opposite of what McPhee’s theory predicts.

One reason why one could expect heavy customers to choose more obscure products is that they somehow “run out of good options” among the more popular products.¹ In that respect, it seems worthwhile to examine the role of genres of cultural products. Most consumers have a preference for certain genres or subcategories within the product category, but they are likely to differ in the strength of their preferences. Some consumers may be omnivores in that they are keen to adopt products from a variety of genres ranging from, say, jazz to hip hop music, while others may solely want to consume products that belong to their most preferred genre. As such, it seems reasonable to expect a relationship between the degree of concentration in choices across genres on the one hand and the choice for obscure products on the other hand. A higher level of concentration, i.e. a more narrow focus on certain genres, could be associated with a higher risk of depleting viable popular options, and thus be correlated with relatively more obscure choices.

Question 2: How Much Do Consumers Appreciate These Obscure Products? McPhee

(1963) also examined how much popular and unpopular products are liked. A phenomenon he introduced as “*double jeopardy*,” McPhee posited that “the larger the proportion of the people that is not familiar with a given alternative, then the less likely are those who are familiar with it to like it especially” (page 133).² In other words, obscure titles are appreciated less than more popular titles. At first glance this seems absurd, and in fact most people are inclined to believe the contrary. In McPhee’s words, “we console ourselves that the obscure alternative is at least appreciated by the people who do know it; for example, the out-of-the-way book is at least a delight to those who find it” (page 136). Those intuitions, he argues, are misplaced.

He based his theory on patterns in the results of Gallup polls on the relative appeal of media alternatives, and specifically a demonstrated link between the proportion of people who had heard a certain NBC radio personality, and the reported liking among those who had heard that personality.

¹ This possibility, in fact, is the rationale behind Brynjolfsson et al (2007)’s control variable.

² In the marketing literature, the double jeopardy phenomenon is typically described as the dual problem of both small and disloyal consumers (e.g., Ehrenberg and Goodhardt 1990). In studies of the behavior of media audiences, loyalty is typically operationalized as either the time spent viewing or the amount of repeat viewing (e.g., Goodhardt, Ehrenberg and Collins 1975, Barwise and Ehrenberg 1984, Barwise 1986, Webster 2005).

McPhee attributes the double jeopardy phenomenon not necessarily to the possibility that obscure products simply lack merit. Rather, he argues that it mostly follows from the fact that obscure products are chosen by people who are familiar with (too) many competitive alternatives, whereas the favored or “obvious” alternative is the one that becomes known to the kind of people who, in making choices, know little else to choose from. The “double jeopardy” for an unpopular alternative therefore is, first, not being known and then, second, when it finally becomes so, being known to people “who know better.”

As the phenomenon to date has only been studied in settings with relatively modest assortments, it is an open question to what extent this phenomenon also exists in contexts with thousands of alternatives that can now so often be found in online environments. Do consumers who choose long-tail products indeed appreciate those less than the hits? And, building on the idea of long-tail consumers “knowing better,” is the double jeopardy effect stronger for those consumers who can be considered to have more expertise, for instance because they consume more products overall or because they “specialize” in certain genres which makes them more skilled at distinguishing, say, an inferior science fiction movie from a superior one?

DATA

The data I use to examine the research questions come from two companies: Rhapsody, an American online music service, and Quickflix, an Australian online DVD rental service. Both firms can be characterized as “content aggregators” that enable their subscribers, in exchange for a monthly fee, to access a large assortment of entertainment products—music titles and DVDs, respectively. **Table 1** provides descriptive statistics.

Rhapsody Run by RealNetworks, Rhapsody offers its subscribers streaming on-demand access to its library of, as of October 2006, over 3.5 million songs. They have the ability to browse the library, search it by artist, album, track, and other title characteristics, and access over 100 radio stations which are programmed by Rhapsody’s editorial team or created by users themselves. They can choose

between three subscription plans: “Rhapsody Unlimited,” “Rhapsody To Go,” and a free trial. For a monthly fee of \$12.99, Rhapsody Unlimited subscribers are able to download an unlimited number of tracks to their computers. Priced at \$14.99 per month, Rhapsody To Go offers subscribers Rhapsody Unlimited’s features plus the ability to transfer an unlimited amount of music to compatible portable devices without purchasing songs on a per-track basis. Rhapsody’s advertising-supported free trial service, “Rhapsody 25,” allows users to play 25 songs on-demand each month and listen to 25 radio stations. Consumers can also purchase individual songs and albums through Rhapsody, generally for \$0.99 per song and \$9.99 per album, but at a 10% discount for subscribers.

The data used in this study are individual transactions, either streaming of songs on demand or playback of subscription music downloaded to subscribers’ PCs, from August 1 to October 31, 2006. Data are available for a random sample of Rhapsody’s “active” customer base, defined as those customers with more than one transaction per week in the three-month study period. The sample consists of 64,463 anonymous customers who together played 1,047,032 tracks and 207,941 albums from 81,226 artists in 20 genres (e.g., “Latin”) and over 500 subgenres (e.g., “Latin Rock”). The data cover 32,164,031 individual transactions.

Quickflix A service comparable to Netflix in the U.S., Australia-based Quickflix allows its customers to compile a list of online of DVDs they would like to receive via regular mail, keep the DVDs that have been delivered to them as long as they want, and receive a new title when they return one. Depending on their plan, users can have up to five DVDs at any given time, and up to ten or an unlimited amount per month. Prices range from AUS\$9.95 for a one-DVD-at-a-time-and-two-DVDs-per-month plan to AUS\$56.95 for a plan that allows customers to rent up to five DVDs at a time without setting limits on the monthly total. The most popular plan is the three-DVDs-at-a-time plan priced at AUS\$36.95. New customers are granted a 14-day free trial option.

The data available for this study cover all transactions for Quickflix’s full customer base of 15,741 customers over a 6-month period from January 1 to June 30, 2006. The number of unique titles shipped to these customers totals 15,768, or virtually Quickflix’s entire catalog as of June 2006, while the number of

transactions is 322,593. Characteristics of titles include the movie release year, DVD release date, 22 genres (e.g., Action, Documentary, and Romance), six Australian movie ratings classifications (the equivalent of MPAA ratings in the U.S., with categories ranging from “G” (suitable for all ages) to “R” (restricted to viewers of 18 years of age and over), and the studio or distributor (such as 20th Century Fox, the BBC, Columbia Pictures, and Universal). Characteristics of transactions include the shipping date and, if provided, a customer’s rating of the title after viewing it.

Content Aggregators and the Appeal of the Long Tail

The distribution of transactions across titles ranked from the most popular to the most obscure, for both Rhapsody and Quickflix, is depicted in **Figure 1**. The figure shows a relatively high level of concentration in plays across the over one million tracks for Rhapsody. The top 10% of titles account for 78% of all plays, and the top 1% of titles 32% of all plays. In other words, roughly 10,000 titles account for nearly one third of all plays. For Quickflix, the concentration is less strong in a relative sense: the top 10% of DVDs account for 48% of all rentals, and the top 1% for 18% of all rentals. In other words, roughly 150 titles account for nearly one fifth of all rentals. However, especially given the vast assortments, it is worth noting the implications in an absolute sense: 1% of Rhapsody titles, or 10,000 songs, far exceeds the number of unique titles played on an average U.S. radio station in a given year and, when expressed in albums, represents roughly the entire music inventory of a typical Wal-Mart store (Anderson 2007). Similarly, 1% of Quickflix DVDs is on par with all movies theatrically released by major U.S. movie studios in a given year. Online retailers like Rhapsody and Quickflix thus facilitate the consumption of a vast number of titles that, especially if Brynjolfsson et al (2006)’s estimations of the assortment sizes of bricks-and-mortar music (15,000 albums) and DVD retailers (1,500 DVDs) are correct, are likely mostly not available in traditional channels.

The concentration of transactions across titles can be described using the Gini coefficient mostly known for its applications in research on wealth inequality (Gini 1921). Brynjolfsson et al (2007) also use the Gini coefficient in their study of online versus catalog clothing sales distributions, while Salganik et al

(2006) apply it to measure success inequality in music downloads in an experimental setting.³ In terms of the present study, if one plots the sales distribution curve in a graph with on the x-axis the cumulative percentage of titles and on the y-axis the cumulative percentage of transactions involving those titles, then the Gini coefficient is the ratio of the area between the curve and a 45-degree line to the total area under a 45-degree line. It is calculated as:

$$Gini = 1 - \frac{2 \sum_{i=1}^n [(n+1-i) popularity_i]}{(n+1) \sum_{i=1}^n popularity_i}$$

where $popularity_1, popularity_2, \dots, popularity_n$, denotes the number of transactions for each of n titles, and $popularity_1 \leq popularity_2 \leq \dots \leq popularity_n$. When transactions are evenly distributed across titles, i.e. when every title is equally popular, the Gini coefficient has a value of zero. As the distribution becomes more concentrated, the Gini coefficient increases, and if all sales are concentrated with one title, the Gini coefficient equals one.

The results for the two data sets are as follows:

Data	Gini
Music (Rhapsody)	0.838
DVDs (Quickflix)	0.604

With a Gini coefficient of 0.84, usage of Rhapsody’s music service is substantially more concentrated than the internet and even the catalog sales channels in Brynjolfsson et al (2007)’s study, which report coefficients of 0.70 and 0.77, respectively. Transactions are more evenly spread across titles in the case of Quickflix’s DVD rental service, with a Gini coefficient of 0.60.

