2017

Is Tom Cruise Threatened? An Empirical Study of the Impact of Product Variety on Demand Concentration

Tom Tan
Southern Methodist University, ttan@cox.smu.edu

Serguei Netessine
INSEAD, Singapore, serguei.netessine@insead.edu

Lorin Hitt
University of Pennsylvania, lhitt@wharton.upenn.edu

Is Tom Cruise Threatened? An Empirical Study of the Impact of Product Variety on Demand Concentration

Tom Fangyun Tan

Cox Business School, Southern Methodist University, Dallas, Texas, U.S.A.

ttan@cox.smu.edu

Serguei Netessine

INSEAD, Singapore

serguei.netessine@insead.edu

Lorin Hitt

The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania, U.S.A.

lhitt@wharton.upenn.edu

Abstract

We empirically examine the impact of expanded product variety on demand concentration using large data sets from the movie rental industry as our test bed. We find that product variety is likely to increase demand concentration, which goes against the "Long Tail effect" theory predicting that demand would become less concentrated on "hit" products due to expanded product variety. We further provide evidence that this finding is not due to introducing many low-selling niche products as the intuition might suggest. Instead, we discover that increasing product variety diversifies the demand away from each movie title, but less significantly for hits than for niche products. In particular, we find that increasing product variety by 1,000 titles may increase the Gini coefficient of DVD rentals by 0.0029, which translates to increasing the market share of the top 1% of DVDs by 1.96% and the market share of the top 10% of DVDs by 0.58%. At the same time the market share of the bottom 1% of DVDs is reduced by 21.29% while the market share of the bottom 10% of DVDs is reduced by 5.28%. We rule out alternative explanations using a variety of "Long Tail" metrics, capturing movie format/distribution channel interaction and customer heterogeneity, while making use of instrumental variables.

Keywords: product variety; demand concentration; movie rental; the Long Tail effect; product rating.

Acknowledgment: We sincerely acknowledge the entire review team for their most diligent and constructive comments. In particular, we are thankful to the reviewers for encouraging us to consider various alternative explanations. Furthermore, we are grateful to the faculty members (especially Katherine Milkman, Bin Gu, *Kartik Hosanagar, and Peter Fader) and the PhD students at the Wharton School, INFORMS and ICIS Annual Conference participants for their valuable suggestions. Finally, we wish to thank the INSEAD-Wharton Alliance and Mack Institute for Innovation Management for their financial support throughout this project.*

1 Introduction and Related Literature

Chris Anderson, former editor-in-chief of *Wired Magazine*, coined the term "Long Tail effect" (Anderson, 2004) predicting that, due to the adoption of information technologies, obscure or "niche" products would comprise increasing market share, while the demand for popular products, such as Tom Cruise's "hit" movies, would continue to decrease, so that demand would become less concentrated over time. The reason is that niche products would continue to better satisfy consumer preferences because consumers would continue to have more and more varying preferences, and the expanded product variety due to advances in information technology would make even the most obscure products available to the masses. 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, for enhancing online recommendation systems, and for supply chain management of these products on the Internet (Brynjolfsson et al., 2010; Jiang et al., 2011; Xu et al., 2012; Gallino et al., 2015). The Long Tail effect has generated widespread interest in academic circles. Brynjolfsson et al. (2010) provide a timely review of the research on the Long Tail effect, where they categorize the plausible drivers of the Long Tail effect into demand-side and supply-side drivers. In particular, the demand-side drivers mainly include search and database technologies, personalization technologies and online communities and social networks, while the supply-side drivers suggest that lowered production and stocking costs in the IT-enabled markets allow more types of products to be available to satisfy consumers' demand.

Subsequent academic papers tended to focus on the impact of lower search costs, especially those enabled by new information technology, on demand concentration. 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. Brynjolfsson et al. (2011) empirically analyze a retailer that offers the same product assortment online and offline and find that the online store exhibits less concentrated demand because of its lower search costs. Likewise, Zentner et al. (2013) conclude that the Internet channel exhibits lower demand concentration because of lower search costs. Moreover, Tucker and Zhang (2011) suggest that information about product popularity online, such as how many people browsed the product, can disproportionately increase the appeal of niche products. Dewan and Ramaprasad (2012) also find that music blogs expose consumers to a wider range of music and encourage them to sample more niche songs, although they are less willing to purchase such songs. Oestreicher-Singer and Sundararajan (2012) empirically find that the recommendation network should lead to less disperse demand. In addition, Kumar et al. (2014) find that information discovery in pay cable broadcast windows allows consumers to discover movies that they did not discover in the theaters, shifting their DVD purchases toward niche titles. All these studies seem to suggest that lowering search costs leads to higher demand for niche products.

On the other hand, several studies have questioned the premise of the Long Tail effect and provided conflicting evidence. Hervás-Drane (2013) provide an analytical model to show that different search processes have mixed impacts on demand concentration. 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. Hosanagar et al. (2014) further find that personalization tools, which are assumed to fragment consumers and therefore to diversify demand, surprisingly create commonality among the consumers.

Although whether or not lower search costs decrease demand concentration remains a hotly debated topic, very limited research has been done to empirically evaluate the other original premise of the Long Tail effect, i.e., the effect of increasing product variety on demand concentration (see Hinz et al., 2011 for an excellent literature review). Zhou and Duan (2012) use the number of downloads of a particular popularity segment to measure the Long Tail (the "Absolute Long Tail", more about this definition in Subsection 3.2) and find that product variety may decrease demand concentration in the context of online software downloading. Hinz et al. (2011) find that product variety has almost no impact on demand concentration for video-on-demand if it is measured in terms of the Gini coefficients, an often used metric in the literature on the Long Tail effect. These conflicting results of using different measures of demand concentration correspond to a critical issue of this stream of literature, that is, different measures of the Long Tail can lead to seemingly contradictory outcomes, thus causing confusion (Brynjolfsson et al., 2010).

