THE ALLIANCE
INSEAD Wharton
UNIVERSITY of PENNSYLVANIA

Alliance Center
for Global Research
and Education

Faculty & Research
Working Paper

Plenty is no Plague, or is it? An Empirical Study of the Impact of Product Variety on Demand Concentration

Tom F. TAN
Serguei NETESSINE
Lorin HITT
2012/128/TOM/ACGRE

Plenty is no Plague, or is it?

An Empirical Study of the Impact of Product Variety on Demand Concentration

Tom F. Tan*

Serguei Netessine**

Lorin Hitt***

* PhD Candidate in Operations and Information Management at Wharton School, University of Pennsylvania, 3901 Locust Walk, MB 253, Philadelphia 19104, USA.
Email: fangyun@wharton.upenn.edu

** The Timken Chaired Professor of Global Technology and Innovation, Professor of Technology and Operations Management, Research Director of the INSEAD-Wharton Alliance at INSEAD Boulevard de Constance 77305 Fontainebleau, France.
E-mail: serguei.netessine@insead.edu

*** Professor at Wharton School, University of Pennsylvania, 571 Jon M. Huntsman Hall
Philadelphia 19104, USA. Email: lhitt@wharton.upenn.edu

Abstract

We empirically examine the impact of expanded product variety due to the adoption of the Internet on demand concentration using two large data sets from the movie rental industry, at both the movie level and the consumer level. We find that increasing product variety diversifies the demand away from both hits and niche products defined in absolute terms (e.g., the top/bottom 1,000 titles), but less significantly for hits than for niche products. Using relative terms (e.g., the top/bottom 10% of titles) to normalize the demand over time, we find that product variety increases the demand for hits and decreases the demand for niche movies, which is contrary to the celebrated \Long Tail" effect. We confirm our aggregate-level results using a consumer-level analysis and we conclude that new products appear much more quickly than consumers discover them, thus causing a \rich get richer" effect. Finally, we find no evidence that niche titles satisfy consumer tastes any better than popular titles and that a small number of heavy users are more likely to venture into niche products than light users are.

Keywords: Product Variety; Demand Concentration; Movie Rental; The Long Tail Effect; Product Rating; Customer Satisfaction; e-commerce; Search Cost; Online Behavior.

1 Introduction and Related Literature

Chris Anderson, editor-in-chief of *Wired Magazine*, coined the term "Long Tail effect" (Anderson, 2004) predicting that, due to the introduction of the Internet, niche products would comprise increasing market share, while the demand for hit products would continue to decrease. Part of the reason, according to Anderson, is that niche products would continue to better satisfy consumer preferences because consumers would continue to have more and more varying preferences, and the Internet's expanded product variety would make even the most obscure products available to the masses. Anderson (2006) also explains that new online recommendation systems help niche products quickly find their demand in the market once they are made available. The potential for the existence of the Long Tail effect is of great importance for product assortment decisions in a variety of industries, for advertising dollars spent on supporting this variety, and for supply chain management of these products on the Internet.

The Long Tail effect has recently generated widespread interest in academic circles. Brynjolfsson et al. (2006, 2010, 2011) present plausible factors that may drive the Long Tail effect, including both supply-side and demand-side effects. On the supply side, they suggest that the Internet reduces the production and distribution costs of niche products, creating more available products. These products can satisfy consumers' heterogeneous needs, thus driving the Long Tail effect. On the demand side, they note that both active and passive search and personalization tools lower search costs and hence facilitate finding niche products.

So far, many theories and evidence tend to focus on the impact of lower search costs, especially those enabled by the Internet, on demand concentration. Brynjolfsson et al. (2011) empirically analyze a retailer that offers the same products online and offline and find that the online store exhibits less concentrated demand because of its lower search costs. Cachon et al. (2008) predict that lowered search cost can further encourage firms to enlarge their assortment, which may contribute to increasing demand for niche products for the aforementioned reason that large variety satisfies heterogeneous demand. Moreover, Tucker and Zhang (2011) suggest that product popularity information, such as the number of people who have browsed the product, can disproportionately increase the appeal of niche products. In addition, Kumar et al. (2011) found that broadcasting movies on pay-TV can increase the awareness of unpopular movies, thus reducing the demand concentration of DVD sales. All these studies seem to suggest that lowering search costs leads to increasing demand for niche products.

On the other hand, several studies have questioned the premise of the Long Tail effect. Ghose and Gu (2006) argue that search costs are even lower for popular products than for niche products, which may limit the Long Tail effect. Hervas-Drane (2009) provide an analytical model to show that different search processes have mixed impacts on demand concentration. Moreover, Fleder and Hosanagar (2009) suggest that selection-biased recommendation systems can reduce sales diversity because these systems tend to recommend products with sufficient historical data (i.e., hits). From a field experiment, Fleder et al. (2009) further found that consumers buy a more similar mix of music after receiving recommendations than before.

Although whether or not lower search costs induce increasing demand for niche products remains a hotly debated topic, little research has evaluated the other original argument about the Long Tail effect, i.e., the effect of increasing product variety on demand concentration. Previous research often has implicitly assumed that expanding product variety will satisfy consumers' increasingly heterogeneous tastes, which causes the Long Tail effect (e.g., Brynjolfsson et al. 2006, 2010, 2011). Whether or not increased product variety really satisfies consumer needs remains an empirical question. In addition, the original Long Tail argument presumes that consumers can necessarily find what they really enjoy as product variety increases. However, it remains unclear whether or not consumers can really immediately discover new products, especially those niche products. In other words, although various online search tools have temporarily reduced search costs, expanded product variety may further escalate search costs. A limited number of studies have noted that omitting the impact of product variety may lead to misleading inferences about the Long Tail effect. Zhou and Duan (2011) study the interaction effect between online user reviews and product variety on software downloads. They find that the impacts of online user reviews, a key demand factor, are weakened by expanding product variety. Hinz et al. (2011) examine the impact of product variety on the online demand for individual video-on-demand sales per customer and find that product variety reduces the demand of each movie per customer. Nevertheless, they find that product variety has almost no impact on the Gini coefficients, a measure of demand inequality. Our data are both larger in terms of product variety and are collected over a much longer time horizon, allowing us to better capture the effects of changing product variety over time.

Although little research has been done to explicitly and rigorously examine the impact of product variety on demand concentration, whether or not the demand for niche products increases amid an ever-changing product variety is a fundamental question for decision-makers in operations, marketing and finance, particularly when they face the prospect of further penetration of the Internet channel, which offers expanding product variety and new recommendation systems.

In this paper we use large data sets from the movie rental industry to empirically evaluate the impact of product variety on the demand concentration at both the movie level and the consumer level. When product variety is large, the demand for any one product tends to be smaller than when product variety is small (Hinz et al., 2011). Likewise, when the consumer base is large, consumers learn about new products more quickly than when the consumer base is small. In this case, two competing effects might be observed: 1) consumers discover the obscure products as they appear and 2) new products appear, possibly so quickly that most consumers have no time to discover them. Our paper empirically addresses the question of which effect dominates.

Our study also contributes to the literature by measuring product popularity in terms of relative ranking. Anderson (2006) primarily refers to product popularity in absolute terms, e.g., the top 100 or the top 1,000 for hits. In his own words, "number one is still number one, but the sales that go with that are not what they once were". This absolute popularity measure has been used in much of the Long Tail literature. This absolute definition of product popularity is static, which implicitly excludes the impact of increasing product variety. The static definition would certainly reflect product popularity in a channel where product variety is relatively stable and where all products are consumed, such as in a brick-and-mortar store. However, product variety has skyrocketed during the Internet age, and more products than ever are not being discovered by consumers (Brynjolfsson et al., 2003). Such a dramatic increase in product variety is likely to create demand diversification. For example, given a choice set of only five movies, people may tend to concentrate their demand on one movie whose popularity rank is number 1 or equivalently on the top 20%. However, out of a wider choice set of 500 movies, the demand may be concentrated on 100 movies whose popularity ranks in the top 100 or also in the top 20%. This example causes a conflicting definition of hits and niche products amid different sizes of product variety at different points in time. Should we classify the top 20%, which is respectively the top one out of five movies and the top 100 out of 500 movies as the hits, or should we restrict the label of hits to only the top one movie regardless of the total variety? We argue that in order to compare the dynamics of demand distribution, we should normalize the definitions of hit and niche products by the changing product variety. Demand for either hit or niche products is composed of both the number of titles and the number of times each of these titles is consumed. The absolute definition measures the demand only in terms of the number of times each title is consumed. The relative definition, however, reflects both components of demand. Therefore, we propose a relative definition of product popularity, e.g., hits are defined as the top 10% of all the titles during a month.