The exact distribution of sales across titles is a critical metric for content aggregators like Rhapsody and Quickflix. They benefit from directing customers to less popular titles in the tail of the

³ Other concentration measures often used in media and marketing contexts include those of “relative entropy” and the “Herfindahl-Hirschman Index” (e.g., see Napoli 1999).

distribution because those tend to generate higher margins. Rhapsody compensates music rights owners in the form of either an agreed-upon amount each time a subscriber plays a certain song or a certain minimum amount per subscriber. Contracts with smaller, independent labels that bring lesser-known artists to the market are often more favorable for Rhapsody than agreements with the major record labels that represent the superstar artists. Holding all else constant, shifting demand to the smaller, independent labels therefore theoretically could lead to increased profits for Rhapsody. Quickflix's costs primarily vary depending on the age of a title. New releases are typically priced anywhere between AUS\$25 and AUS\$35, catalogue titles (older than two months) below AUS\$15, and back catalogue titles (older than six months) below AUS\$10.⁴ Certain budget titles can cost below AUS\$5. Quickflix fully depreciates its DVDs in a matter of months. The video rental firm can thus directly improve its profitability by shifting demand to titles that were cheaper to buy or have depreciated more over time.

MODELING APPROACH

My approach to testing the two core hypotheses regarding the consumption and appreciation of obscure versus popular products consists of three stages.

Stage 1: Regression: The Average Rank of Titles

In order to examine the drivers of customers' inclination to consume hit versus obscure products, first, I estimate a regression model with the log of the average rank of the titles consumed by a customer as the dependent variable,⁵ and a set of customer descriptors as independent variables. Specifically:

$$\log(\text{average_rank}_i) = \beta_0 + \beta_1 \log(\text{frequency}_i) + \beta_2 \log(\text{genre_concentration}_i) + \beta_3 \text{subscription}_i + \varepsilon_i \quad (1)$$

⁴ In the last month of the Quickflix study period, June 2006, 1 Australian dollar averaged 0.74 U.S. dollar.

⁵ I opt for this dependent variable to facilitate direct comparisons with Brynjolfsson et al (2007). Benchmark analyses with dependent variables that better account for the ordinal nature of the underlying rank measure, such as the median of the rank of the titles consumed, generate similar substantive results.

where, for customer i , $average_rank_i$ denotes the average rank of the titles that customer i chose in the study period (with higher values reflecting a greater focus on obscure titles), $frequency_i$ reflects the total number of titles consumed by that customer in the study period, $genre_concentration_i$ represents a measure of the inequality of the distribution of titles chosen by customer i across their genres, $subscription_plan_i$ denotes a vector of covariates related to that customer's subscription, and ε_i represents the error term. In line with the theory that light consumers “monopolize” the most popular products, and that consumers with a more narrow focus on certain genres face a higher risk of depleting viable popular options, I expect the estimates for both parameters β_1 and β_2 to be positive.

I measure the variable $genre_concentration_i$ by assessing, for each customer separately, the relative number of titles consumed in each of the available genre categories—20 for the Quickflix data, and 22 for the Rhapsody data. Specifically, I calculate the concentration of a customer's transactions across genres using a new application of the Gini coefficient (Gini 1921):

$$genre_concentration_j = 1 - \frac{2 \sum_{g=1}^h [(h+1-g) popularity_g]}{(h+1) \sum_{g=1}^h popularity_g}$$

where $popularity_1, popularity_2, \dots, popularity_h$, denotes the number of transactions for each of h genres, and $popularity_1 \leq popularity_2 \leq \dots \leq popularity_h$, calculated for each customer j separately.

Like the Gini coefficient, the $genre_concentration_i$ measure varies between zero and one, with a score just above zero reflecting a low level of concentration or, in this context, choices spread out relatively evenly across genres, and a score close to one representing a high level of concentration, or a few genres receiving a disproportionate amount of attention. I illustrate the measure for two customers in **Figure 2**.

For estimations with the Rhapsody music data, $subscription_plan_i$ includes the log of the length of a customer's subscription, measured by the number of days between that customer's first registration date and the end of the study period (October 31, 2006), as well as two dummies for the Rhapsody Unlimited

and Rhapsody To Go partner plans, with the omitted third option being the 7-day free pass. For the Quickflix DVD data, *subscription_plan_i* also includes the log of the length of a customer's subscription, again measured by the number of days between that customer's first registration date and the end of the study period (June 30, 2006), as well as dummies for the maximum number of DVDs a customer can rent at any given time and in any given month, and a dummy for a free trial plan.

I estimate the model for each of the two data contexts separately, using ordinary least squares. I generate heteroskedasticity-robust standard errors using MacKinnon and White's (1985) 'HC3' method. The log-log form aids the interpretation of the results in that the estimated coefficients directly represent the elasticity of the right-hand-side variable with respect to changes in the left-hand-side variable. In other words, each coefficient describes the percentage change in *average_rank_j* for each percentage change in the corresponding covariate. The equation is modeled after work by Brynjolfsson et al (2007), who in their aforementioned study of a women's clothing retailer constructed a log-log model with the log of the average rank of the products chosen by a customer regressed on measures of his or her prior internet experience, recency, monetary value, and purchase frequency. This facilitates a comparison of the results, most notably regarding their finding of a negative relationship between a customer's average rank and frequency of transactions.

A disadvantage of the model is that the dependent variable only allows for a rudimentary examination of long tail consumption. For instance, customers with a strong interest in both the most obscure and popular products cannot be distinguished from customers with a taste for moderately popular titles. Therefore, second, to better understand the composition of audiences for products with varying degrees of popularity and the drivers of the demand for those products, I turn to a count model.

Stage 2: Negative Binomial Regression: The Number of Titles in A Popularity Decile

I estimate a set of models with the number of titles chosen by a customer within a particular popularity threshold as the dependent variable. That is, I split the titles into ten groups: the popularity thresholds are defined as the decile to which a title *j* belongs based on that title's total number of transactions in the sample period. (In **Figure 1**, the deciles are obtained by dividing the titles ranked on the x-axis into ten

groups, with the leftmost group representing the 10% of titles with the lowest number of transactions, and the rightmost group representing the 10% most popular titles). I estimate one model for each decile.

Because the dependent variable is a count variable and likelihood ratio tests provide evidence of overdispersion, I opt for an extension of the Poisson regression model—the negative binomial regression model (Coleman 1964).⁶

The specified model is:

$$\begin{aligned} titles_{id} = & \beta_0 + \beta_1 frequency_i + \beta_2 genre_concentration_i \\ & + \beta_3 subscription_i + \varepsilon_{id} \end{aligned} \quad (2)$$

for $d = 0, 1, 2, \dots, 9$, where $titles_{id}$ is the number of titles chosen by customer i that belong to a certain decile d . The lowest decile contains the 10% of titles with the fewest total transactions, while the highest decile consists of the 10% of titles with the highest total transactions in the study period. The main independent variables in these models again are $frequency_i$, the total number of titles chosen by customer i , and $genre_concentration_i$, the measure of the equality of that customer's choice of genres. The vector related to that customer's subscription plan, $subscription_i$, contains the same covariates as that in equation (1), albeit none in log form.

Again, as in equation (1), I expect the estimates for parameters β_1 and β_2 to be positive, as light consumers “monopolize” the most popular products, and consumers with a more narrow focus on certain genres face a higher risk of depleting viable popular options. One key advantage of the model in equation (2) is that it allows for the examination of a possible disproportionate relationship—a relatively strong effect of, say, a customer's frequency of consumption on the number of the most obscure products chosen.

To facilitate the interpretation of the results, I transform the estimated coefficients and report incidence rate ratios, i.e. $\exp(\beta)$ instead of β . These ratios directly show the factor by which the number of

⁶ In a general form, for a discrete random variable, Y , and observed frequencies, y_i , $i = 1, 2, \dots, N$, where $y_i \geq 0$, and regressors x_i , the probability distribution for the negative binomial regression model can be expressed as $P[Y = y_i | \varepsilon] = e^{-\lambda_i \exp(\varepsilon)} \lambda_i^{y_i} / y_i!$, whereby $\ln \lambda_i = \delta' x_i + \varepsilon$ and $\exp(\varepsilon)$ has a gamma distribution with mean 1 and variance a . (Greene 1997).

titles chosen in a particular decile changes as a result of changes in an independent variable. For example, an estimate of 3 indicates the number of titles consumed triples with a one-unit increase in the corresponding independent variable, while an estimate of 0.5 implies the number of titles consumed in a particular decile is reduced by 50%.

Stage 3: Regression: The Rating of Titles Third and finally, to assess the relationship between customer's consumption and appreciation of products of varying degrees of popularity, I use a regression model with the customer rating as the dependent variable:

$$\begin{aligned} rating_{ij} = & \beta_0 + \beta_1 popularity_j \\ & + \beta_2 frequency_i + \beta_3 (popularity_j \cdot frequency_i) \\ & + \beta_4 genre_share_{ij} + \beta_5 (popularity_j \cdot genre_share_{ij}) \\ & + \beta_6 subscription_i + \beta_7 release_j + \varepsilon_{ij} \end{aligned} \quad (3)$$

where $rating_{ij}$ denotes the rating given by customer i to title j after consuming it. Among the independent variables, $popularity_j$ denotes the number of transactions involving title j (measured in thousands), $frequency_i$ the total number of titles chosen by customer i , $genre_share_{ij}$ the share of customer i 's transactions dedicated to the genre to which title j also belongs. Both $(popularity_j * frequency_i)$ and $(popularity_j * genre_share_{ij})$ are interaction terms. The vectors of customer ($subscription_i$) and title ($release_j$) characteristics are control variables, and ε_{ij} is the error term.

If the double jeopardy phenomenon holds, the coefficient for the variable $popularity_j$, β_1 , is expected to be greater than zero: a title's ratings are expected to increase the more popular the title is or, in McPhee's (1963) words, "the larger the share of the audience that is not familiar with a given alternative, then the less likely are those who are familiar with it to like it especially."