A significant challenge of understanding the true causal effect of product variety on demand concentration lies in potential endogeneity and alternative explanations. For example, retailers may anticipate the demand for hit or niche products and thus decide on the size of their product offering, making it difficult to disentangle the direction of any causal effect. In addition, when it comes to the media industry (a favorite example in Anderson, 2004), movie rentals are available in different formats (e.g., VHS, DVD) and in different channels (e.g., online, offline). If one format or channel cannibalizes the demand for a particular popularity segment, i.e., hit or niche movies in another format/channel, the demand concentration will change, which can confound the true effect of product variety. Consumers who favor a particular popularity segment may also enter or exit a movie format/channel at different time, creating another confounding factor. None of the previous studies of the effect of product variety on demand concentration explicitly consider these endogeneity issues or alternative explanations. The goal of our paper is to identify the causal effect of product variety on demand concentration while alleviating such possible concerns.

In this paper we use large data from the movie rental industry as a test-bed to empirically evaluate the impact of product variety on the demand concentration, with particular attention to distinguishing the direction of causal effects and ruling out many of the plausible alternative explanations. Our identification strategy relies on an likely exogeneous shock to supply in the form of new agreements with Program Suppliers, which we use as an instrument and in a regression discontinuity design. Multiple models and robustness checks consistently show that higher product variety is likely to increase the demand concentration, contrary to the predictions made regarding a long-tail effect. In particular, we find that increasing product variety by 1,000 titles may increase the Gini coefficient of DVD rentals by 0.0029, which translates to increasing the market share of the top 1% of DVDs by 1.96% and the market share of the top 10% of DVDs by 0.58%. At the same time the market share of the bottom 1% of DVDs is reduced by 21.29% while the market share of the bottom 10% of DVDs is reduced by 5.28%. We further provide evidence that this main finding is not due to introducing many low-selling niche products as the intuition might suggest. Instead, it is likely to be caused by uneven demand diversification for each movie. In particular, as product variety increases, we disover that the demand for each movie title (measured by movie's market share) drops. This demand diversification turns out to be less significant for hits than for niche movies, thus increasing relative demand for all the hits and reducing demand for the niches.

2 Conceptual Framework

Since we are interested in the effect of product variety on consumers as opposed to on firms, we build our theoretical foundation primarily on consumer behavior literature. Classical theories suggest that larger product variety helps consumers meet their diverse preferences (see Lancaster, 1990 for a review). First of all, some consumers clearly know their ideal preferences and search for products that are closest to those preferences (Chernev, 2003). Therefore, a large product variety is more likely to allow consumers to find the product that matches their tastes and satisfies their heterogeneous preferences (Baumol and Ide, 1956; Lancaster, 1990; Anderson, 2004). Similar to other information goods industries, the movie industry generally has highly heterogeneous consumers enjoying different types of movies (Caves, 2000). Second, consumers often seek variety, i.e., they look for products with attributes different from their old favorites, probably out of satiation, curiosity or fluctuating needs (McAlister, 1982; Simonson, 1990; Kahn, 1995, 1998). Research also shows that variety seeking is more likely to happen in experiential attributes such as tastes than non-experiential attributes such as brand names (Inman, 2001). Offering a large product variety allows firms to follow the variety seeking inclinations of consumers. Movies are a type of experiential goods, within which the movie consumers are found to be more likely to seek variety than, say, in beer or soft drinks categories (Trivedi et al., 1994). The consumers may often seek another type of a movie to maintain an optimal level of stimulation (Raju, 1980), and therefore they should benefit from a larger variety of movies offered in the market. These two reasons seem to predict that product variety should diversify the demand, thus reducing demand concentration.

However, recent studies have highlighted some downsides of having "too much choice", which may counter the expected effect of product variety on demand diversification (Gourville and Soman, 2005). First, having many choices may induce various types of negative emotions. For example, choosing from a large choice set may demand more consumers' cognitive resources to evaluate the alternatives, causing confusion and anxiety (Lehmann, 1991; Huffman and Kahn, 1998). Second, too much variety may make the choice more difficult because the differences among the options become smaller and the amount of information about them may overload consumers (Iyengar and Lepper, 2000; Berger et al., 2007; Fasolo et al., 2009). Large product variety makes it even more difficult to evaluate experiential products like movies because their qualities are not fully revealed up front. Assessing these movies by searching Internet resources, such as Variety.com or Rottentomatoes.com takes extra time and requires additional cognitive effort. As a result, a large product variety makes an exhaustive consideration of all alternatives undesirable and infeasible from a time-and-effort perspective (Schwartz, 2004). Consumers may therefore choose to consider fewer choices and to process a smaller amount of information available regarding the choices using simpler heuristics (Hauser and Wernerfelt, 1990; Payne et al., 1993). For example, consumers may restrict their choices to the products for which they have ex-ante knowledge (Stigler, 1961; Rothschild, 1974). They may also consider only those easily justifiable choices (Sela et al., 2009) which involve utilitarian options over hedonistic ones. When renting movies, consumers may rely on some simple heuristics that may logically concentrate on well-known movies because most consumers have ex-ante knowledge about them. In addition, consumers are more likely to consider those movies that appeal to the general public in a larger product variety because those movies may function as a public topic instead of merely as a hedonistic consumption (McPhee, 1963). All of these reasons suggest that demand might concentrate more around hit movies.