Our relative popularity measure also differs from relevant studies by Zhou and Duan (2011) and Elberse and Oberholzer-Gee (2008). In their quantile regressions, the authors define movie popularity as the particular sales percentiles of all the movies across the entire study span. For example, in Elberse and Oberholzer-Gee (2008), the most popular category sold more than 125 copies per week, which corresponded to the top 90th percentile of weekly sales across the six-year period. Our study also differs from Hinz et al. (2011). They focus on the impact of product variety on demand for an average product, while we study its impact on particular parts of the demand distribution, i.e., hits and niches.

Among our findings, we highlight that when movie popularity is measured in absolute terms, e.g., the top 10 or the top 100 for hits, there is only partial evidence to support the Long Tail effect: product variety reduces the demand for hits (consistent with the Long Tail effect) but also for niche products (contrary to the Long Tail effect). The magnitude of this demand diversification effect, when taking product variety into account, is stronger for niches than for hits, which is against the Long Tail prediction. In addition, demand for the top 0.1% of movies increases approximately six times as fast as demand for the top 10%, indicating that demand for the "hits of the hits" continues to skyrocket. Unlike Hinz et al. (2011), who find no impact of product variety on Gini coefficients, we find that Gini coefficients increase as product variety changes: 1,000 new movies per month may increase demand inequality by 0.19%. We also find that product variety is positively associated with the monthly demand for the top deciles of products, unlike the demand for lower deciles, which is either negatively correlated or uncorrelated.

At the consumer level, we find that new products appear so quickly that most consumers have no time to discover or consume them, causing them to watch more and more hits. Furthermore, while Anderson (2004; 2006) argues that more and more consumers will choose niche products because they will tend to better satisfy consumer preferences, we reveal that consumers tend to be less satisfied with niche and less popular movies than with popular ones. We also find that it is mostly the heavy users or the "movie buffs", a small fraction of all consumers, that venture into niche movies.

To summarize, the contributions of this paper are three-fold. First, we explicitly examine the impact of product variety on demand concentration. Second, we propose to delineate two effects: demand diversification due to expanding product variety and consumers learning about new products. We suggest that, when measuring product popularity, one has to use relative measures to adjust for instantaneous active product variety. With this definition, we find that product variety increases demand concentration on hits, which is against the Long Tail prediction. Third, we study demand at the consumer level and find that new movies appear so quickly that most consumers have no time to discover them, and that niche movies do not satisfy consumer tastes better than hit movies.

2 Hypotheses Development

Previous studies argue that producers and retailers have increasing incentives to produce and stock niche products because of lowered production and inventory costs (e.g., Brynjolfsson et al. 2006, 2011). This increased product variety is further assumed to better and better satisfy consumer preferences because consumers will continue to have increasingly varying preferences, thus leading to the Long Tail effect (e.g., Anderson 2006). While there is little doubt that product variety generally increases over time and that technology such as the Internet and drop-shipping techniques allows companies to economically offer an even wider variety of products, it is less clear that consumers necessarily quickly discover these products, let alone actually consuming them. Large product variety is likely to make it more difficult to notice new products. In other words, a change from one to two options in the choice set can be easily noticed, but it takes a lot more effort to notice a change from 2,000 to 2,001 options. New products that have limited associated advertising budgets to create "buzz", particularly niche products, may disproportionately stay unnoticed, thus jeopardizing their demand.

In addition, although consumers have varying tastes and like to seek variety, they are less likely to examine all choices to find their "true" fit of tastes when they are faced with large product variety. Too many choices require more cognitive efforts to evaluate the attractiveness of alternatives (see Kuksov and Villas-Boas 2010 for a review), thus increasing search costs. When search costs are high, consumers tend to restrict their choice consideration to the products for which they have ex ante knowledge (Stigler, 1961; Rothschild, 1974). These products tend to be popular because they are more likely to have more buzz from advertising, promotion and word-of-mouth. As a result, demand is likely to be even more concentrated on these hits.

Finally, high-quality products may seem even more attractive within larger product variety. Simonson and Tversky (1992) suggest that adding extremely low-quality products into a consideration set increases the attraction of the higher-quality products. Therefore, introducing more lower-quality products, which tend to be niche products, is likely to make higher-quality products even more popular. In addition, when product variety is large, consumers are found to be more discriminating in terms of product quality (Bertini et al., 2012), which can further increase the demand for the hits.

For these reasons, we hypothesize that

HYPOTHESIS: *Product variety is positively associated with demand for hits and negatively associated with demand for niche products.*

3 Data

3.1 Research Setting and Data Collection

To examine our research hypothesis, we gathered data available from a company (a distributor) that leases and delivers movies to retailers for rental. Its clients represent approximately 30% of the entire U.S. movie rental retailers. This company collects the related rental information for the movies on a revenue-sharing basis. Our data consist of the DVD rental turns and movie characteristics at the movie level from January 2001 to July 2005.

In addition to the movie-level data (Dataset I), we also collected consumer-level data from

Netflix (Dataset II), a major U.S. online movie/TV series[1] rental service with annual revenues in excess of $1 billion in 2008. Dataset II was made available to the public during the Netflix Prize competition, which offered $1 million to the team that could use the data to create the most accurate movie recommendation system. The data set consists of the movie ratings submitted by consumers through the Netflix website from 2000 to 2005, encompassing the time period for which we have movie-level rental data. Netflix encourages its users to rate the movies they have watched both outside and within Netflix to improve its recommendations for them, so users have direct incentives to provide truthful and complete ratings. As a result, Shih et al. (2007) suggest that Netflix has the world's largest collection of accurate movie ratings.

We believe that our data provide rich grounds to study the impact of varying product variety on market concentration patterns. First, Dataset I is one of the most extensive sources of information on the movie rental industry among all related studies, as it includes the vast majority of movie titles distributed in the U.S. Second, the revenue-sharing contract ensures the accuracy of the reported movie rental turns through considerable computer monitoring and external verification of the results. Finally, Dataset II allows us to observe temporal changes in the popularity of the movies and in evolving customer preference. The combination of the two data sets allows observation at the movie and the consumer level, which is quite rare.

It is important to stipulate here that Dataset II only reflects the number of movies rated, but not all customers rate all movies that they watch. On the other hand, customers do not have to watch the movie at Netflix to be able to rate it. Admittedly, using rating data as a proxy for actual rental demand at Netflix can be inconclusive, but the rating data can provide insights into what movies consumers are aware of and are interested in. First, previous literature (see Chen et al. 2004) suggest a strong connection between product demand and the number of consumer reviews. In our own data, we find a correlation of about 0.5 between the monthly number of ratings and the rental turns among the matched movies. Second, unlike other review data that are known to have selection bias because users tend to review items they particularly like or dislike (see Hu et al. 2009; Dellarocas et al. 2010 ; Dellarocas and Wood 2008 and citations therein), pure ratings may avoid this bias because giving a rating is much less costly to a user than writing a review. In our data, we plot the histogram of the rating values on a scale from one to five and find the rating of

[1] In the rest of the paper, we refer to movies and TV series as well as other DVD content as "movies" for simplicity.

four to be the most frequent, followed by the ratings of three, five, two and one (see Figure 3 for more detail). The bell-shaped histogram seems to suggest that Netflix users may be less biased toward rating the movies that they extremely like or detest. Third, the recommendation system of Netflix directly incentivizes and facilitates its users to reveal their truthful and complete preference for movies to improve their recommendations. Finally, the movie-level results of the Netflix data endorse the main results of the rental data as a robustness check. For all these reasons, we proceed with utilizing ratings as a proxy for movie appeal and consumer preference to complement Dataset I at the consumer level, although it should be understood that we imply the number of ratings.