The variable $frequency_i$ (with coefficient β_2) potentially plays an important role in this relationship. As McPhee's natural monopoly phenomenon predicts, more frequent customers form a larger share of the audience for more obscure products. If the more frequent customers are, on the whole, also more critical, natural monopoly in itself could therefore cause a positive relationship between a title's rating and its

popularity. After all, lower ratings given by this group of customers would have a disproportionately large negative impact on the more obscure products. On the other hand, a positive coefficient for β_2 —more frequent customers giving, on average, higher ratings—would help strengthen a possible double-jeopardy result. A positive relationship between $rating_i$ and $frequency_i$ ($\beta_2 > 0$), seems intuitive in its own right: the more transactions customers engage in, the more they can be expected to appreciate the experience in general. The same holds for a possible positive relationship between $rating_i$ and $genre_share_{ij}$ ($\beta_4 > 0$); it seems likely that customers who dedicate a larger share of their choices to a certain genre will, on average, appreciate titles in that genre more than those who rarely pick the genre.

The interaction terms with the variables $frequency_i$ and $genre_share_{ij}$ may help understand McPhee’s explanation for the double jeopardy idea, namely that customers who consume and therefore rate obscure products “know better” than to like those products. The two variables can be thought of as operationalizations of a customer’s expertise: people who transact with greater frequency could develop a better ability to discern various levels of title “quality,” however defined, and people who spend relatively more of their time watching, say, romantic comedies may become better at distinguishing a superior romantic comedy from an inferior one. Consequently, $frequency_i$ and $genre_share_{ij}$ may moderate the relationship between a title’s popularity and an individual customer’s appreciation for it. I expect that customers with a higher level of “expertise,” i.e. higher scores on $frequency_i$ and $genre_share_{ij}$, will give disproportionately low ratings to obscure products and disproportionately high ratings to popular products. Thus, I expect the coefficients β_3 and β_5 belonging to the interaction terms to be greater than zero.

I can only estimate the model in equation (3) using the Quickflix DVD data—the Rhapsody data do not include ratings data. A total of 4,443 customers, or 28% of the subscriber base, have provided ratings for one or more titles. Covariates for the vector $subscription_j$ include the length of a customer’s subscription as well as dummies for the maximum number of DVDs a customer can rent at any given time and in any given month and a dummy for a free trial plan. The covariates belonging to the vector $release_j$ include one dummy for distribution by a major Hollywood studio, four dummies for official

Australian Classification Board rating categories⁷ (“PG,” “M,” “MA15+,” and “R,” with “G” being the excluded dummy), and the year of the theatrical release.

FINDINGS

I describe the main results organized along the three stages in the modeling approach.

Average Rank of Titles Consumed Equation (1) regresses the average rank of all the titles involved in a customer’s transactions on several customer characteristics. The estimates are reported in **Table 2**.⁸ The table shows that, for both the music and DVD data, the model fit is significant at a 1% level, with an adjusted R^2 of 0.17 for music and 0.21 for DVDs. The reported estimates support the view that, controlling for various customer characteristics, the frequency of consumption is a driver of the consumption of relatively unpopular products. As illustrated by the positive and, at a 1% level, statistically significant value for β_1 , the parameter belonging to variable *frequency*, customers with a higher frequency of consumption on average choose less popular titles. Both results are consistent with McPhee’s natural monopoly concept.

For music, β_1 is 0.17, meaning a 1% change in frequency leads to a 0.17% change in the rank of titles consumed (recall that the coefficients, due to the log-log form, can be interpreted as elasticities). For example, the findings suggest that if a customer that plays an average of about 500 tracks with an average rank of 77,000 increases her frequency with 100% to around 1,000 tracks played, the average rank of her chosen tracks will move to close to 90,000. The move to less popular products is substantial in this example, but the average track consumed remains in the top decile of just over 100,000 tracks (recall that the total number of tracks in the database is 1,047,032).

⁷ The categories are G (general), PG (parental guidance recommended), M (recommended for mature audiences), MA 15+ (not suitable for people under 15; under 15s must be accompanied by a parent or adult guardian), and R (restricted to people 18 and over).

⁸ A correlation analysis reveals that Pearson correlation coefficients between the continuous customer characteristics, *frequency*, *genre_concentration*, and *subscription_duration*, are significantly different from zero in both the music and DVD data sets, but not high enough to warrant concerns about multicollinearity. The highest coefficient is 0.44.

For DVDs, β_1 is 0.52, meaning a 1% change in frequency leads to 0.52% change in the rank of titles consumed. This suggests, for example, that if a customer that rents an average of about 21 DVDs with an average rank of 2,600 increases its frequency with 100% to around 40 rentals, the average rank of titles rented will move to close to 4,000. Considering that each decile contains about 1,500 titles, this implies a shift from the average title being in the second decile to one being in the fourth decile—a far more dramatic shift to the tail. This corresponds to the more equal distribution of transactions across a smaller assortment of titles for the DVD data, as compared to the music data, as illustrated in **Figure 1**.

The role of frequency in these two data contexts stands in sharp contrast with Brynjolfsson et al’s (2007) finding of a negative relationship between the natural logs of sales rank and frequency, measured as number of historical transactions in years before their study period. They report a coefficient of -0.20, statistically significant at a 1% level.

Similarly, as hypothesized, the estimate for β_2 , the parameter for *genre_concentration_j*, is statistically significant and positive for both data contexts. The findings thus support the existence of a positive relationship between a customer’s concentration on a select set of genres and that customer’s tendency to consume obscure products: the more a consumer is focused on one or just a few genres, the deeper he or she is likely to move into the tail with less popular products. The estimate for β_2 is 0.10 for the music data, and 0.57 for the DVD data. That suggests, for example, that compared with a Rhapsody customer with a *genre_concentration* score of 0.46 and average rank of tracks consumed of 77,000, a customer with a score that is 50% higher will consume tracks that have, on average, a 5% higher rank of nearly 81,000. Similarly, compared with a Quickflix customer with a *genre_concentration* score of 0.40 and average rank of tracks consumed of 2,600, a customer with a score that is 50% higher will consume tracks that have, on average, a 28% higher rank of over 3,300.

As far as the control variables are concerned, the findings also show that the consumption of obscure products increases with less restrictive subscription plans, in terms of the titles that customers can access per period (*subscription_unlimited_j* for music and DVDs), the number they can transfer onto other devices (*subscription_togo_j* for music), and the number they can have in their possession at any given time

(*subscription_3-or-more-at-a-time*_j for DVDs). In both the DVDs and music context, unlimited plans appear particularly strongly related to a choice for more obscure products, given the relatively high positive coefficients of 0.45 for music and 0.20 for DVDs. Free trial customers (*subscription_free-trial*_j for music, and the excluded dummy for Rhapsody) appear relatively more focused on the popular products.

Evidence for a relationship between a customer's choice for products of various levels of popularity and the duration of his or her subscription is inconclusive. For the music context, the coefficient for *subscription_duration*, β_3 , is statistically significant and positive, with a value of 0.03. For the DVD data, the coefficient for *subscription_duration*, β_3 , is statistically indistinguishable from zero. The difference may be a result of the extremely strong growth that the DVD rental firm Quickflix was undergoing during the sample period, leading to a relatively large number of customers that had only recently joined the company's customer base.

Number of Obscure and Popular Titles Consumed Before moving to the estimation results for the negative binomial models, it is useful to consider the consumption patterns in the underlying raw data. **Figure 3** depicts the distribution of obscure and popular titles across customers for both Quickflix (the top graph) and Rhapsody (the bottom graph).

The top graph reveals that for customers renting at least one DVD in the highest decile, i.e. at least one of the 10% most popular titles, in the study period, on average 61% of that customer's rentals come from that top decile, and another 13% from the second-highest decile. Less than 1% of all titles rented by these customers come from the lowest decile, i.e. the 10% most obscure titles. By comparison, for the smaller group of customers renting at least one DVD in the lowest decile, i.e. at least one of the 10% most obscure titles, on average 8% of that customer's total rentals is comprised of titles in this lowest decile, and just over a quarter are titles in the bottom half of the distribution. Even for this group, the biggest share of rentals, on average just over a third of that customer's rentals, originate from the top decile. The wide appeal of these top titles is, of course, what makes them popular in the first place.

Offering anecdotal evidence for the natural monopoly phenomenon, the downward sloping pattern implied in **Figure 3** generally becomes more pronounced the more the focus is on heavy users of

the category. Consider the second graph for music consumption, where the focus is on customers that play tracks belonging to at least one hundred different albums in a given decile. Here, for customers playing tracks from at least one hundred different albums in the highest decile, on average 87% of that customer's plays come from that top decile, and another 7% from the second-highest decile. Less than 1% of their consumption consists of the 50% lowest-ranked titles. In contrast, for customers playing tracks from at least one hundred albums in the lowest decile, on average 2% of those customers' total plays is comprised of titles in this lowest decile, and just below 15% are titles in the bottom half of the distribution. Again, over a third of those customers' plays—the largest share—comes from the top decile.

The negative binomial regression model allows for a closer examination of these phenomena and the factors that shape them. For both contexts, I model the number of titles that customers choose in a given decile as a function of the customer characteristics also featured in the log-log regression model above. **Table 3** lists the estimates. Note that the table reports incidence rate ratios which directly show the factor by which the number of titles consumed in a particular decile changes as a result of a one-unit change in the corresponding independent variable.