The aforementioned theories seem to suggest conflicting effects of product variety on demand concentration. On one hand, a larger product variety may satisfy heterogeneous consumers' increasingly varying tastes and allow them to follow their variety seeking inclinations, thus diversifying the demand from hits to niches. On the other hand, consumers facing huge product variety may restrict their choice consideration to only the movies for which they have ex ante knowledge or those movies that can be easily justified, i.e., popular hits. Hence, whether demand concentration increases or reduces demand concentration is an important empirical question, which we rigorously examine in the following sections.

3 Data

3.1 Research Setting and Data Description

We gathered data available from a distributor (we call it the Company hereafter) that leases and delivers movies to retailers for subsequent rental to consumers. Its clients include home video specialty stores, grocery stores and convenience stores, which represent approximately 30% of the entire U.S. movie rental retailers. Note that our data represents actual rental transactions conducted by consumers. The Company implemented an innovative information system to collect the rental information for the movies because they are rented to consumers on a revenue-sharing basis with the retailers. Our data consist of the monthly aggregate DVD rental turns and movie characteristics at the movie level from January 2001 to July 2005.

We believe that our data provide rich grounds to study the impact of varying product variety on demand concentration patterns. First, this data set is one of the most representative and 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 for a relatively long time span. In particular, the U.S. DVD rental turns reached 1.75 billion turns in 2004 (Association, 2015), while in our sample the DVD rental turns were 545 million turns, approximately 31% of total market turns.

Second, our sample characteristics are comparable to the industry-level characteristics, thus providing confidence for the generalizability of our results. For example, we find that the composition of the movie genres in the DVDs released in the U.S.[1] is congruent with the composition of the genres in our sample. In addition, the DVD rental market typically exhibits seasonal peaks in early summer and Christmas because distributors tend to release hit titles during those periods. In our sample, we also observe similar seasonal peaks during the same time of the year.

Third, the revenue-sharing contract ensures the accuracy of the reported movie rental turns through considerable computer monitoring and external verification of the results. Moreover, the Company sells their movie rental information recorded in their revenue-sharing systems to their content providers, retailers and market researchers to be used as business intelligence. This business model provides further assurance that their information should be representative of the market characteristics.

Fourth, the fact that all the transactions happened at the brick-and-mortar stores controls for the similar business model and industry trend. Between 2001 and 2005 brick-and-mortar movie rental retailers dominated the home video rental industry, representing the majority of consumers' preferences. This particular industry background during our study period alleviates the concern that those consumers who self-selected into brick-and-mortar DVD rental market may have been systematically different from those consumers in other distribution channels. Admittedly, although online streaming or mobile streaming were unavailable until after the end of our study period, online movie rental companies like Netflix grew between 2001 and 2005. In addition, online DVD rentals and video cassettes (VHS) rentals may have interactive effects with offline DVD rentals, creating possible alternative explanations to our results. For these reasons, we introduce additional data of VHS rentals and consumer-level online DVD rentals in Subsection 4.3 to alleviate such concerns. Furthermore, our data are at the movie-level, which allows for the potential alternative explanations of the entry and exit of consumers having heterogeneous tastes together with the changes of product variety. To address this issue, in Subsection 4.3, we conduct a consumer-level analysis of a balanced cohort and we find robust results.

Table 1 presents the descriptive statistics of the rentals by year. The active product variety, which is the number of distinct DVDs rented by consumers at least once, substantially increased from 7,246 in 2001 to

[1]The market level information comes from Hometheaterinfo.com, which claims to include over 99.95% of all the DVD titles having a Universal Product Code.

25,488 in 2005, up approximately three 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. The skewness of the turns increased from 5.37 in 2001 to 6.94 in 2005, suggesting that the most popular titles are likely to constitute 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.

Table 1: Descriptive Statistics of Movie Rentals

Year	Product Variety	Rental Turns (in MN)	Skewness of Turns	Min Turns	Median Turns	Max Turns	Newly Released Titles	Newly Rented Back Catalog
2001	7,246	162	5.37	23	895	728,526	1,639	5,607
2002	10,975	245	5.40	1	845	933,998	2,468	1,762
2003	15,681	369	5.62	1	1,054	1,122,852	2,994	2,214
2004	23,255	546	6.02	1	2,461	1,019,122	3,953	3,270
2005*	25,488	352	6.94	1	1,555	682,397	2,354	1,064
Mean/year	16,529	335	5.87	5	1,362	897,379	2,681	2783
Stdev/year	7,798	145	0.65	10	676	188,235	859	1770

* We only observe seven months in 2005.

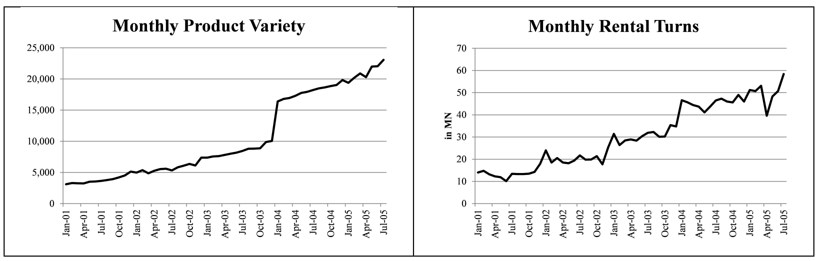
Figure 1 shows that the monthly product variety increased quickly from January 2001 to July 2005, and that the rental turns increased linearly during the same period[2]. The product variety expanded because more and more DVDs were converted from VHS during that period and because the Company steadily lowered the ordering costs for the retailers in exchange for their commitment to order more titles from suppliers. Note that product variety increased sharply in 2004. According to the Company's 10K, in 2004 it implemented new agreements with a major new supplier to increase the available titles of DVDs. This exogenous jump in product variety is an important factor that we will use to identify causal effects. In addition, to assure that our results are not mechanistically caused by this jump in product variety in 2004, we randomly selected two samples that respectively have one third and two thirds of the original product variety levels from the rental data. We use these two random samples separately to repeat our main analysis. The results remain qualitatively the same.