3.2 Measures and Controls

In much of our analysis, we first elect to work with monthly (instead of weekly or yearly) data and therefore we aggregate all variables at the monthly level. By doing so, we ensure both an adequate sample size in each month for each movie and enough observations over time for statistically significant estimates.

We are interested in studying varying demand for movies having different popularity levels. To reflect how popular movie j is at time t within a particular product offering set, we first rank the rental turns of each movie within each month in a descending order and use this rank as a proxy for movie popularity. Note that a higher (lower) rank indicates a less (more) popular movie. Then we propose two measures to categorize whether or not a movie is a hit. The first measure is the absolute ranking, e.g., the top 100, the top 1,000 movies, which is used in previous literature including Anderson (2004). Alternatively, we propose to rank movies in relative terms, e.g., the top 1%, the top 10%, thus adjusting for current product variety (the total number of movies rented this month).

In addition, we define $Variety_t$ as the total number of different movies that were rented during month t. Unlike the product assortment size in Hinz et al. (2011), which includes all product offerings, the $Variety_t$ variable reflects active product variety as many movies are not rented in a given month. We believe that active product variety is a more relevant variable than total variety, which includes DVDs with no rentals because 1) products that are not discovered by consumers (or that are discovered but forgotten) should not be taken into account when ranking popularity, and 2) it accounts for both product offering and consumer demand. Clearly, by using active rather than total product variety, we are more likely to find evidence of the Long Tail effect so, if anything, we are biasing our results against our hypothesis. Furthermore, we measure the demand for individual movies with the proxy $Share_{jt}$, which reflects movie j's market share of rental turns among all the rented movies within month t. This measurement allows us to adjust for possible changes in the consumer base. As a robustness check, we also compute the monthly Gini coefficient $Gini_t$, which is often used in social sciences as a measure of inequality in a distribution (e.g., Yitzhaki 1979; Lambert and Aronson 1993). A $Gini_t$ of zero indicates total equality during month t, while a value of one suggests maximal inequality.

To summarize, all variable definitions at the movie level are presented in Table 1.

Table 1: Movie-level Analysis Variable Definitions

Variable	Definition
$Variety_t$	Total number of rented movies during time period t.
$Share_{jt}$	Share of the number of times that movie j is rented among all the rented movies during time period t.
$Gini_t$	The Gini coefficient of demand distribution during month t.

In addition to the movie-level variables, we further define variables for our consumer-level analysis. We define $ProductRank_{it}$ as the average absolute ranking of the movies that consumer i rates in a given month t. In essence, $ProductRank_{it}$ i s a summary statistic of movie-level popularity. A high value of $ProductRank_{it}$ means that consumer i tends to watch more niche movies. We calculate the mean, the median, the top 10%, and the bottom 10% of the rankings to obtain more complete information about consumer choices. Furthermore, we divide these metrics by monthly product variety to obtain relative measurements. These relative measurements adjust for both increasing product variety and the skewness of demand distribution.

In order to control for consumer heterogeneity over time, we define $Frequency_{it}$ as the number of movies that user i rated in month t. In marketing, certain theoretical constructs such as the Dirichlet model suggest a strong link between purchase frequency and brand choice. In particular, it is often found that most consumers of a brand are low-frequency buyers (Chatfield and Goodhardt, 1975; Goodhardt et al., 1984). These light buyers often constitute the majority of the customers who purchase the popular brand (McPhee, 1963) because of the "superstar" effect (Rosen, 1981). In addition, McPhee (1963) explains that consumers who are familiar with the alternatives tend to consume the niche products. Therefore, consumers with high-consumption frequency are likely to consume more niche products than those with low-consumption frequency because the former may be better informed of the variety of products than the latter.

Furthermore, we define $RatingAverage_{it}$ and $RatingVariance_{it}$ as the average and the variance of the ratings that user i gives in month t. These two measurements are likely to reflect people's tastes and movie acceptance. For example, Clemons et al. (2006) demonstrate the relationship between variance of ratings and demand for products. Further, Hu et al. (2009) recommend controlling for the standard deviation of ratings as well as for two modes to overcome the consumer under-reporting bias. Since in our case the distribution of ratings is symmetric, we do not control for the modes. All user-level variables are defined in Table 2.

Table 2: Consumer-Level Analysis Variable Definitions

Variable	Definition
$ProductRank_{it}$	Popularity of the movies that user i rated in month t, measured as the mean and the median, the top 10%, and the bottom 10% of the movie rankings, both in absolute and relative terms.
$Frequency_{it}$	Number of movies rated by user i in month t.
$RatingAverage_{it}$	Average rating given by user i in month t.
$RatingVariance_{it}$	Variance of the ratings given by user i in month t.

3.3 Descriptive Statistics

Table 3 presents the descriptive statistics of the rentals by year. The product variety, which is the number of distinct movies rented at least once, substantially increased from 7,246 in 2001 to 25,488 in 2005, up approximately two and a half times. The total rentals also saw more than a threefold increase, going from 162 million turns in 2001 to 546 millions turns in 2004. For each title, the average turns seems to be stable. The skewness of the turns increased from 5.37 in 2001 to 6.94 in 2005, suggesting that the most popular movies are likely to contribute to an increasing market share. Furthermore, we observe that the minimum yearly turns per title dropped from 23 in 2001 to one in the following years, while the maximum yearly turns per title seem to be increasing from 728,526 in 2001 to over 1 million in 2004. This difference may suggest that the popular movies may be even more popular, while the obscure ones become even more obscure.

Table 3: Descriptive Statistics of Movie Rentals

Year	Product Variety	Total Rental Turns (in MN)	Avg Turns per Title	Skewness of Turns	Min Turns	Median Turns	Max Turns
2001	7,246	162	22,381.79	5.37	23	895	728,526
2002	10,975	245	22,319.38	5.40	1	845	933,998
2003	15,681	369	23,502.76	5.62	1	1,054	1,122,852
2004	23,255	546	23,459.19	6.02	1	2,461	1,019,122
2005*	25,488	352	13,812.03	6.94	1	1,555	682,397
Mean/year	16,529	335	21,095	5.87	5	1,362	897,379
Stdev/year	7,798	145	4,110	0.65	10	676	188,235

* We only observe seven months in 2005.

Figure 1 shows that the monthly product variety in Dataset I increased exponentially from January 2001 to July 2005, and that the rental turns increased linearly during the same period[2]. A relevant question is whether product variety is growing because a lot of brand new movies are being released or because consumers keep discovering previously released titles. Our data indicate that the number of brand new titles increased from 1,639 in 2001 to 3,879 in 2004, while the newly rented back catalog titles decreased from 5,607 in 2001 to 3,707 in 2004. These observations suggest that product variety growth is primarily due to the introduction of brand new products. The more precise answer to this question is complicated by the fact that many movies are released on DVD later than in theaters, but this gap continues to decrease over time. Further, most popular movies are released in several DVD versions at different points in time which makes it hard to exactly delineate rentals of “old” vs. “new” movies.

[2]Some movies may be removed from the market over time. Even though some movies were available in the market throughout the year, they may have been rented at least once only in a few months. Consequently, the yearly product variety in Table 3 is greater than or equal to the monthly product variety in the same year.