Not surprisingly, the results show that a customer's overall frequency of consumption is a positive predictor of the number of titles consumed in a given decile in both the music and DVD data contexts. The estimate for *frequency*, β_1 , is significantly greater than one in every instance. More interestingly, and in line with expectations, the estimates for β_1 are highest in the lower deciles, indicating that high-frequency customers are more likely to consume obscure titles in the tail of the distribution. The values are only slightly above one, but the differences are meaningful considering the scale of the variables. For music, a one-unit increase in customer frequency, measured by the total number of plays for that customer, increases the number of titles consumed in the bottom decile with 0.6%, versus only 0.1% in the two top deciles, or at a ratio of six to one. These findings provide direct support for the natural monopoly idea that heavy category consumers constitute a disproportionately large portion of the audience for obscure titles or, correspondingly, that light category consumers comprise a disproportionately large portion of audience for popular titles.

The same pattern of higher estimates for lower deciles and lower estimates for the higher deciles occurs for the *genre_concentration* covariate (β_2), again in both data contexts, suggesting that customers that are focused on a smaller set of genres tend to consume more obscure products. For the music data, the number of titles consumed increases more than six-fold with a one-unit increase in the *genre_concentration* variable in the lowest decile, as opposed to more than four-fold in the highest decile, in the music data. This, again, is in line with expectations. A surprising result occurs for DVDs: whereas the estimates for the *genre_concentration* coefficient are significantly greater than one in decile two through six, indicating a positive relationship between genre concentration and consumption of titles, estimates significantly lower than one are found in decile zero, one, and seven through nine. The latter suggests that the number of the most obscure and the most popular titles a customer chooses actually *decreases* the more that customer concentrates on certain genres. It is not clear exactly what causes this u-shaped pattern, but the relatively weak presence of genre specialists rather than “omnivores” in the lowest two deciles may be related to the smaller assortments in some genres—it could be that customers would prefer to pursue titles within their favorite category, but are in some instances running out of viable options. The DVD rental firm’s smaller assortment relative to the music service could be the underlying cause.

Regarding the customer control variables, the estimates for β_3 , which includes *subscription_unlimited* and *subscription_to-go* in the music context as well as *subscription_3-or-more-at-a-time* and *subscription_unlimited-per-month* in the DVD setting, are significant, greater than one, and generally decrease as the deciles increase. The results confirm the expected pattern that, first, less restrictive plans generally go hand-in-hand with a higher level of consumption across deciles and, second, such plans are particularly closely associated with higher levels of consumption of the more obscure products. In the music context, the estimate for the *subscription_duration* covariate, measuring the length of a customer’s subscription, is significant but not meaningfully different from one. In the DVD context, the covariate is only slightly higher than one, albeit not in the highest deciles, suggesting that customers with a longer tenure may be slightly heavier consumers of obscure products. Finally, also in the DVD context, the estimates for *subscription_free-trial* suggests that customers engaged in a free trial consume fewer titles than those who do

not, judging by the finding of estimates that are lower than one. For instance, the coefficient is 0.51 for the highest decile, which implies that free-trial customers consume, on average, about 50% fewer titles in that decile than other customers. In the lower deciles, however, the estimates for *subscription_free-trial* are not statistically significant.

Appreciation Partly by means of summary of the above results, and partly as a transition from a focus on consumption to one on appreciation, **Figure 4** depicts three, as this study shows related, patterns in the raw DVD data. Naturally, first, the percentage of customers renting a title in a given decile increases the higher the decile: it gradually jumps from 11% in the lowest decile to 91% in the highest decile. Second, the average number of titles shipped to these customers decreases from roughly 50 titles in the six-month study period for customers in the lowest decile to just over 20 titles for those in the highest decile. This is in correspondence with the natural monopoly concept. Now focusing on the pattern in appreciation, third, the average rating given by these customers increases the more the focus is on higher deciles. Specifically, the rating jumps from roughly 3.00 for titles in the lowest two deciles to roughly 3.40 for titles in the highest decile. The raw data thus suggests that, in line with the double jeopardy phenomenon, obscure titles are on average appreciated less than more popular titles.

Could this finding be explained by the fact that some of the heavier consumers who choose more obscure products are more critical and therefore give lower ratings across the board, thereby having a disproportionate large negative effect on ratings for obscure products? Offering more detailed insights, **Table 4** presents average ratings and corresponding standard deviations for titles in a particular decile, as given by customers with at least one rental in that decile. Although lower average ratings go hand in hand with slightly higher variances, several patterns can be detected in the average scores. First, as already seen in **Figure 4**, titles in the lower deciles (the left-hand columns in the table) receive lower ratings than titles in the higher deciles (the right-hand columns). Second, both customers renting in the bottom deciles (the upper rows in the table) and those renting in higher deciles (the bottom rows) rate titles in this fashion. Each group gives lower ratings to titles in the lower deciles compared with titles in the higher deciles. Third, although the difference is subtle, it appears that customers of the more obscure products give

disproportionately low ratings to those products, and disproportionately high ratings to hit titles. That is, they appear to have a somewhat bigger magnitude in their scores.

The question remains to what extent these patterns are statistically significant, and what are the underlying drivers. **Table 5** presents estimation results for the regression model captured in equation (3), again for the DVD data. Two model variations are reported: Model I without, and Model II with interaction effects, with an adjusted R^2 of 0.11 and 0.12, respectively. In Model I, the coefficient for *popularity*, β_1 , is 0.08 and significant at a 1% level. The results thus confirm that, as the double jeopardy theory dictates, customer ratings are positively related to the popularity of a title across the customer base. The estimates for *frequency*, the total number of titles chosen by a customer as well as, in Model II, the interaction term of (*popularity * frequency*) are also positive and significant at a 1% level. On the one hand, the frequency with which a person consumes therefore is a direct predictor of her ratings—heavier consumers give slightly higher average ratings. As explained in the “Modeling Approach” section, the positive coefficient for β_2 , the coefficient for *frequency*, is intuitive in its own right, and helps strengthen the double-jeopardy result. On the other hand, as indicated by the value for β_3 , the coefficient for (*popularity*frequency*), frequency is a moderator of the relationship between a title’s popularity and its customer rating: the higher the frequency, the stronger the link between the demand and appreciation for a title. This suggests that people who transact with greater frequency may indeed develop a better ability or greater tendency to differentiate products of varying levels of popularity.⁹ The combination of popularity, frequency, and their corresponding interaction term has a convex relationship with ratings.

The same findings apply to the role of the *genre_share* variable, which measures what share of a customer’s transactions belong to a given title’s genre. As indicated by the positive estimates for the coefficients β_4 (in Model I and II) and β_5 (in Model II), the variable is both a direct influence on customer ratings as well as a moderator of the relationship between a title’s popularity and its rating. Thus, customers who dedicate a larger share of their choices to a certain genre will, on average, appreciate titles in that genre more than those who choose the genre relatively infrequent, and differentiate more between

⁹ It is generally seen as desirable that the moderator and dependent variable are not correlated (Baron and Kenny 1986). The variable should therefore strictly speaking, be treated as a “quasi” moderator.

the most obscure and most popular products. In other words, they “know better” than to like the obscure products as much as the hits.

The coefficients for the variables related to a customer’s subscription, β_6 , reveal that customer ratings are not significantly related to *subscription_duration*, the length of a customer’s subscription, but are generally higher the less restrictive the customer’s subscription plan is. Parameter estimates for *subscription_3-or-more-at-a-time* and *subscription_unlimited-per-month* are 0.08 and 0.04, respectively, in Model I, and significant at a 1% level. The estimates for the title characteristics, β_7 , show that, on average, customers rate movies that qualify for a wider audience higher than those that are restricted to certain age groups. The coefficients for *release_rating_M* and *release_rating_R* are both significantly lower than zero, while *release_rating_PG* is significantly greater than zero, which means M and R-rated titles score lower ratings than G (the excluded dummy) and PG-rated titles. Additionally, the coefficient for *release_major_studio*, the dummy indicating that the title is released by a major studio, is positive and statistically significant from zero at a 1% level, suggesting that such titles garner higher ratings on average. The estimate for the year of a title’s first release, *release_year*, is not significantly different from zero.

CONCLUSION

Is the interest in popular “hit” and unpopular “niche” titles equally distributed among the customers of an online retailer, or is the attention for obscure products concentrated among a small group of fans? What are the characteristics of customers in the “head” and “tail” of the distribution? And how do customers’ choices for popular or obscure titles relate to their appreciation of those titles? In this study, inspired by the surge in attention for what has been dubbed the “long tail” phenomenon, I analyze data from online music service Rhapsody and DVD rental business Quickflix to examine this set of related questions.

As also shown in other recent studies on the impact of online channels (Brynjolfsson et al 2003, Brynjolfsson et al 2007, Elberse and Oberholzer-Gee 2007), the data reveal that customers collectively consume a vast array of products ranging from the most popular to the most obscure. In both the music

and DVD contexts, the distribution of transactions across titles shows a long, relatively flat tail. This study's main contribution lies in the insights it offers about the way in which individual consumers distribute themselves across the wide range of available products. Encouraging for proponents of the long tail idea, the analyses reveal that the majority of online consumers regularly opt for obscure products only chosen by a small group of others and (given the average assortments of bricks-and-mortar retailers) likely not available in offline stores. As hypothesized, consumption of the more obscure products is particularly prevalent among those people who consume with a higher frequency and who concentrate on a narrow selection of genres. Long tail consumers, it appears, are more likely to be heavy rather than light consumers, and more adequately labeled "genre fanatics" rather than "genre omnivores" in search of content that satisfies their larger appetite and matches their preference. These findings correspond with a view that decreased search costs cause less concentrated sales distributions: heavy consumers and those with a deeper knowledge of certain genres may be able to rely more on collaborative filtering tools such as recommendation engines that can point them to undiscovered titles based on their past choices, or may generally face fewer hurdles in searching for obscure titles that fit their tastes.