A relevant question is whether product variety is growing because many brand new movies are being

[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 1 is greater than or equal to the monthly product variety in the same year.

released or because consumers keep discovering previously released titles. Table 1 indicates that the number of brand new titles increased from 1,639 in 2001 to 3,953 in 2004, while the newly rented back catalog titles decreased from 5,607 in 2001 to 3,270 in 2004. Although the number of newly rented back catalog rebounded in 2004, it was mainly due to the aforementioned new agreements with a major new supplier, which tended to introduce a significant number of its back catalog products upon signing. Hence, 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.

Figure 1: Monthly Product Variety and Rentals



3.2 Measures and Controls

Ideally, we would like to adhere to the same weekly-level analysis as performed in some of prior works (e.g., Hinz et al., 2011). Unfortunately, we are limited to working with monthly data because our raw movie rental data were collected on a monthly basis. Nevertheless, our data contain larger product variety, and cover not only many more retailers and consumers (30% of the entire U.S. market) but also significantly longer study periods than prior works. In addition, by aggregating our analysis on the monthly basis, we ensure both an adequate sample size in each month for each movie and enough observations over time for statistically significant estimates. In addition, our monthly-level analysis may provide a conservative estimate of the effect of product variety on demand concentration because longer time intervals tend to smooth out the variations and create a more stable pattern.

First, we define $Variety_t$ as the total number of movies (in 1,000 titles) that were rented at least once 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 at all in a given month. In a related study, Brynjolfsson et al. (2011) include only the products available in the most recent clothing catalog in their definition of product variety, although consumers can still purchase from other items from back issues of catalogs. According to the authors, this definition ensures that any evidence that there is a longer tail of products on the Internet than in the catalog channel cannot be explained by some items not appearing in the catalog. Similarly, in our setting, those movies that consumers did not rent should not be taken into account when ranking popularity. An additional consideration is that movies that have no rentals should be considered "niche" movies by definition. Including those movies may mechanistically deflate the demand for niches. Hence, our definition of product variety as active variety, if anything, should be upwardly biasing the effect of product variety on the demand for niches and therefore make our results more conservative. Under this definition $Variety_t$, it is true that a particular movie which is rented in one month but not in another will be included in the product variety of only one month even though that movie is available during both months. We do not believe that this will cause significant bias in our findings because the demand for any particular movie tends to decay quickly over time and therefore we should separately treat the markets in those two months. Nevertheless, we run a robustness check by counting the number of distinct movies during the current month and the two previous months as the product variety for the current month. The results of this alternative product variety definition are still consistent.

We are interested in studying demand concentration, i.e., how demand is distributed among popular and niche movies. Brynjolfsson et al. (2010) suggested three measures of the demand concentration – the "Absolute Long Tail", the "Relative Long Tail" and the exponent of Power-law distribution. We do not adopt the "Absolute Long Tail" which measures the changes in terms of the absolute number of rentals. As argued by Brynjolfsson et al. (2010), this definition of the Long Tail "is not always intuitive to apply across different markets" or time periods because it is scale variant. As shown in Table 1, total market demand dramatically increased from 162 million turns in 2001 to 546 million turns in 2004 (up close to two and half times). In other words, the significance of an absolute number of rentals, e.g., 1,000 turns, in 2001 should be greater than the same number of rentals in 2004. Nevertheless, we conduct a robustness check using this measure and the results are qualitatively consistent with our main results.

We focus on the "Relative Long Tail" definition which measures the relative share of demand above or below a certain rank. This definition is appealing to apply in our setting because it is scale invariant, allowing us to adjust for possible changes in the consumer base and total market demand over time. In particular, we 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 uniform distribution of demand during month *t*, while a value of one suggests maximal inequality with all demand allocated to one product (Brynjolfsson et al. (2011)). As a secondary measure, which is similar to the classic Pareto Principle that 20% of the products often generate 80% of the sales, we calculate 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*. Accordingly, under the definition of the "Relative Long Tail", if the sum of the market shares of the movies above a certain rank drops (or if the sum of the market shares of the bottom movies rises), we conclude that the Long Tail effect is significant. We consider the cutoff points to be in percentage terms, e.g., top 1%, bottom 1% of all the movies during month *t*, thus adjusting for current active product variety Following the conventional categorization of product popularity in the literature (e.g., Anderson, 2004; Brynjolfsson et al., 2010), we refer to the movies in the top ranks as hits, and the movies in the bottom ranks as niches. One disadvantage of this "Relative Long Tail" definition as argued by Brynjolfsson et al. (2010) is that introduction of a large number of new niche products, each with very low demand, will trivially cause increasing demand for hits by the relative metric. We address this issue in the Subsection 4.3 and show that newly added movies in our data are not necessarily niche or hit products.

For the third measure of the Long Tail suggested in Brynjolfsson et al. (2010), we estimate $PowerCoeff_t$, which is the coefficient of the log-linear relationship between the product rank and its market share during month *t*. We present the results that show consistent findings in Subsection 4.3.

In addition to these main variables of interest, we consider several control variables. The home video market is susceptible to economic trends and seasonality. For example, hit movies tend to be introduced during early summer and Christmas. In order to adjust for these temporal factors, we consider a continuous variable $Trend_t$, which is a series from one to 55 with an increment of one every month. We also introduce a categorical control variable $Month_t$, which is the calendar month of month *t*. We make *Trend* continuous because we have only one observation each month. We further control for movie characteristics (*Genre, MPAA, BoxOffice)* when analyzing individual movie demand.