Figure 1: Monthly Product Variety and Rentals

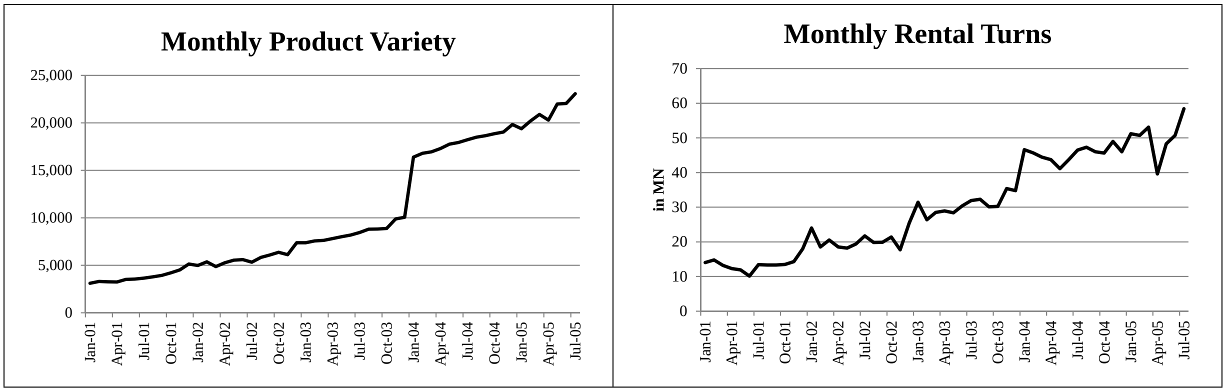


Figure 2 illustrates the distribution of rental turns after pooling all observations from 2001 to 2005. Demand is highly concentrated on a few titles, i.e., the hits: the top 10% of movies constitute close to 80% of total rental turns; the top 20% of movies contribute to slightly over 92%.

Figure 2: Rental Distribution (2001-2005)

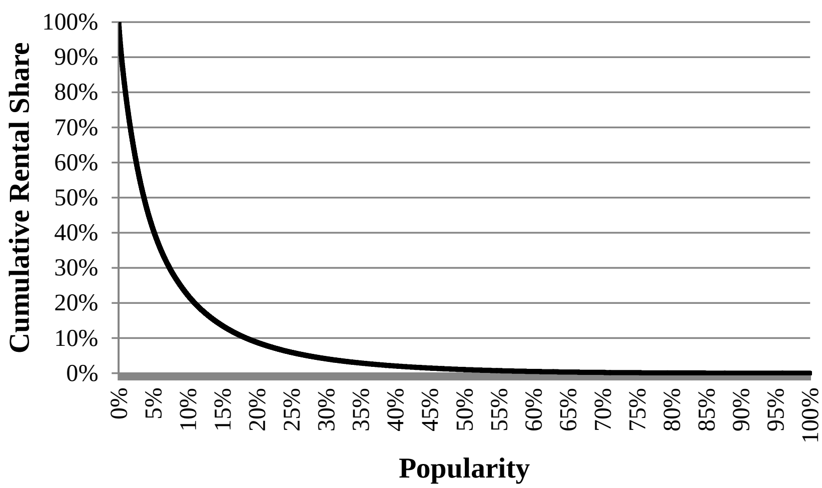


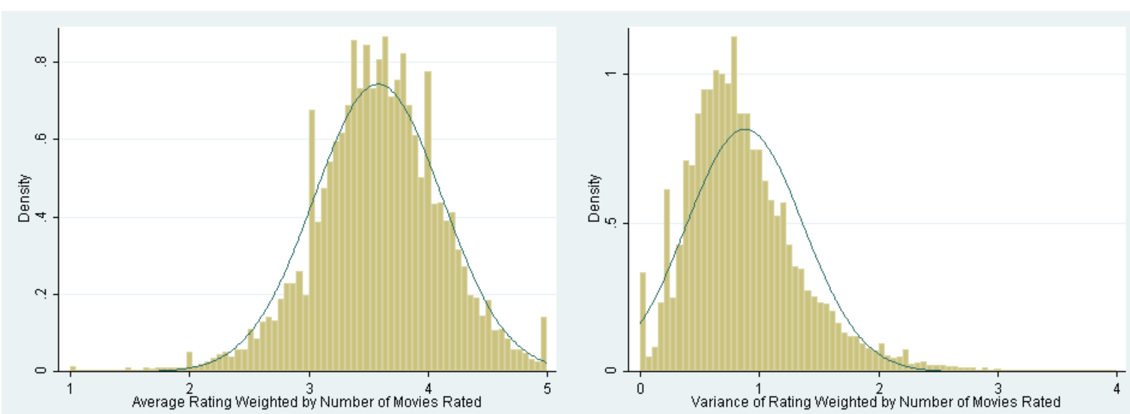
Table 4 shows the summary statistics of Dataset II at the consumer level. The number of movies rated per person every month is highly skewed toward the high percentiles, indicating that a small group of people rate a large number of movies each month. It is possible that the large number of movies rated, such as 41 for the 90th percentile and 219 for the 99th percentile, contain a large number of ratings given by users to train the recommendation system because a user can only watch a limited number of movies every month. Ratings submitted during the training process can result in a "contamination" of the data because the ratings of previously watched movies may not reflect the current popularity of a movie. In order to alleviate this issue and provide a robustness check, we tried purging data with a monthly number of rated movies over 30. We choose the cutoff point of 30 because watching 30 movies a month is probably the maximum number of movies that a heavy user is technically allowed to watch within Netflix rental system. Our results remain qualitatively and quantitatively similar, so the results reported in the paper do not drop ratings.

Furthermore, it does not appear that heavy users have a tendency to give higher or lower ratings because the frequency of ratings very weakly correlates with the average rating (correlation = 0.0038) and with the variance of ratings (correlation = 0.1027). Figure 3 further shows that consumer ratings are almost normally distributed except that the right tail is censored at the rating of 5 because of the limit of the rating scale. This nearly normal distribution of consumer ratings provides statistical evidence that the users at Netflix may be unbiased toward rating the movies about which they feel strongly. Furthermore, from the average variance of ratings (0.7) and the mean average rating (3.57), we compute that the coefficient of variation is approximately 0.23, suggesting that consumers tend to be stable in their ratings.

Table 4: Monthly Consumer Breakdown (N=4,041,165)

	Mean	Std	Skewness	1%	10%	25%	50%	75%	90%	99%
Number of Movies Rated	19	49	11	1	1	3	7	16	41	219
Average Rating	3.57	0.72	-0.42	1.33	2.75	3.12	3.6	4	4.5	5
Variance of Ratings	0.70	0.64	1.42	0	0	.22	.6	1	1.55	2.88

Figure 3: Monthly Rating Distribution



4 Estimation and Results

We test our hypothesis about the impact of product variety on demand concentration at two levels. First at the movie level, we estimate regression models using both absolute and relative measures to examine the shift of demand concentration. As robustness checks, we first analyze the impact of product variety on the Gini coefficients. Then we examine the demand impact of product variety across monthly deciles of movies. Finally, we use an instrumental variable approach to address the potential omitted variable bias. In order to gain insights into the movie-level analysis, we then turn our attention to the impact of product variety at the consumer level. We employ fixed-effect models to understand how the propensity of each consumer to discover niche movies evolves and how the composition of his/her "basket" is affected by product variety.

4.1 Movie-Level Analysis

4.1.1 Absolute and Relative Measures of Hit and Niche Movies

We use both absolute and relative measures of popularity to examine changing demand concentration using Dataset I. We specify the models as follows:

$$\log(\sum_{j \in k} Share_{jt}) = \alpha_0 + \alpha_1 \, Variety_t/1000 + \varepsilon_{kt} \quad \forall k \in \text{absolute cutoff points} \tag{1}$$

$$\log(\sum_{j \in l} Share_{jt}) = \beta_0 + \beta_1 \, Variety_t/1000 + \xi_{lt} \quad \forall l \in \text{relative cutoff points} \tag{2}$$

In these models, we divide $Variety_t$ by 1,000 and logarithmically transformed $Share_{jt}$ for interpretation purposes. We also use Huber-White estimation to correct standard errors.

Table 5 shows the results of Models 1 and 2. We observe that product variety is negatively associated with the demand for the top 10, the top 100, and the top 1,000 movies, while it is positively associated with the demand for the top 0.1%, the top 1%, and the top 10% of the movies. All coefficients are highly significant and these parsimonious regression models have significant exploratory powers. These results imply that, as product variety increases, the demand for hits decreases in absolute terms, but increases in relative terms, which partially supports our hypothesis. In particular, the demand for the top 0.1% of movies increases approximately six times as fast as the demand for the top 10% of movies (coefficients 0.0304 and 0.0054, correspondingly), indicating that with product variety increasing the demand for the "hits of the hits" significantly outpaces the demand for less popular movies when popularity is measured in relative terms.