Casting doubt on the "democratizing" nature of online channels, however, I also show that even for consumers who regularly choose the most obscure products, hit products typically constitute the lion's share of their choices. In that respect, the findings provide compelling support for a mass behavior "theory of exposure" first described by McPhee (1963) nearly half a century ago, and in particular the theory that a disproportionately large share of the audience for popular products consists of relatively light consumers and, correspondingly, that a disproportionately large share of the audience for obscure products consists of relatively heavy consumers. The pattern that hit products "naturally monopolize" light consumers is evident in both the music and DVD context. Additionally, in line with what McPhee termed "double jeopardy" and described as "the larger the proportion of the people that is not familiar with a given alternative, then the less likely are those who are familiar with it to like it especially," consumers of obscure products generally appreciate those products less than the more popular products. While McPhee based his theories on research in settings with typically only up to a dozen competing

alternatives, I believe this study is the first to demonstrate the existence of these phenomena in the context of the massive assortments often found in online channels.

The results further shed light on forces fueling the double jeopardy pattern. The findings reveal that the positive relationship between a customer's reported liking for a title and that title's popularity is moderated by two possible operationalizations of that customer's expertise, namely the frequency with which that customer transacts and his or her exposure to the title's genre. In turn, this suggests that a customer's familiarity with comparable alternatives may drive ratings for superior popular products up, and those for inferior obscure products down.

Overall, this study thus paints a balanced picture of the impact that online channels are having on market demand. On the one hand, in line with the long tail view, there is abundant evidence for a highly fragmented market with a large collection of obscure products that each command some attention from consumers. On the other hand, however, hit products remain a dominant force—even for customers that venture deep into the tail of obscure products, more popular titles constitute the lion's share of titles consumed—and hits are, on average, appreciated more than obscure products.

Managerial Implications The findings have important managerial implications, particularly for online retailers or aggregators who attempt to profit from long tail demand. Some of the most successful online businesses claim to derive a large share of their income from the collective demand for obscure items (The Economist 2005). For example, a former employee of Amazon, an online retailer best known for its massive collection of books, was quoted as saying: “We sold more books today that didn't sell at all yesterday than we sold today of all the books that did sell yesterday” (Anderson 2006). Netflix, a DVD-rental-by-mail business, touts its large assortment of video titles, many of which were never released in theaters or on television. One key take-away of this study is that online retailers' best customers—those that consume with a higher frequency—likely are more interested in wide assortments. However, the caveat is that online retailers should not expect their “long tail” customers to predominantly opt for obscure products and like those more than the hits—in fact, the opposite is more likely. Online

businesses should take this into account when managing customer expectations and satisfaction, particularly because the latter is likely correlated with long-term profitability.

The continued dominance of hit products is noteworthy in light of their business models, which rely in large part on stimulating demand for obscure products that, as discussed, typically come with higher profit margins. In media and entertainment industries, less popular products are often cheaper to acquire for retailers, or such products are older and therefore have more fully depreciated, while popular products are frequently used as loss leaders,¹⁰ leading to higher per-unit profits for the more obscure products (also see Brynjolfsson et al 2003). But even though less popular products thus often are more attractive for retailers from a margin perspective, building a business revolving solely around a wide collection of niche products may, the findings suggest, be difficult. The extremely low demand for a large group of the most obscure products coupled with the non-zero costs of making those products available—in the case of Rhapsody, for instance, checking who has the rights to make older music available can be very time-consuming and therefore costly—points to further difficulties in successfully executing a long-tail only model. All in all, even for online retailers that make much of their large assortments and their ability to tap into the long tail, a select group of hit products likely remains critical to their bottom line.

Future Research Directions

The focal products in this study, music and movies, can both be characterized as “fad” or “fashion” products with relatively short product lifecycles. They share many similarities but also some important differences, including that music is typically played repeatedly whereas movies are mostly rented once. One obvious future research avenue is to examine other product contexts, so as to better understand to what extent and how these differences impact hit and long-tail consumption in online businesses. While a continued focus on subscription models like those of Rhapsody and Quickflix is certainly called for, a second possible future research direction is to examine the behavior of customers under different business models and settings, perhaps most notably settings in

¹⁰ The seventh book in the Harry Potter series is a recent example. According to the New York Times (2007), Scholastic, its American publisher, had set a suggested retail price of \$34.99, but Barnes & Noble priced the book at \$20.99 (a 40% discount) and Amazon offered it for \$17.99 (a saving of 49%).

which consumers pay for each product consumed, or business-to-business settings such as Google's paid search business which reportedly heavily relies on long tail demand. Such research may help explain why this study's findings on the role of frequency differ from those by Brynjolfsson et al (2007). Third, while this study cannot make any claims about the causality of certain findings, most notably the relationship between frequency and genre concentration versus the choice for obscure products as well as the correlation between popularity and appreciation, a possible research extension would be to examine these relationships over time so as to assess causes and consequences. Finally, it may be worthwhile to more explicitly relate customer demand dynamics to profitability, ideally using data that considers the exact profit margins per product. Anecdotal evidence provided by media executives suggests a strong link between a title's rank and profit margin, but more comprehensive testing is called for.

REFERENCES

- Alba, Joseph, John Lynch, Barton Weitz, Chris Janiszewski, Richard Lutz, Alan Sawyer, and Stacey Wood (1997). Interactive Home Shopping: Incentives for Consumers, Retailers, and Manufacturers to Participate in Electronic Marketplaces. *Journal of Marketing* 61, 38-53.
- Anderson, Chris (2004). The Long Tail. *Wired Magazine*. Issue 12.10, October 2004.
- Anderson, Chris (2006). *The Long Tail: Why The Future of Business Is Selling Less of More*. New York, NY: Hyperion.
- Anderson, Chris (2007). Personal Communications. August 18, 2007.
- Bakos, J. Yannis (1997). Reducing Buyer Search Costs: Implications for Electronic Marketplaces. *Management Science* 43(12), 1676-1708.
- Bakos, J. Yannis (2001). The Emerging Landscape for Retail E-Commerce. *Journal of Economic Perspectives*, 15(1), 69-80.
- Bakos, Y. and Brynjolfsson, E. (2000). Bundling and Competition on the Internet. *Marketing Science* 19(1), 63-82.
- Baron, Reuben M., David A. Kenny. 1986. The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6), 1173–1182.
- Barwise, T. Patrick (1986). Repeat-Viewing of Prime-Time US Programs. *Journal of Advertising Research*, 28 (August-September), 9-14.
- Barwise, T. Patrick and Andrew S.C. Ehrenberg (1984). The Reach of TV Channels. *International Journal of Research in Marketing*, 1(1), 37-49.
- Bear Stearns (2006). Entertainment Industry: The Long Tail. Bear Stearns Equity Research, November 2006.
- Bear Stearns (2007). Entertainment Industry: A Longer Look at the Long Tail. Bear Stearns Equity Research, July 2007.
- Boatwright, Peter and Joseph C. Nunes (2001). Reducing Assortment: An Attribute-Based Approach. *Journal of Marketing* 65(July), 50–63.
- Broniarczyk, Hoyer and McAlister (1998). Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction. *Journal of Marketing Research* 35(2), 166-176.
- Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith (2003). Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers. *Management Science*, 49 (11), 1580–1596.
- Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith (2006). From Niches to Riches: Anatomy of the Long Tail. *MIT Sloan Management Review* 47(4), 67-71.

- Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Duncan Simester (2007). Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales. Working Paper, MIT, February 2007.
- Chernev, Alexander (2003). When More is Less and Less is More: The Role of Ideal Point Availability and Assortment in Consumer Choice. *Journal of Consumer Research* 30 (September), 170–183.
- Coleman, J.S. (1964). *Introduction to Mathematical Sociology*. Glencoe: The Free Press.
- Dhar, Ravi (1997). Consumer Preference for a No-Choice Option. *Journal of Consumer Research* 24 (September), 215–231.
- Ehrenberg, Andrew S. C., Gerald J. Goodhardt and T. Patrick Barwise (1990). Double Jeopardy Revisited. *Journal of Marketing* 54(3), 82-91.
- Elberse, Anita, and Felix Oberholzer-Gee (2007). Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales. Harvard Business School Working Paper, 07-015.
- Frank, Robert H., and Cook, Philip J. (1995). *The Winner-Take-All Society*. New York: The Free Press.
- Gini, Corrado (1921). Measurement of Inequality and Incomes. *The Economic Journal*, 31, 124-126.
- Goodhardt, Gerald J., Andrew S. C. Ehrenberg, and M.A. Collins (1975). *The Television Audience: Patterns of Viewing*. Aldershot, UK: Gower Press.
- Gourville, John T. and Dilip Soman (2005). Overchoice and Assortment Type: When and Why Variety Backfires. *Marketing Science*, 24 (3), 382–395.
- Greene, William H. (1997). *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ.
- Häubl, Gerald and Valerie Trifts (2000). Consumer Decision Making In Online Shopping Environments: The Effects of Interactive Decision Aids. *Marketing Science* 19(1), 4-21.
- Hoch, Stephen J., Eric T. Bradlow and Brian Wansink (1999). The Variety of an Assortment. *Marketing Science* 18(4), 527–546.
- Huffman, Cynthia and Barbara E. Kahn (1998). Variety For Sale: Mass Customization or Mass Confusion. *Journal of Retailing* 74(4), 491–513.
- Iyengar, Sheena S. and Mark R. Lepper (2000). When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology* 79(6), 995–1006.
- Kahn, Barbara E. (1995). Consumer Variety-Seeking Among Goods and Services. *Journal of Retailing and Consumer Services* 2(3), 139–148.
- Kahn, B., D. R. Lehmann (1991). Modeling Choice Among Assortments. *Journal of Retailing* 67(3) 274–299.
- Lehmann, Donald R. (1998). Customer Reactions to Variety: Too Much of a Good Thing? *Journal of the Academy of Marketing Science* 26(1), 62–65.
- Lynch, John G. Jr. and Dan Ariely (2000). Wine Online: Search Costs Affect Competition of Price, Quality, and Distribution. *Marketing Science* 19 (1), 83-103.