To summarize, Table 2 presents a list of variable definitions.

Table 2: Movie-level Analysis Variable Definitions

Variable	Definition
$Gini_t$	The Gini coefficient of demand distribution during month *t*.
$Share_{jt}$	Market share of movie *j* during month *t*, i.e., the number of rentals of movie *j* divided by all the rentals of all movies during month *t*.
$PowerCoeff_t$	The regression coefficient of the log-linear relationship between the product rank and its market share during month *t*.
$Variety_t$	Total number of movies (in 1,000 titles) with at least one rental during month *t*.
$Month_t$	Categorical variable indicating the calendaer month *t=1,...,12*.
$Trend_t$	A trend from 1 to 55 with an increment of one every month.
$Genre_j$	Categorical variable indicating the genre of movie *j*.
$MPAA_{jt}$	Categorical variable indicating the MPAA rating of movie *j*.
$BoxOffice_j$	Cumulative box office revenue of movie *j* in $1,000.

4 Empirical Analysis and Results

Our objective is to examine how product variety affects the demand concentration. We showed in Table 1 that the skewness of the DVD rentals increased over time, suggesting increasing trend in demand concentration. Meanwhile, product variety also expanded quickly. Although these two trends seem to suggest that there is a positive correlation between product variety and demand concentration, we cannot use this positive correlation to infer causality. In particular, confounding factors can make our causal inference spurious. For example[3], consumers may have still rented most popular movies in VHS format in 2001, while they shifted to renting popular movies in DVD format in 2005, causing the demand concentration of DVD rentals to rise in 2005. In addition, the consumers who like popular "mass" movies may have been late to shift to DVD market, which would create an alternative explanation to the observed increasing demand concentration. In order to alleviate such concerns and identify the causal effect, we first propose an econometric model using time series techniques to capture trend, seasonality and potential serial correlations. Then we use an instrumental variable approach to address the potential endogeneity biases. Finally, we introduce additional data and conduct various robustness checks to further alleviate aforementioned alternative explanations.

[3]We thank the SE for the valuable advice to consider these two alternative explanations.

4.1 Relative Long Tail

4.1.1 Time Series Regression Model

We first estimate the following ordinary least square (OLS) time series regression model:

$$Gini_t = \beta_0 + \beta_1 Variety_t + \beta_2 Gini_{t-1} + \beta_3 Month_t + \beta_4 Trend_t + \beta_5 Trend_t^2 + \varepsilon_t. \tag{1}$$

In this model, we include one month lagged dependent variable $Gini_{t-1}$, i.e., first order autoregressive component AR(1), to model the potential serial correlations of the errors (Kennedy, 2003). The serial correlations may arise because of unobserved exogenous shocks to demand concentration. For example, some particularly popular movies may have been released in a month and their popularity may have sustained over months because of rising social media and its word-of-mouth effect, which will cause positive serial correlation. In addition, we use $Month_t$ to control for the seasonality of the rental demand. We further use both $Trend_t$ and $Trend_t^2$ to adjust for industry and economic trends. We elect not to use simply the linear term of *Trend* because the variance inflation factors (VIFs) of the linear trend and *Variety* are both above 10, which is a rule of thumb for multicollinearity (Kennedy, 2003). Previous research has also used the polynomial specification to control for the trend in recorded music (Waldfogel, 2012). Alternatively, we apply first-differencing of the dependent variable to remove the trend and make the time series stationary (Makridakis et al., 2008). In particular, we estimate

$$\Delta Gini_t = \alpha_0 + \alpha_1 Variety_t + \alpha_2 Gini_{t-1} + \alpha_3 Month_t + \xi_t, \tag{2}$$

where $\Delta Gini_t = Gini_t - Gini_{t-1}$. After controlling for serial correlation, seasonality and trend, the coefficients of interest β_1 and α_1 represent the impact of increasing product variety by 1,000 on the Gini coefficient. If they are positive, increasing product variety may increase demand concentration; however, if they are negative, increasing product variety may reduce demand concentration.

4.1.2 Two-Stage-Least-Squares (2SLS) Model

Although Model 1 is useful as a preliminary estimator, it may not address two potential endogeneity issues. The first one is omitted variable bias. The omitted variable bias from OLS is given by $r_{Tx}\beta_x s_x/s_T$, where r_{Tx} is the correlation between the omitted variable x and the endogenous variable T (product variety), β_x is the relationship between the omitted variable x and the dependent variable (demand concentration), s_x and s_T are the standard deviations of x and T. In our setting, any monthly market condition that is systematically related to both product variety and demand concentration is a potential omitted variable. For example, one potential omitted variable is the effectiveness of search/personalization enabled by technology advances, which evolves over time. It might be negatively associated with the demand concentration because those advanced search tools generally enable the heterogeneous consumers to find their niches, i.e., $\beta_x < 0$ (e.g., Brynjolfsson et al., 2011). In addition, the effectiveness of search should be limited by product variety, namely $r_{Tx} < 0$ because large product variety may create distractions from effective search. The standard deviations of both x and T should be positive. Hence, OLS may overestimate the true impact of product variety on the demand concentration.

The second potential endogeneity source is simultaneity bias: We have theories predicting different potential effects of product variety on demand concentration. It is also likely that during the months when firms anticipate low demand concentration (i.e., increasing demand for niche movies) they may choose to increase the variety of available movies because a large product variety enables the firms to leverage the demand for those products that would not otherwise be offered, i.e., niches, and to maximize their total revenues. This loop of causality between product variety and demand concentration may cause a simultaneity bias, which specifically may cause a downward bias for the impact of product variety on demand concentration.