Further, as product variety decreases, so does demand for niches movies, whether popularity is measured in absolute or relative terms. In particular, with product variety increasing, the bottom 0.1% movies lose demand faster than the bottom 1% and the bottom 10% of movies (coefficients -0.1324, -0.1229 and -0.0319, correspondingly), indicating that the demand for the "niches of the niches" deteriorates faster than for more popular movies. These results support our hypothesis: no matter how popularity is measured, product variety is negatively associated with demand for niche movies. Moreover, note that the magnitudes of *Variety* coefficients for niche movies in absolute measures are systematically larger than those for hit movies. For example, the *Variety* coefficient for the bottom 1,000 is over ten times as large as that for the top 1,000, suggesting that as product variety increases, demand for niche movies decreases even faster than demand for hits even in absolute measures.

Table 5: Regression of Hit and Niche Movie Market Shares

	Top 10	Top 100	Top 1,000	Top 0.1%	Top 1%	Top 10%
Variety	-0.0622***	-0.0397***	-0.0192***	0.0304***	0.0140***	0.0054***
	(0.0035)	(0.0018)	(0.0005)	(0.0022)	(0.0014)	(0.0004)
Constant	-2.2271***	-0.7497***	-0.0644***	-3.4028***	-1.3883***	-0.3581***
	(0.0440)	(0.0219)	(0.0065)	(0.0278)	(0.0173)	(0.0053)
Hypothesis	Not supported	Not supported	Not supported	Supported	Supported	Supported
R^2	0.853	0.906	0.962	0.776	0.656	0.756

	Bottom 10	Bottom 100	Bottom 1,000	Bottom 0.1%	Bottom 1%	Bottom 10%
Variety	-0.2328***	-0.2480***	-0.2004***	-0.1324***	-0.1229***	-0.0319***
	(0.0226)	(0.0043)	(0.0044)	(0.0218)	(0.0048)	(0.0023)
Constant	-11.1396***	-7.5805***	-4.7857***	-12.3709***	-8.9307***	-6.7860***
	(0.2802)	(0.0530)	(0.0548)	(0.2709)	(0.0598)	(0.0284)
Hypothesis	Supported	Supported	Supported	Supported	Supported	Supported
R^2	0.668	0.985	0.975	0.410	0.925	0.786

1) *p-value<0.05. **p-value<0.01, ***p-value<0.001

2) The rows below the estimates are standard errors.

4.1.2 Robustness Checks

To provide robustness checks, we first analyze the impact of a product on the Gini coefficient of the movie demand. The Gini coefficient is often used to measure distribution dispersion, with a large

Gini coefficient suggesting a highly concentrated distribution, i.e., a large Gini coefficient implies that the demand is highly concentrated on the most popular movies. We employ the following model:

$$Gini_t = \theta_0 + \theta_1\, Variety_t/1000 + \nu_t. \tag{3}$$

We find that product variety is positively associated with the Gini coefficient (coefficient is significant and equal to 0.0019). This result, once again, goes against the notion of the Long Tail effect that product variety increases the demand for niche movies.

Furthermore, we divide the movies into deciles for each month and examine the impact of product variety on these deciles. We use the following model,

$$\log(\sum_{j \in d} Share_{dt}) = \gamma_0 + \gamma_1\, Variety_t/1000 + \tau_{dt} \quad d \in \text{deciles},$$

to show the impact of product variety on movie demand across the entire distribution.

Table 6 presents the results of the demand regressions across monthly deciles. As can be seen, only the *Variety* coefficient for the top decile is significantly positive (0.0054), while the coefficients for the rest of the deciles are either statistically insignificant or significantly negative. Supporting the results in Table 5, these results further suggest that as product variety increases, the demand for the most popular movies increases, whereas the demand for less popular movies either remains stable or decreases.

Table 6: Regression of Demand Across Deciles

	1st Decile	2nd Decile	3rd Decile	4th Decile	5th Decile
Variety	0.0054***	-0.0144***	-0.0229***	-0.0209***	-0.0102***
	(0.0004)	(0.0012)	(0.0015)	(0.0017)	(0.0017)
Constant	-0.3581***	-1.9439***	-2.5701***	-3.1801***	-3.8957***
	(0.0053)	(0.0148)	(0.0187)	(0.0208)	(0.0216)
R^2	0.756	0.732	0.813	0.745	0.394

	6th Decile	7th Decile	8th Decile	9th Decile	10th Decile
Variety	0.0014	0.0032	-0.0001	-0.0091***	-0.0319***
	(0.0020)	(0.0018)	(0.0017)	(0.0018)	(0.0023)
Constant	-4.6189***	-5.1741***	-5.6556***	-6.1561***	-6.7860***
	(0.0247)	(0.0226)	(0.0214)	(0.0222)	(0.0284)
R^2	0.009	0.054	0.000	0.327	0.786

1) *p-value<0.05. **p-value<0.01, ***p-value<0.001

2) The rows below the estimates are standard errors.

So far, our econometric model is limited to capturing the correlation between product variety and demand. Many other unobserved demand factors, such as intrinsic movie quality, may bias our estimation. In order to address this omitted variable bias issue, we adopt an instrument variable 2SLS estimation approach (Angrist and Krueger, 1994). We propose using the monthly variety of all the DVDs in the market as a candidate for a valid instrument variable. In particular, we collect this market-wide DVD variety from Hometheaterinfo.com, which claims to include over 99.95% of all the DVD titles having a Universal Product Code. This market-wide variety is highly correlated with the product variety at the company that provided our movie rental data (the coefficient is about 97%), thus satisfying the relevance condition of a valid instrument. In addition, we expect market-wide variety to be exogenous to the unobserved demand factors for the movies provided by our company because some movies that were available in the market were unavailable at our company. In other words, the market-wide variety should affect demand only through the product variety provided by our rental company, so it should satisfy the exclusion restriction assumption of a valid instrument.

Table 7 shows the instrument variable 2SLS estimation results. As is clear from the tables, even after considering the additional variables, the results of our analysis are very consistent with the results in Table 5. Namely, *Variety* coefficients for the top 10, top 100, and top 1,000 movies are -0.064, -0.0405 and -0.0196, suggesting that the demand for hits decreases with product variety increasing in absolute terms. However, the same coefficients for the top 0.1%, top 1%, and top 10% of movies are 0.0302, 0.0138 and 0.0055, implying that the demand for hits tends to increase with product variety in relative terms. For the niche movies, we see that demand decreases in both absolute and relative terms as product variety increases, which supports our hypothesis and goes against the Long Tail prediction.

Table 7: Instrument Variable 2SLS Estimation of the Demand for Hit and Niche Movies

	Top 10	Top 100	Top 1,000	Top 0.1%	Top 1%	Top 10%
Variety	-0.0640***	-0.0405***	-0.0196***	0.0302***	0.0138***	0.0055***
	(0.0036)	(0.0018)	(0.0005)	(0.0022)	(0.0014)	(0.0004)
Constant	-2.2082***	-0.7409***	-0.0596***	-3.4011***	-1.3861***	-0.3582***
	(0.0441)	(0.0219)	(0.0065)	(0.0277)	(0.0173)	(0.0053)
Hypothesis	Not supported	Not supported	Not supported	Supported	Supported	Supported
R^2	0.852	0.905	0.962	0.776	0.656	0.756
	Bottom 10	Bottom 100	Bottom 1,000	Bottom 0.1%	Bottom 1%	Bottom 10%
Variety	-0.2471***	-0.2433***	-0.2023***	-0.1465***	-0.1179***	-0.0301***
	(0.0227)	(0.0043)	(0.0044)	(0.0220)	(0.0049)	(0.0023)
Constant	-10.9894***	-7.6295***	-4.7657***	-12.2220***	-8.9825***	-6.8048***
	(0.2807)	(0.0535)	(0.0548)	(0.2714)	(0.0603)	(0.0285)
Hypothesis	Supported	Supported	Supported	Supported	Supported	Supported
R^2	0.665	0.984	0.975	0.405	0.923	0.783

1) *p-value<0.05. **p-value<0.01, ***p-value<0.001

2) The rows below the estimates are standard errors.