- MacKinnon, James G. and Halbert White (1985). Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. *Journal of Econometrics*, 29 (September), 305-326.
- McPhee, W. N., 1963. *Formal Theories of Mass Behaviour*, The Free Press of Glencoe, New York.
- Napoli, Philip M. (1999). Deconstructing the Diversity Principle. *Journal of Communication*, 49(4), 7-34.
- Napoli, Philip M. (2003). *Audience Economics: Media Institutions and the Audience Marketplace*. New York: Columbia University Press.
- Neuman, W. Russell (1991). *The Future of the Mass Audience*. New York: Cambridge.
- Oestreicher-Singer, Gal and Arun Sundararajan (2006). *Network Structure and the Long Tail of Electronic Commerce*. Working Paper, New York University, August 2006.
- Rosen, S. (1981). The Economics of Superstars. *The American Economic Review*, 71(5), 845-858.
- Salganik, Matthew J., Peter Sheridan Dodds, Duncan J. Watts (2006). Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science* 311, 854-856.
- The Economist (2005). *Profiting From Obscurity*. May 5, 2005.
- The New York Times (2007). 'Harry' and the Strange Logic of Book Discounters. July 28, 2007.
- The Wall Street Journal (2006a). It May Be a Long Time Before the Long Tail Is Wagging the Web. July 26.
- The Wall Street Journal (2006b). Many Companies Still Cling To Big Hits To Drive Earnings. August 2.
- Webster, J. (2005). Beneath the Veneer of Fragmentation: Television Audience Polarization in a Multichannel World. *Journal of Communication*, 366-382.

Table 1: Descriptive Statistics

Music (Rhapsody)

Variable	N	Mean	Median	SD	Min	Max
Customers						
<i>average_rank_i</i>	64,463	76,639.3	47,780.3	87,022.4	1	781,754.0
<i>frequency_i</i>	64,463	498.95	57.00	1,146.97	1	37,580.00
<i>genre_concentration_i</i>	64,463	0.46	0.49	0.23	0	0.92
<i>subscription_length_i</i>	64,463	331.84	268.00	290.46	0	1,793.00

Variable	N	%
Customers		
<i>subscription_unlimited</i>	64,463	39%
<i>subscription_to-go</i>	64,463	16%

DVDs (Quickflix)

Variable	N	Mean	Median	SD	Min	Max
Customers						
<i>average_rank_i</i>	15,736	2,605.15	2,209.36	2,101.32	1.00	14,798.00
<i>frequency_i</i>	15,736	20.50	9.00	26.86	1.00	286.00
<i>genre_concentration_i</i>	15,736	0.40	0.36	0.22	0.05	1.00
<i>subscription_length_i</i>	15,736	265.10	177.00	236.81	0.00	1,178.00

Titles

<i>year_of_theatrical_release_j</i>	15,741	1992	1999	14	1914	2006
---	--------	------	------	----	------	------

Transactions

<i>rating_{ij}</i>	65,064	3.34	3.00	1.10	0.00	5.00
----------------------------	--------	------	------	------	------	------

Variable	N	%
Customers		
<i>subscription_3-or-more-at-a-time_i</i>	15,736	41%
<i>subscription_unlimited-per-month_i</i>	15,736	60%
<i>subscription_free-trial_i</i>	15,736	3%
Titles		
<i>release_major_studio_j</i>	5,595	16%
<i>release_rating_pg_j</i>	15,741	25%
<i>release_rating_m_j</i>	15,741	44%
<i>release_rating_m15_j</i>	15,741	1%
<i>release_rating_r_j</i>	15,741	4%

Note: The tables present descriptive statistics for the music and DVD data. In subsequent analyses, missing values for the variable *release_major_studio* are set to zero.

Table 2: Regression Model for the Average Rank of Titles Consumed

Music

Dependent variable = $\log(\text{average_rank})$				
Symbol	Coefficient of ...	Estimate	SE	
β_0	<i>intercept</i>	9.908	0.032	***
β_1	$\log(\text{frequency})$	0.168	0.003	***
β_2	$\log(\text{genre_concentration})$	0.101	0.006	***
$\beta_{3.1}$	$\log(\text{subscription_duration})$	0.032	0.005	***
$\beta_{3.2}$	<i>subscription_unlimited</i>	0.452	0.017	***
$\beta_{3.3}$	<i>subscription_to-go</i>	0.268	0.019	***
N = 64,463				
Adjusted R² = 0.17				

DVDs

Dependent variable = $\log(\text{average_rank})$				
Symbol	Coefficient of ...	Estimate	SE	
β_0	<i>intercept</i>	5.435	0.059	***
β_1	$\log(\text{frequency})$	0.524	0.015	***
β_2	$\log(\text{genre_concentration})$	0.565	0.026	***
$\beta_{3.1}$	$\log(\text{subscription_duration})$	-0.002	0.009	
$\beta_{3.2}$	<i>subscription_3-or-more-at-a-time</i>	0.052	0.025	**
$\beta_{3.3}$	<i>subscription_unlimited-per-month</i>	0.198	0.027	***
$\beta_{3.4}$	<i>subscription_free-trial</i>	-0.127	0.054	**
N = 15,819				
Adjusted R² = 0.21				

Note: The tables present log-log regression coefficients and standard errors for the model captured in equation (1) for the music and DVD data. The dependent variable is the log of the average rank of the titles involved in a customer's transactions. Three asterisks (***) denote significance at a 1% level, two (**) at a 5% level, and one (*) at a 10% level.

Table 3: Negative Binomial Regression Model for the Number of Titles Consumed in a Given Decile

Music

Dependent variable = titles		Decile									
Symbol	Coefficient of ...	(0) 0-10%	(1) 10-20%	(2) 20-30%	(3) 30-40%	(4) 40-50%	(5) 50-60%	(6) 60-70%	(7) 70-80%	(8) 80-90%	(9) 90-100%
β_1	<i>frequency</i>	1.006 (.000)***	1.006 (.000)***	1.004 (.000)***	1.003 (.000)***	1.002 (.000)***	1.003 (.000)***	1.002 (.000)***	1.002 (.000)***	1.001 (.000)***	1.001 (.000)***
β_2	<i>genre_concentration</i>	6.674 (.688)***	6.024 (.752)***	5.990 (.958)***	6.598 (.834)***	5.149 (.788)***	5.293 (.735)***	4.425 (1.036)***	4.288 (.790)***	4.729 (.565)***	4.506 (.900)***
$\beta_{3,1}$	<i>subscription_duration</i>	1.000 (.000)***	1.000 (.000)***	1.000 (.000)***	1.000 (.000)***	1.000 (.000)***	1.000 (.000)***	1.000 (.000)***	1.000 (.000)***	1.000 (.000)***	1.000 (.000)***
$\beta_{3,2}$	<i>subscription_unlimited</i>	3.814 (.108)***	3.838 (.109)***	3.889 (.099)***	4.009 (.095)***	4.011 (.089)***	3.955 (.081)***	3.919 (.073)***	3.978 (.067)***	3.773 (.055)***	3.034 (.027)***
$\beta_{3,3}$	<i>subscription_to-go</i>	2.941 (.094)***	2.939 (.095)***	2.995 (.088)***	3.107 (.085)***	3.077 (.079)***	3.154 (.075)***	3.158 (.069)***	3.411 (.067)***	3.366 (.058)***	3.075 (.032)***
	N =	64,463	64,463	64,463	64,463	64,463	64,463	64,463	64,463	64,463	64,463
	Pseudo R² =	0.13	0.13	0.13	0.12	0.12	0.12	0.12	0.12	0.12	0.14

Note: The table presents negative binomial regression coefficients for the model captured in equation (2) for the music data. The dependent variable is the number of titles consumed that belong to a given decile. The reported coefficients are incidence rate ratios. Standard errors are in parentheses. Three asterisks (***) denote significance at a 1% level, two (**) at a 5% level, and one (*) at a 10% level.

Table 3: Negative Binomial Regression Model for the Number of Titles Consumed in a Given Decile (Continued)

DVDs

Dependent variable = titles		Decile									
Symbol	Coefficient of ...	(0) 0-10%	(1) 10-20%	(2) 20-30%	(3) 30-40%	(4) 40-50%	(5) 50-60%	(6) 60-70%	(7) 70-80%	(8) 80-90%	(9) 90-100%
β_1	<i>frequency</i>	1.018 (.000)***	1.019 (.000)***	1.018 (.000)***	1.016 (.000)***	1.016 (.000)***	1.013 (.000)***	1.014 (.000)***	1.013 (.000)***	1.010 (.000)***	1.007 (.001)***
β_2	<i>genre_concentration</i>	0.975 (.067)***	0.912 (.099)***	1.464 (.078)***	1.362 (.062)***	1.308 (.076)***	1.586 (.064)*	1.207 (.048)***	0.907 (.054)***	0.724 (.067)***	0.727 (.087)***
$\beta_{3,1}$	<i>subscription_duration</i>	1.002 (.000)***	1.001 (.000)***	1.001 (.000)***	1.001 (.000)***	1.001 (.000)***	1.001 (.000)***	1.001 (.000)***	1.001 (.000)***	1.000 (.000)***	1.000 (.000)***
$\beta_{3,2}$	<i>subscription_3-or-more-at-a-time</i>	1.984 (.081)***	2.428 (.095)***	2.361 (.093)***	2.167 (.082)***	2.207 (.088)***	1.825 (.073)***	1.860 (.080)***	1.869 (.083)***	1.685 (.081)***	1.491 (.084)***
$\beta_{3,3}$	<i>subscription_unlimited-per-month</i>	2.842 (.047)***	2.438 (.055)***	2.248 (.061)***	1.995 (.062)***	1.980 (.060)***	1.633 (.062)***	1.645 (.067)***	1.600 (.069)***	1.435 (.070)***	1.251 (.085)***
$\beta_{3,4}$	<i>subscription_free-trial</i>	0.779 (.323)	0.702 (.173)	0.706 (.179)	0.717 (.132)*	0.702 (.132)*	0.666 (.105)**	0.711 (.111)**	0.638 (.085)***	0.537 (.057)***	0.511 (.027)***
	N =	15,819	15,819	15,819	15,819	15,819	15,819	15,819	15,819	15,819	15,819
	Pseudo R² =	0.14	0.15	0.15	0.15	0.15	0.16	0.16	0.17	0.18	0.19

Note: The table presents negative binomial regression coefficients for the model captured in equation (2) for the DVD data. The dependent variable is the number of titles consumed that belong to a given decile. The reported coefficients are incidence rate ratios. Standard errors are in parentheses. Three asterisks (***) denote significance at a 1% level, two (**) at a 5% level, and one (*) at a 10% level.