In order to address the aforementioned potential endogeneity issues, we adopt an instrumental variable 2SLS approach (Angrist and Krueger, 1994), which is widely used to alleviate the endogeneity problem (Kennedy, 2003). A valid instrumental variable should be uncorrelated with the error, i.e., satisfy the exclusion restriction and correlated with the endogenous regressor, i.e., satisfy the relevance condition (Wooldridge, 2002). In other words, the instrument should affect the dependent variable (demand concentration) only through the endogenous regressor (product variety).

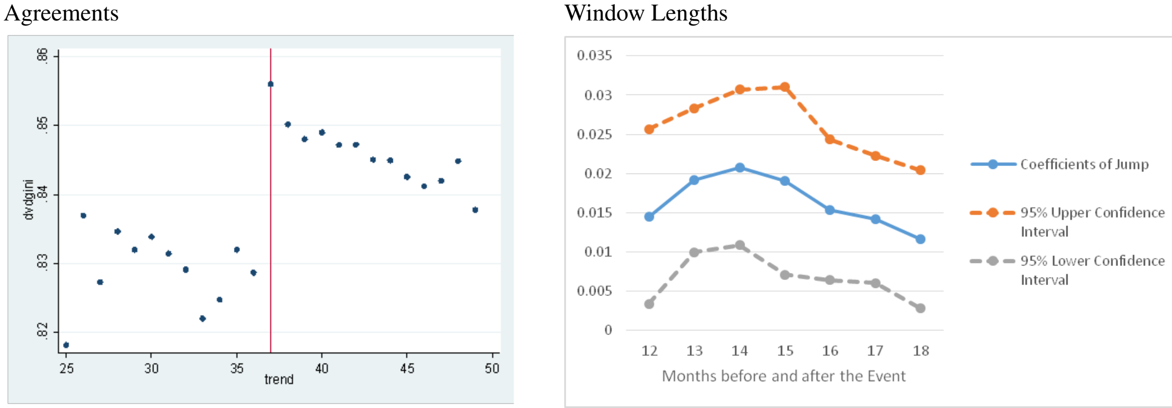
We propose an exogenous shock to product variety as a candidate for a valid instrumental variable: the implementation of new agreements with suppliers to increase product offerings from January 2004 onwards. In particular, we create a dummy variable $Jump_t$, which equals one for all the observations after January 2014 and equals zero for all other observations. It is highly correlated with $Variety_t$ (correlation is 0.95) because product variety greatly expanded after the implementation of the new agreements as shown in Figure 1. In addition, as expected, the coefficient of *Jump* is significant and positive in the first-stage regression (coefficient is 6.11). The F-statistics from the first stage is 1,651 significantly higher than 10, the suggested rule of thumb for weak instruments (Staiger and Stock, 1997). For these reasons, *Jump* as an instrumental variable is not weak and should satisfy the relevance condition of a valid instrument.

Furthermore, the implementation of the new agreements should satisfy the exclusion restriction assumption of a valid instrumental variable because the timing of the new agreements is an exogenous shock. Specifically, the 10K does not mention any other reasons for the new agreements than offering more movies. In addition, the number of total market rental turns seems to have grown steadily from 2001 to 2005, suggesting that the agreement timing is not related to the total market demand. Similarly, Figure 2a shows that the Gini coefficients close to the implementation of the new agreements (month 37, marked with a vertical line). If the trends of Gini are different before and after the event, they may indicate some endogeneity of signing the agreements (e.g., the Company wants to reverse the increasing demand concentration). However, as can be seen, the trends remain in the same declining pattern; the agreement simply shifts the trend upwards. For these reasons, the implementation of the new agreements should be exogenous. Admittedly, a peak appeared in the total rental turns in January 2004, but there are similar peaks during every January because of seasonality, which is already controlled for in the model.

Figure 2:

a) Monthly Gini Coefficients Close to the New Variety Agreements

b) Regression Discontinuity Design Results with Varying Window Lengths



4.1.3 Regression Discontinuity Design (RDD)

In the previous subsection, we explained that the implementation of the new agreements is likely to be an exogenous shock to expand product variety. In Figure 1, we also observe that product variety sharply jumps after the new agreements with Program Suppliers (mainly studios), which provides us with an opportunity to use a regression discontinuity design (RDD) to identify the causal effect of variety expansion on demand concentration. The RDD is a quasi-experiment pretest-posttest design that has become increasingly popular in the studies of statistics, econometrics, political science and epidemiology (Imbens and Lemieux, 2008). This method assigns a cutoff above or below which a random intervention happens. The observations closely lying on one side of the cutoff are the control group, while the observations closely lying on the other side of the cutoff are the treatment group. Since the intervention is exogenous, observations barely received treatment are comparable to those who just barely did not receive treatment, all else being equal, which enables researchers to evaluate the average treatment effect of the intervention. In our setting, the new agreements are an exogenous intervention, so those months shortly before (after) the new agreements compose a control (treatment) group, after we control for time-series components. In addition, new agreements significantly increased product variety. Hence, the effect of the agreements reflects the effect of product variety expansion.

Specifically, we estimate the following regression model:

$$Gini_t = \beta_0 + \beta_1 Jump_t + \beta_2 Gini_{t-1} + \beta_3 Month_t + \beta_4 Trend_t + \beta_5 Trend_t^2 + \varepsilon_t, \tag{3}$$

where β_1 is the average treatment effect of product variety expansion, all else being equal. We drop the observation on the 37th month (new agreement implmentation month) to reduce the contamination of the agreements implementation (but the result is the same if we include this month).When selecting the sample size, we choose the beginning of the window to be 25th and the end to be 49th month, 12 months before and after the 37th month, so that we can estimate the 12-month categorical variable to control for seasonality. We further estimate the magnitude of the agreements with varying observation window lengths to show consistency. Ideally, we would have liked to examine even shorter window length. But since our model has the 12 monthly dummies and the analysis is at the month level, we need at least 24 data points to estimate all the coefficients. Nevertheless, as a robustness check, we examine shorter window lengths (two months, three months, and four months before and after the implementation month) including only variable $Jump_t$ in the regression, and find both quantitatively and qualitatively consistent results.