4.2 Consumer-Level Analysis

We turn to the Netflix Dataset II to understand how product variety changes individual consumers' preferences for movies in order to gain insights into the movie-level analysis. Admittedly, ratings are not actual rental demand, but the rating data can provide insights into what movie consumers are aware of and are interested in (Chintagunta et al., 2010), serving as a proxy for movie appeal. Moreover, we replicate the movie-level rental analysis on the Netflix data and find consistent qualitative results. Additionally, rating data are useful in understanding whether niche movies satisfy consumers better than hits, which is one of the premises of the Long Tail effect.

We now proceed to examine how the propensity of each consumer to discover niche movies evolves, while controlling for observed user heterogeneity, such as rating frequency and variance. The data that we have lack other potentially significant consumer characteristics, such as demographics. In order to cope with this issue, we introduce a time-invariant preference for each consumer's movies through the panel data analysis and we further assume that preference correlates with the observed characteristics of the consumer. This correlation is likely to be caused by the recommendation systems, which can influence an individual's preference based on his/her observed characteristics. The Hausman test further provides strong evidence of this correlation. Therefore, we employ the following fixed-effect regression to predict consumer propensity to rate movies:

$$\begin{aligned}\log(\mathit{ProductRank}_{it}) &= \beta_0 + \beta_1 \mathit{Variety}_t/1000 + \beta_2 \mathit{Frequency}_{it} \\ &\quad + \beta_2 \mathit{RatingAverage}_{it} + \beta_3 \mathit{RatingVariance}_{it} + \mu_i + \varepsilon_{it}. \end{aligned} \tag{4}$$

The top of Table 8 shows the results of the absolute movie rankings while the bottom presents the same results of relative movie rankings using Model 4. As is evident from the top of the table, all $\mathit{Variety}_t$ coefficients are significantly positive, suggesting that the absolute popularity rankings of the movies watched by the average consumer consistently increase except for the very popular movies. In other words, consumers tend to discover more and more niche movies over time when movies are ranked in absolute terms. In particular, the $\mathit{Variety}_t$ coefficient for the bottom 10th percentile of the movies that a person rates (i.e., the obscure titles) is 0.0519, which is approximately 30 times as much as the same coefficient for the top 10th percentile of the movies (i.e., the popular titles). This comparison suggests that consumers are likely to discover niche products much faster than they move away from the hits (again, if popularity is measured in absolute terms).

However, the picture completely changes when popularity is measured in relative terms. The bottom part of Table 8 shows that $\mathit{Variety}_t$ coefficients are consistently negative, suggesting that, relative to the product variety that is available at that point of time, consumers tend to be interested in more and more popular movies. In particular, the $\mathit{Variety}_t$ coefficient of the top 10th percentile of movies is -0.0919, which is over twice as much as the coefficient of the bottom 10th percentile of movies, suggesting that a consumer's attention shifts toward more popular hits faster than it shifts away from less popular niche products.

Taken together, the results of Model 4 are consistent with the movie-level analysis in Subsection 4.1. They suggest that consumers do venture into more niche movies as product variety increases, but the growth rate of product variety is substantially higher than the speed at which consumers discover niche products. We suggest that such quickly expanded product variety may cause consumers to anticipate high search costs and watch those movies that they are easily aware of. A high concentration of advertising expenditures, word-of-mouth effects and theatrical release allow those popular movies to enjoy more exposure, thus enhancing their demand at the expense of niche movies. In addition, this finding is consistent with the results of Fleder and Hosanagar (2009) that recommendation systems guide similar consumers to the same products, which does not effectively help consumers discover products at the tail of the distribution.

Table 8: Fixed-Effect Model of Niche-seeking Behavior

	Mean	Median	Top 10%	Bottom 10%
Variety	0.0519***	0.0432***	0.0017***	0.0519***
	(0.0002)	(0.0002)	(0.0003)	(0.0002)
Frequency	0.0007***	0.0009***	0.0029***	0.0009***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
RatingAverage	-0.1146***	-0.1563***	-0.2609***	-0.0737***
	(0.0010)	(0.0011)	(0.0014)	(0.0011)
RatingVariance	0.2559***	-0.0056***	-0.6601***	0.5156***
	(0.0010)	(0.0012)	(0.0015)	(0.0011)
Overall R^2	0.031	0.012	0.048	0.059

	Relative Mean	Relative Median	Relative Top 10%	Relative Bottom 10%
Variety	-0.0414***	-0.0504***	-0.0919***	-0.0417***
	(0.0002)	(0.0003)	(0.0003)	(0.0003)
Frequency	0.0006***	0.0009***	0.0028***	0.0008***
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
RatingAverage	-0.1184***	-0.1544***	-0.2590***	-0.0718***
	(0.0010)	(0.0011)	(0.0014)	(0.0011)
RatingVariance	0.2440***	-0.0064***	-0.6609***	0.5148***
	(0.0010)	(0.0012)	(0.0015)	(0.0012)
Overall R^2	0.030	0.016	0.064	0.062
Observations	4,740,731	4,740,731	4,740,731	4,740,731

1) * p-value<0.05, **p-value<0.01, ***p-value<0.001

2) Standard errors are in parentheses.

Furthermore, the consistently positive and highly significant coefficients of $Frequency_{it}$ indicate that heavier users tend to discover more niche movies. In particular, the coefficients for both absolute and relative means are about 0.0006 and for both medians are 0.0009. In other words, if an average consumer watches five more movies per month, the mean of his/her propensity for niche products is likely to increase by 0.3% on average. Accordingly, the median is likely to increase by 0.45% on average, holding other factors constant. Thus, it appears that heavy users are the ones that drive the demand for niche products, which can cause the Long Tail effect. Nevertheless, these heavy users constitute a relatively small segment of the entire population: as we demonstrated earlier, heavy users with a monthly frequency over the mean constitute less than 25% of all users. Although this small group of people tends to discover more niche movies, it does not seem to shift the entire demand from hits to niche products. A comparison of the coefficient for the top 10th percentile (0.0028) and the coefficient for the bottom 10th percentile (0.0008) suggests that heavier users shift away from the hits approximately three times as fast as they discover niche products. That is, even heavy users are not as fast in discovering niche products as they are in moving away from about hits.

The consistently negative and highly significant coefficients of $RatingAverage_{it}$ suggest that consumers who, on average, give higher ratings and may therefore be more satisfied tend to be interested in more popular movies. For example, the coefficients for both absolute and relative means are around -0.11, which suggests that increasing the average rating by one unit is associated with an 11% increase in average movie popularity. In other words, the more popular movies generally satisfy people better than do the obscure titles. Of course, it is possible that consumers who watch popular movies are systematically different from consumers who watch niche movies in that the former tend to rate all movies more highly than the latter do. Since we are unable to observe characteristics of individual consumers, our findings are subject to this limitation.

Finally, we note that $RatingVariance_{it}$ is negatively associated with the median and the top 10th percentile, while this variable is positively associated with the mean and the bottom 10th percentile (in either absolute or relative terms). We interpret these mixed signs to imply that consumers having highly disperse rating tend to watch extreme hits and niche products. In other words, the extreme hits and the extreme niche products receive more polarized ratings from those consumers. Further, the popularity of the movies that those consumers watch tends to skew toward niche products. In other words, consumers having highly disperse ratings watch a larger quantity of hits than niche products, but the niche products that they watch are generally extremely obscure.