Table 4: The Relationship Between Consumption and Appreciation: Ratings and Variances

DVDs

			Ratings Given By These Customers For Titles in each Decile						
		N	0-10%	10-20%	20-30%	...	70-80%	80-90%	90-100%
Customers Renting At Least One Title in a Decile	0-10%	1,687	3.00	2.99	3.00		3.33	3.36	3.47
			(1.09)	(1.04)	(1.01)	...	(0.92)	(0.89)	(0.83)
	10-20%	2,875	3.01	2.94	3.02		3.29	3.41	3.45
			(1.09)	(1.08)	(0.98)	...	(0.87)	(0.92)	(0.83)
	20-30%	3,932	3.01	3.03	3.00		3.30	3.36	3.44
			(1.06)	(0.99)	(1.05)	...	(0.91)	(0.94)	(0.84)

	70-80%	8,164	3.04	3.04	3.02		3.29	3.33	3.42
			(1.08)	(1.07)	(1.04)	...	(0.98)	(0.98)	(0.85)
	80-90%	9,896	3.02	3.03	3.03		3.29	3.31	3.38
			(1.06)	(1.07)	(1.05)	...	(0.95)	(1.02)	(0.85)
	90-100%	14,472	3.03	3.04	3.00		3.28	3.31	3.40
			(1.07)	(1.08)	(1.05)	...	(0.98)	(1.01)	(0.88)

Note: The table lists, for customers renting at least one title in a given decile, the average ratings given by these customers for titles in each decile. The standard deviations of these average ratings are captured in parentheses. The table presents only the top and bottom three deciles.

Table 5: Regression Model for the Appreciation of Titles Consumed

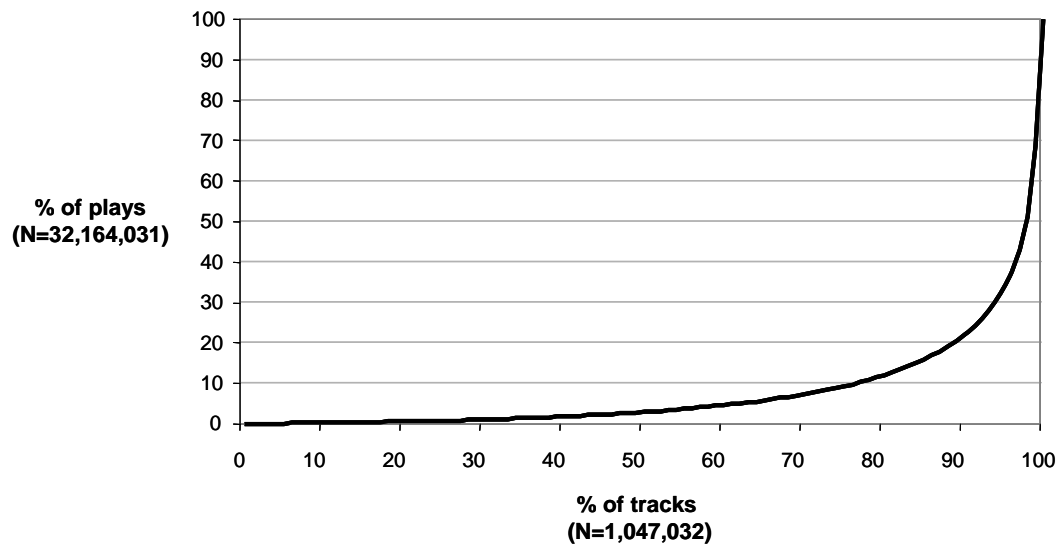
DVDs

Dependent variable = rating		Model I			Model II		
Symbol	Coefficient of ...	Estimate	SE		Estimate	SE	
β_0	<i>intercept</i>	2.249	0.750	***	2.197	0.751	***
β_1	<i>popularity</i>	0.082	0.014	***	0.047	0.016	***
β_2	<i>frequency</i>	0.002	0.000	***	0.001	0.000	***
β_3	<i>popularity * frequency</i>	--	--		0.002	0.001	***
β_4	<i>genre_share</i>	0.462	0.021	***	0.446	0.025	***
β_5	<i>popularity * genre_share</i>	--	--		0.171	0.130	***
$\beta_{6.1}$	<i>subscription_duration</i>	0.001	0.000		0.001	0.000	
$\beta_{6.2}$	<i>subscription_3-or-more-at-a-time</i>	0.081	0.014	***	0.081	0.014	***
$\beta_{6.3}$	<i>subscription_unlimited-per-month</i>	0.043	0.016	***	0.039	0.016	***
$\beta_{6.4}$	<i>subscription_free-trial</i>	-0.063	0.132		-0.054	0.132	
$\beta_{7.1}$	<i>release_major_studio</i>	0.074	0.010	***	0.074	0.010	***
$\beta_{7.2}$	<i>release_rating_PG</i>	0.108	0.015	***	0.107	0.015	***
$\beta_{7.3}$	<i>release_rating_M</i>	-0.028	0.014	**	-0.029	0.014	***
$\beta_{7.4}$	<i>release_rating_MA15+</i>	0.008	0.037		0.007	0.037	
$\beta_{7.5}$	<i>release_rating_R</i>	-0.410	0.025	***	-0.410	0.025	***
$\beta_{7.6}$	<i>release_year</i>	-0.000	0.000		-0.000	0.000	
		N=64,480			N=64,480		
		Adjusted R² =0.11			Adjusted R² =0.12		

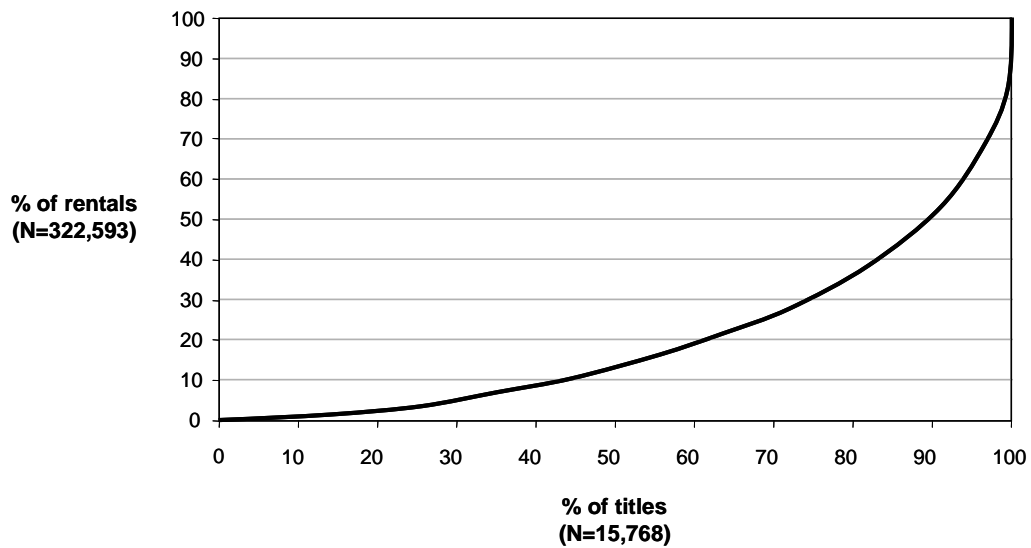
Note: The table presents regression coefficients and standard errors for the model captured in equation (3) for the DVD data. Model I excludes and Model II includes interaction effects. The dependent variable is a customer's rating of a title. For scaling purposes, the independent variable *popularity* is expressed as the total rentals for a title in thousands. Three asterisks (***) denote significance at a 1% level, two (**) at a 5% level, and one (*) at a 10% level.