4.1.4 Individual Movie Demand

Besides examining the demand concentration at the aggregate level, we are also interested in understanding the effect of product variety on the demand for each individual movie title that belongs to a particular popularity segment, namely, a hit or a niche movie. The individual movie level analysis allows us to control for movie characteristics, although the "classical" Long Tail literature makes no predictions about specific movie categories, such as an action or a drama movie. We perform the following regression analysis on hit and niche movies, separately:

$$\begin{aligned}\log(Share_{it}) &= \beta_0 + \beta_1 Variety_t + \beta_2 \log(Share_{it-1}) + \beta_3 Characteristics_i + \\ &\quad \beta_4 Month_t + \beta_5 Trend_t + \beta_6 Trend_t^2 + \varepsilon_{it} \quad i \in \text{a hit or a niche movie.} \end{aligned} \tag{4}$$

In these models, *Characteristics* include the control variables *BoxOffice, MPAA* and *Genre* defined in Table 2. We also include one month lagged variable $Share_{it-1}$, $Month_t$, $Trend_t$ and $Trend_t^2$ to adjust for auto-correlated movie demand, seasonality and trend, respectively. We logarithmically transform $Share_{it}$ for interpretation purposes.

We use top 1% as a cutoff to define a hit and bottom 15%to define a niche movie. A larger cutoff for the niche movies allows us to collect two samples of comparable sizes (3,148 vs. 1,573 movies) because many bottom ranked movies were not released to theaters, thus having no box office information. As a robustness check, we also use the cutoffs 1% and 10% to define a hit or a niche movie in fixed-effects models[4], replacing movie-characteristics with movie fixed effects, which also yields comparable sample sizes. Furthermore, we use Huber-White estimation to correct standard errors. We finally conduct two-sample t tests of the coefficients of β_1 within different samples to compare the effects of product variety on the demand for a hit and a niche movie. Please note that our definitions of hits and niches allow for the popularity of a DVD to vary from month to month. For example, a DVD that is popular in 2001 may become a niche title in 2005 because DVDs tend to have short-lived popularity cycles (Zentner et al., 2013). .

4.2 Results

4.2.1 Relative Long Tail - Gini Coefficient Analysis Results

Table 3 shows the results of measuring the effect of product variety on demand concentration using the Gini coefficient. We observe that the coefficients of *Variety* are consistently positive and significant across models (they are 0.0028, 0.0013, 0.0029 and 0.0013, respectively), suggesting that product variety may

[4]Adopting the 0.1% cutoff reduces the sample size to less than 30, which is insufficient for statistical inference. We therefore stick with cutoffs 1% and 10% to define a hit or a niche.

increase demand concentration. Interpreting the coefficient estimated by 2SLS, we find that increasing product variety by 1,000 titles may increase the Gini coefficient of DVD rentals by 0.0029. The effect of product variety expansion from the beginning of our study period to the end is estimated to be 0.06 ($Variety_1$= 3.097, $Variety_{55}$ = 23.072). Although our model is linear, we do not extrapolate our results to infinite product variety because more data are warranted in case of extremely large variety. Furthermore, the coefficient of *Jump* is significant and positive, suggesting that variety expansion due to new agreements is likely to increase demand concentration by 0.0145, controlling for everything else.

The coefficients of $Trend_t$ and $Trend_t^2$ are statistically significant in Model 1, but they become insignificant in the RDD mostly because they lose power in a smaller sample (24 observations). In addition, in both OLS- and 2SLS-estimated Model 1 results, the coefficients of $Trend_t$ are positive (0.0008), while the coefficients of $Trend_t^2$ are negative (-0.00002), suggesting that there may have been a flat concave trend in demand concentration (the inflection point is around the $0.0008/(2 \times 0.00002)$ =20th month). Both the trend effect and product variety expansion may have contributed to the rise of the Gini coefficient from 0.806 in January 2001 to 0.8427 in July 2005. Finally, the coefficients of $Gini_{t-1}$ are insignificant across models except in the Model 2. The negative sign of $Gini_{t-1}$ in Model 2 is unsurprising because a large demand concentration from the previous month may reduce the increment of the demand concentration during the subsequent month.

Table 3: Regression of Individual Movie Market Share

	Model 1 Estimated by OLS	Model 2 First Differencing	Model 1 Estimated by 2SLS	Model 2 First Differencing Estimated by 2SLS	Model 3 Regression Discontinuity
$Variety_t$	0.0028***	0.0013***	0.0029***	0.0013***	
	(0.0007)	(0.0003)	(0.0006)	(0.0003)	
$Gini_{t-1}$	0.1004	-0.6543***	0.0940	-0.6822***	-0.1812
	(0.1479)	(0.1450)	(0.1249)	(0.1422)	(0.1601)
$Trend_t$	0.0008**		0.0008***		0.0012
	(0.0003)		(0.0002)		(0.0011)
$Trend_t^2$	-0.0000**		-0.0000***		-0.0000
	(0.0000)		(0.0000)		(0.0000)
$Jump_t$					0.0145*
					(0.0048)
$Month_t$	Included	Included	Included	Included	Included
Observations	54	54	54	54	24
R^2	0.866	0.524	0.837	0.523	0.978

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

2) Standard errors are shown in the parentheses.