5 Conclusion and Discussion

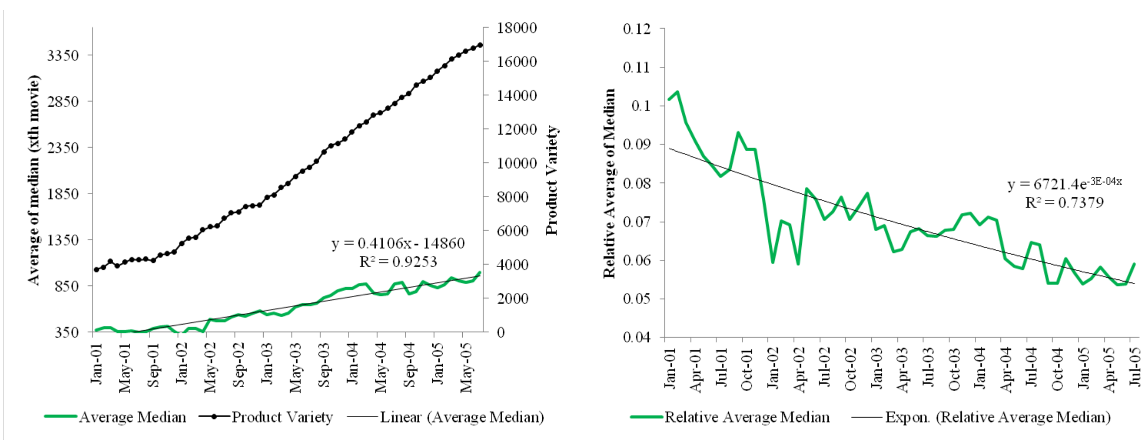
The reasoning behind the Long Tail effect presumes that expanding product variety due to the adoption of the Internet will satisfy consumers' increasingly heterogeneous tastes, thus causing the demand for niche products to rise. In this paper we empirically examine the impact of product variety on demand concentration. We argue that one has to be careful about defining hits and niche products in the Internet era. In a brick-and-mortar world, where product variety is relatively stable and all products are consumed at some rate, hits and niche products are typically defined in absolute terms (e.g., the top 100, the bottom 100 movies). However, product variety has been skyrocketing in the Internet age and therefore more and more products can be left unnoticed by consumers, or are being discovered very slowly, even though the customer base is also expanding. To evaluate the consumer propensity to discover niche products and to separate this effect from the entirely different effect of increasing product variety on the Internet, we suggest that product popularity should be measured in relative terms, thus dynamically adjusting for the "active" product variety at that point of time. By doing this, we bring the distribution of demand to a common scale.

We use two large data sets from the movie rental industry and we analyze the data at both the movie level and consumer level. In these large data sets we find that, when the popularity of a movie is defined dynamically, i.e., in relative terms, product variety is positively associated with the demand for hits, but negatively associated with the demand for niche products. Additionally, even in the absolute ranking definition, the Long Tail effect is only partially present: product variety diversifies the demand for both hits and niche products, although this is less significant for hits than for niche products. This evidence is mostly consistent with our hypothesis. We further find that product variety is positively associated with monthly Gini Coefficients, a measure of demand concentration.

In order to gain insights into these movie-level findings, we further examine changes in the demand distribution at the consumer level. Once again, we find that product variety indeed diversified consumers into more niche movies in absolute terms, but we also discover that the rate at which consumers shift demand from the hits to the niche products is considerably lower than the growth rate of product variety. Therefore, if we normalize demand for currently active product variety and measure popularity in relative terms, we find that consumers tend to watch more and more hits as product variety grows. In other words, expanded product variety may threaten the attraction of one popular movie because of demand diversification, but it may favor more and more popular movies.

Figure 4 visually illustrates the comparison between the absolute and relative popularity of the movies that consumers watch. In Figure 4 (left, bottom line), we observe the ranking of a median movie that the average consumer discovers over time, which has a linear upward trend, indicating that consumers increasingly discover niche products. In particular, in 2001, the median movie rated by an average consumer was ranked slightly above 350, while in 2005 the median movie ranking had increased more than twice to over 850. However, Figure 4 (left, top line) also indicates that product variety increased even more quickly, which creates an impression of the lengthening tail of demand distribution: there are more and more obscure movies over time. However, once we bring distribution to the common scale by dividing by current product variety, the claim of increasing demand for niche products disappears. Not surprisingly, Figure 4 (right) shows that, when we look at median popularity in relative terms, the average consumer gravitates more and more toward hits. In fact, in 2001 the average consumer was interested in the movies in the 10th percentile of product variety while in 2005 the average consumer was interested in the movies in the 6th percentile. Hence, we conclude that although consumers do venture into niche products, new movies appear more quickly than people can actually discover them.

Figure 4: Average Median and Relative Average Median Popularity



We make a number of additional observations based on our consumer-level analysis. We find evidence that the consumers who give high average ratings tend to watch more popular movies. Hence, we do not find any evidence that niche products satisfy consumer tastes better and better over time, which is suggested by Anderson (2004; 2006). Furthermore, we find that the consumers who discover niche products tend to be heavy users, constituting only a small portion of the entire user base. Light users, on the other hand, tend to focus on the popular items and since most users are in this category, hits continue to drive the market.

Our findings have a number of managerial implications as they shed new light on the controversy surrounding the Long Tail effect. First, the promise of the Long Tail effect became a basis for many new business models and business ideas (Anderson, 2006). Our findings suggest that caution needs to be used when assessing the potential benefits of focusing a business on supplying niche products. While it may be true that niche products are much more profitable for companies (e.g., Anderson 2006 rightfully suggests that niche movies cost a fraction of hit movies to make), this argument does not account for the fact that for each niche product that consumers demand, there might be several that are never discovered, thus potentially adding to the costs but not to the revenues. Irrational expansion into niche products will also increase operational difficulties, such as maintaining the level of service (Fisher et al., 1994; Randall and Ulrich, 2001). In fact, to compete against Netflix-like companies that stock a large product variety of niche movies, companies like Redbox successfully remain profitable by focusing only on a selected number of hit movies and capturing 34.5% of the rental market share in 2011 (NPD, 2011). In addition, Amazon.com, which is often cited as an example of offering numerous long tail products on its platform, is found to directly sell only a small percentage of all products listed on its website, with most products being sold by third-party sellers because of insufficient demand for those niche products (Jiang et al., 2011).

Further, a large number of products might take a while to be discovered. This finding seems to suggest that much more attention needs to be paid to recommendation systems, review forums and other means of aiding product discovery. Although Netflix employs what is widely considered to be a sophisticated recommendation system, even this system does not allow numerous consumers to discover titles as fast as they appear. This raises an important issue of carefully forecasting how long it will take for a given title, once it is added to the inventory, to begin accumulating demand. More improvements to the recommendation systems, such as through the Netflix Prize and the algorithm proposed by Park and Tuzhilin (2008) should be implemented.

Insights from our consumer-level analysis suggest that consumers are generally much more satisfied by hit products than by niche products. This is an important consideration: while Netflix currently achieves extremely high customer satisfaction, we do not find any evidence to suggest that customers watching obscure titles find them more satisfactory than other movies. We can speculate that many consumers over time will learn that niche products are called niche products for a reason and might start ignoring them altogether. Our other observation that heavy users tend to venture into more obscure movies suggests that the presence of the Long Tail effect might be moderated by the frequency of service. In the case of Netflix, it is physically impossible to rent more than a few DVDs per month (due to mail processing time). However, Netflix and other companies, such as Amazon.com and Hulu.com, have started allowing customers to stream content and watch movies on their computers at home right away, which may increase the number of heavy users who discover niche products. In this case, one will have to re-examine the demand concentration.