Figure 1: The Distribution of Transactions across Titles

Music

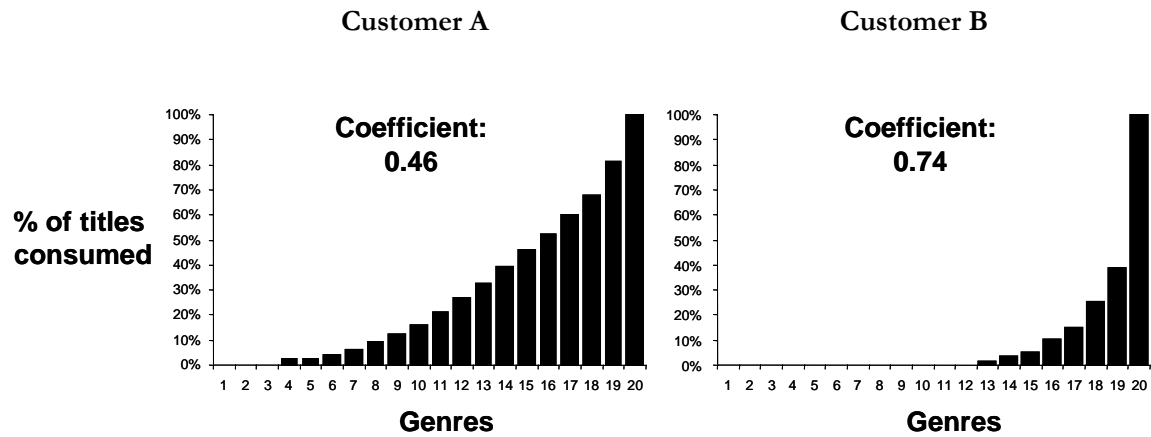


DVDs



Note: The figures plot the percentage of titles on the x-axis and the percentage of either plays (for the music data) or rentals (for the DVD data) on the y-axis.

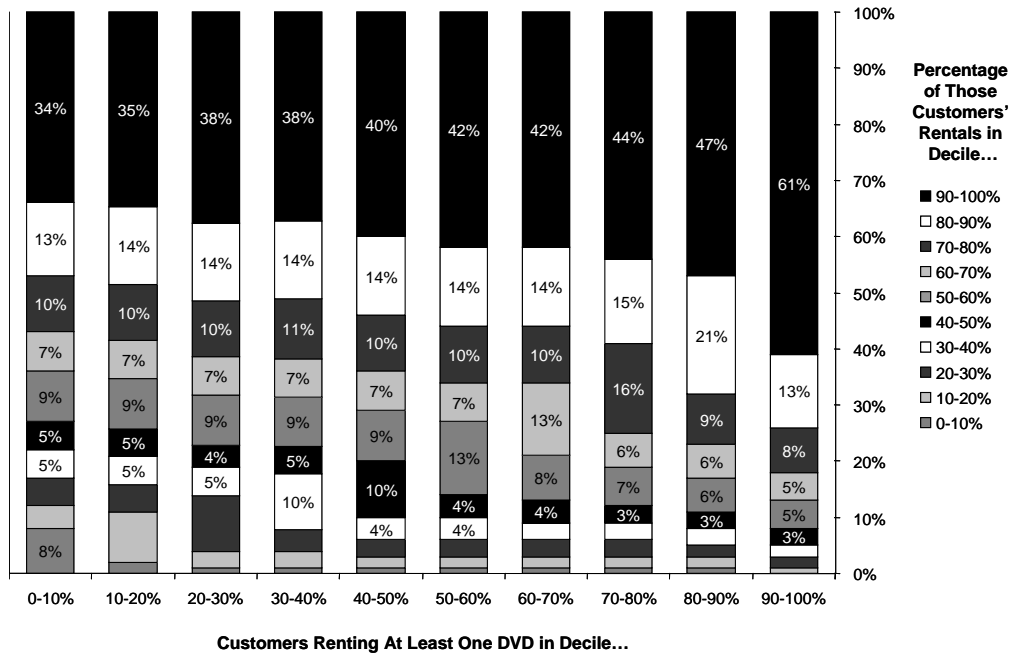
Figure 2: The Genre Concentration Measure: An Illustration



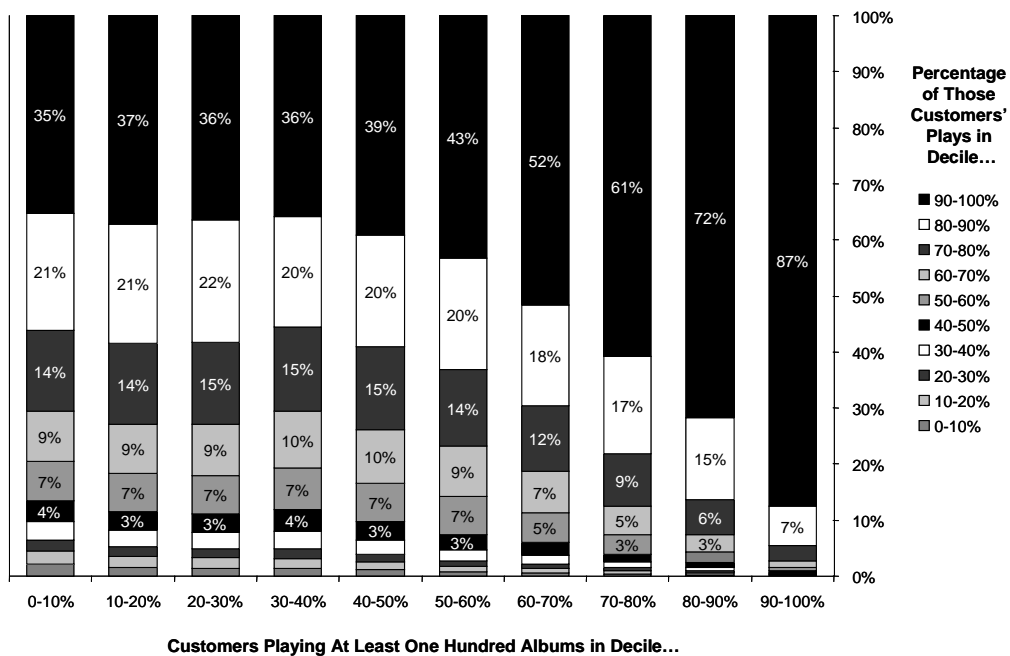
Note: The left graph represents Rhapsody Customer A who played a total of 314 tracks in 17 different genre categories, while the right graph represents Rhapsody Customer B who played a total of 535 titles in eight genre categories in the three-month study period. Both graphs plot the distribution of plays across genres. Customer A's distribution is more equal than Customer B's, hence Customer A's genre Gini coefficient is closer to zero than Customer B's. This coefficient forms the basis for the *genre_concentration* measure.

Figure 3: The Distribution of Obscure and Hit Titles across Consumers

DVDs



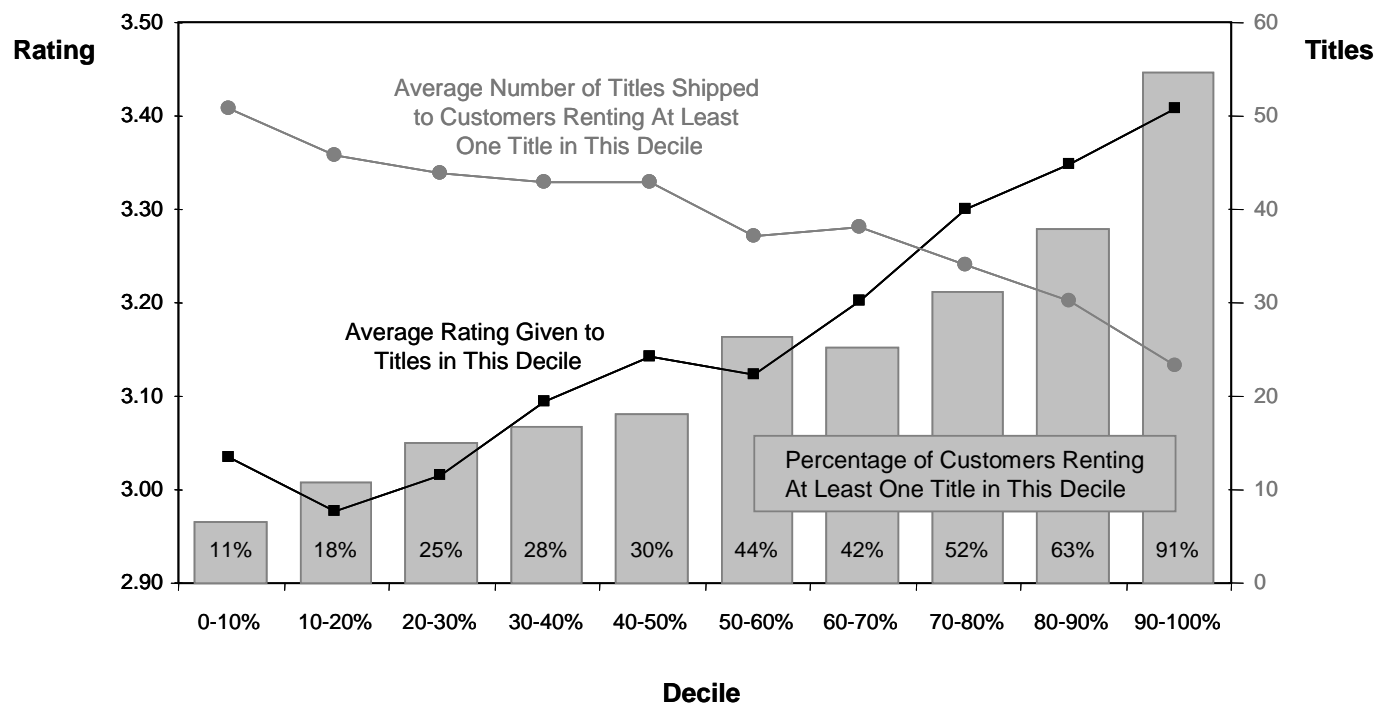
Music



Note: For the music and DVD data, the figures depict the distribution of customers across products with varying levels of popularity. The black bar on the top right of the top figure is to be read as “Of the customers renting at least one DVD in the top decile, i.e. at least one of the 10% most popular DVDs, 61% of their transactions consist of titles that belong to this top decile.”

Figure 4: Consumption and Appreciation of Obscure and Hit Titles

DVDs



Note: For the DVD data, the figures depict (1) the percentage of customers renting at least one title in a given decile, (2) the average number of titles shipped to those customers, and (3) the average rating given to titles in this decile.