It is important to remember the limitations of our findings. First, our study does not directly compare the search costs between brick-and-mortar and Internet companies (e.g., Brynjolfsson et al. 2011) and therefore we are unable to comment on this aspect of the Long Tail effect. Rather, our findings need to be interpreted as a study about the impact of product variety on demand concentration only. Comparing the effects of product variety across channels would be of considerable interest for research. Second, our study has focused on the movie rental industry. It is possible and worthwhile to confirm that demand concentration in other product categories may respond differently to varying product variety levels. Third, our consumer-level analysis is restricted to ratings data. To address this issue, we confirm that the main results of rental data are consistent with those of the ratings data and we use consumer-level rating data to offer additional insights into the movie-level analysis. Admittedly, the consumer-level results cannot be taken as exact evidence for individual rental behavior on Netflix.com, since some consumers probably do not rate the movies that they watched. Nevertheless, consumers also rate movies that they watched elsewhere, providing a richer picture of demand for movies which reflects interest, attention and satisfaction. An interesting venue for research, particularly for behavioral economics, would be to compare ratings data with time-stamped individual-level rental data to understand possible behavior biases. For example, Milkman et al. (2009) analyzed online video rental data to study the present bias between *should* movies and *want* movies. In addition, a data set with time-stamped individual-level information may be potentially used to study consumer purchase patterns and their life-time value.

Further research opportunities also include linking recommendation system metrics, such as product ratings, with operations management and marketing strategies (see Netessine et al. 2006 for some initial work in this direction). Finally, incorporating the empirical findings of the product variety effects on demand concentration and evolving consumer preferences into the analytical models, such as dynamic assortment (Caro and Gallien, 2007), is highly warranted.

References

Anderson, C. 2004. The long tail. *Wired Magazine* **12**(10) 170–177.

Anderson, C. 2006. *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.

Angrist, J., A.B. Krueger. 1994. Why do World War II veterans earn more than nonveterans? *Journal of Labor Economics* **12**(1) 74–97.

Bertini, M., L. Wathieu, S.S. Iyengar. 2012. The discriminating consumer: Product proliferation and willingness to pay for quality. *Journal of Marketing Research* **49**(1) 39–49.

Brynjolfsson, E, Y. Hu, D. Simester. 2011. Goodbye Pareto Principle, hello Long Tail: The effect of search costs on the concentration of product sales. *Management Science* **57**(8) 1373–1386.

Brynjolfsson, E., Y. Hu, M. 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, E., Y. Hu, M. D. Smith. 2006. From niches to riches: The anatomy of the Long Tail. *Sloan Management Review* **47**(4) 67–71.

Brynjolfsson, E., Y. Hu, M.D. Smith. 2010. Research commentary—long tails vs. superstars: The effect of information technology on product variety and sales concentration patterns. *Information Systems Research* **21**(4) 736–747.

Cachon, G. P., C. Terwiesch, Y. Xu. 2008. On the effects of consumer search and firm entry in a multiproduct competitive market. *Marketing Science* **27**(3) 461–473.

Caro, F., J. Gallien. 2007. Dynamic assortment with demand learning for seasonal consumer goods. *Management Science* **53**(2) 276.

Chatfield, C., G. Goodhardt. 1975. Results concerning brand choice. *Journal of Marketing Research* **12**(1) 110–113.

Chen, P.Y., S. Wu, J. Yoon. 2004. The impact of online recommendations and consumer feedback on sales. International Conference on Information Systems (ICIS).

Chintagunta, P.K., S. Gopinath, S. Venkataraman. 2010. The effects of online user reviews on movie box-office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science* **29**(5) 944–957.

Clemons, E. K., G. Gao, L. M. Hitt. 2006. When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems* **23**(2) 149–171.

Dellarocas, C., G. Gao, R. Narayan. 2010. Are consumers more likely to contribute online reviews for hit or niche products? *Journal of Management Information Systems* **27**(2) 127–158.

Dellarocas, C., C.A. Wood. 2008. The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias. *Management Science* **54**(3) 460–476.

Elberse, A., F. Oberholzer-Gee. 2008. Superstars and underdogs: An examination of the Long Tail phenomenon in video sales. HBS Working Paper.

Fisher, M.L., J.H. Hammond, W.R. Obermeyer, A. Raman. 1994. Making supply meet demand in an uncertain world. *Harvard Business Review* (May-June) 83–93.

Fleder, D., K. Hosanagar. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science* **55**(5) 697–712.

Fleder, D., K. Hosanagar, A. Buja. 2009. Recommender systems and their effects on consumers: The fragmentation debate. Net Institute Working Paper.

Ghose, A., B. Gu. 2006. Search costs, demand structure and long tail in electronic markets: Theory and evidence. NET Institute Working Paper.

Goodhardt, G. J., A. S. C. Ehrenberg, C. Chatfield. 1984. The Dirichlet: A comprehensive model of buying behaviour. *Journal of the Royal Statistical Society. Series A (General)* **147**(5) 621–655.

Hervas-Drane, A. 2009. Word of mouth and taste matching: A theory of the Long Tail. Working Paper.

Hinz, O., J. Eckert, B. Skiera. 2011. Drivers of the Long Tail phenomenon: An empirical analysis. *Journal of Management Information Systems* **27**(4) 43–70.

Hu, N., J. Zhang, P.A. Pavlou. 2009. Overcoming the j-shaped distribution of product reviews. *Communications of the ACM* **52**(10) 144–147.

Jiang, B., K. Jerath, K. Srinivasan. 2011. Firm strategies in the "Mid Tail"? of platform-based retailing. *Marketing Science* **30**(5) 757–775.

Kuksov, D., J.M. Villas-Boas. 2010. When more alternatives lead to less choice. *Marketing Science* **29**(3) 507–524.

Kumar, A., M.D. Smith, R. Telang. 2011. Information discovery and the Long Tail of motion picture content. Working Paper.

Lambert, P.J., J.R. Aronson. 1993. Inequality decomposition analysis and the Gini coefficient revisited. *The Economic Journal* 1221–1227.

McPhee, W. N. 1963. *Formal Theories of Mass Behavior*. Glencoe, NY: Free Press.

Milkman, K.L., T. Rogers, M.H. Bazerman. 2009. Highbrow films gather dust: Time-inconsistent preferences and online DVD rentals. *Management Science* **55**(6) 1047–1059.

Netessine, S., S. Savin, W. Xiao. 2006. Revenue management through dynamic cross selling in e-commerce retailing. *Operations Research* **54**(5) 893–913.

NPD. 2011. Coinstar inc. q2 2011 earnings presentation. Tech. rep., NPD Group.

Park, Y.J., A. Tuzhilin. 2008. The long tail of recommender systems and how to leverage it. ACM Conference on Recommender Systems.

Randall, T., K. Ulrich. 2001. Product variety, supply chain structure, and firm performance: An empirical examination of the U.S. bicycle industry. *Management Science* **47**(12) 1588–1604.

Rosen, S. 1981. The economics of superstars. *The American Economic Review* **71**(5) 845–858.

Rothschild, M. 1974. Searching for the lowest price when the distribution of prices is unknown. *The Journal of Political Economy* **82**(4) 689–711.

Shih, W., S. Kaufman, D. Spinola. 2007. Netflix. *Harvard Business School Case* (November) 9–607–138.

Simonson, I., A. Tversky. 1992. Choice in context: Tradeoff contrast and extremeness aversion. *Journal of Marketing Research* **29**(3) 281–295.

Stigler, G.J. 1961. The economics of information. *The Journal of Political Economy* **69**(3) 213–225.

Tucker, C., J. Zhang. 2011. How does popularity information affect choices? A field experiment. *Management Science* **57**(5) 828–842.

Yitzhaki, S. 1979. Relative deprivation and the Gini coefficient. *The Quarterly Journal of Economics* 321–324.

Zhou, W., W. Duan. 2011. Online user reviews, product variety, and the long tail: An empirical investigation on online software downloads. *Electronic Commerce Research and Applications* **11** 275–289.

Europe Campus
Boulevard de Constance
77305 Fontainebleau Cedex, France
Tel: +33 (0)1 60 72 40 00
Fax: +33 (0)1 60 74 55 00/01

Asia Campus
1 Ayer Rajah Avenue, Singapore 138676
Tel: +65 67 99 53 88
Fax: +65 67 99 53 99

Abu Dhabi Campus
Muroor Road - Street No 4
P.O. Box 48049
Abu Dhabi, United Arab Emirates
Tel: +971 2 651 5200
Fax: +971 2 443 9461

www.insead.edu

INSEAD

The Business School
for the World®

Printed by INSEAD