In Table 3, we show the RDD results with 12 months before and after the new agreements implementation. We then repeat this RDD analysis with varying window lengths, the results of which are shown in Figure 2b. As can be seen, the coefficients of *Jump* are consistently above zero, when the window length varies from 12 months (the minimum required to estimate all the coefficients) before and after the agreements implementation to 18 months (the maximum in our sample). Moreover, the 95% confidence intervals are also all positive, lending further support that variety expansion due to new agreements is likely to increase demand concentration.

4.2.2 Individual Movie Demand

Table 4 presents the regression results of individual movie market shares. The coefficients of *Variety* are consistently negative for the market share of either a hit or a niche movie, which implies that increasing product variety dilutes the demand for any movie regardless of its popularity segment. However, after conducting two-sample *t*-tests, we notice that the coefficients in the models of hits are significantly less negative than those in the models of niches. For example, between top 1% and bottom 15% movies (first two columns), the coefficient of *Variety* in the niche model (-0.0852) is about twice as negative as that in the hit model (-0.0445). In the 1% cutoff FE models (middle two columns), the coefficient of *Variety* in the niche model (-0.1189) is close to four times as negative as that in the hit model (-0.0376). In the 10% cutoff FE models (last two columns), the coefficient of *Variety* in the niche model (-0.0863) is approximately four times as negative as that in the hit model (-0.0204). The sharp differences between the hit models and the niche models suggest that the dilution of the demand because of the increased product variety is less significant for hits than for niche movies. In other words, the demand gap between a hit and a niche DVD is widened by increasing product variety, thus contributing to our previous finding that product variety increases the demand concentration on hits (relatively speaking).

4.3 Robustness Checks

4.3.1 Alternative "Long Tail" Measures

To provide robustness checks, we first use some alternative measures of Long Tail which were introduced in the literature on demand concentration.

Table 4: Regression of Individual Movie Market Share

	$\log(Share_{it})$ $i \in$Top 1%	$\log(Share_{it})$ $i \in$Bottom 15%	$\log(Share_{it})$ $i \in$Top 1% Estimated by FE Model	$\log(Share_{it})$ $i \in$Bottom 1% Estimated by FE Model	$\log(Share_{it})$ $i \in$Top 10% Estimated by FE Model	$\log(Share_{it})$ $i \in$Bottom 10% Estimated by FE Model
$Variety_t$	-0.0445***	-0.0852***	-0.0376***	-0.1189***	-0.0204***	-0.0863***
	(0.0035)	(0.0102)	(0.0043)	(0.0238)	(0.0012)	(0.0034)
$\log(Share_{it-1})$	0.1336***	0.2137***	0.0252***	-0.0003	0.3763***	0.0895***
	(0.0069)	(0.0174)	(0.0063)	(0.0179)	(0.0026)	(0.0048)
$Trend_t$	-0.0284***	-0.0249***	0.0079	-0.0523**	-0.0289***	-0.0279***
	(0.0019)	(0.0036)	(0.0052)	(0.0180)	(0.0008)	(0.0020)
$Trend_t^2$	0.0003***	0.0004***	-0.0007***	0.0004*	-0.0000	0.0002***
	(0.0000)	(0.0001)	(0.0001)	(0.0002)	(0.0000)	(0.0000)
$Characteristics_i$	Included	Included	Replaced by movie FE	Replaced by movie FE	Replaced by movie FE	Replaced by movie FE
$Month_t$	Included	Included	Included	Included	Included	Included
Observations	3,148	1,573	4,364	2,457	54,429	32,931
R^2	0.756	0.714	0.531	0.341	0.758	0.350

1) *: p-value<0.05, **: p-value<0.01, ***: p-value<0.00

2) Standard errors are shown in the parentheses.

Relative Long Tail Using Percentage Popularity Segments instead of Gini Coefficients Similar to dividing movies into deciles for each month (e.g., Zhou and Duan, 2012), we employ the following model to examine the impact of product variety on the particular popularity segment:

$$\begin{aligned} \log(\sum_{j\in l} Share_{jt}) &= \beta_0 + \beta_1 Variety_t + \beta_2 \log(\sum_{j\in l} Share_{jt-1}) + \beta_3 Month_t + \\ &\quad \beta_4 Trend_t + \beta_5 Trend_t^2 + \xi_{lt} \quad \forall l \in \text{relative cutoff points.} \end{aligned} \tag{5}$$

In these models, $\Sigma_{j\in l} Share_{jt}$ represents the total market share of a particular popularity segment, hits or niches. We include the lagged dependent variables and same other time series variables to control for serial correlation, trend and seasonality. Coefficient β_1 captures the effect of product variety on demand for a particular popularity segment. Since the OLS estimation of this model may be susceptible to the similar endogeneity issues to Model 1, we use the same instrumental variables approach as in Subsection 4.1.2.

Table 5 shows the results of impact of product variety on demand share for different percentage popularity segments. We observe that the coefficients of *Variety* are both significant and positive in the models of the hit segments (first two columns). In particular, if product variety increases by 1,000 titles, the share of the top 1% of movies increases close to three times as fast as the share of the top 10% of movies (coefficients are 0.0196 and 0.0058, respectively), indicating that with product variety increasing the demand for the "hits of the hits" significantly outpaces the demand for less popular movies. In other words, the gap between the demand for the top 1% and the demand for the top 10% of the movies is likely to shrink, a phenomenon of a more intense demand concentration. In contrast, the coefficients of *Variety* are both significant and negative in the models of the niche segments (last two columns). Specifically, with product variety increasing, the bottom 1% of the movies lose demand faster than the bottom 10% movies (coefficients are -0.2129 and -0.0528, respectively), suggesting that the gap between the demand for the bottom 1% movies, i.e., the "niches of the niches", and the demand for the bottom 10% movies is likely to be widened by product variety, also a phenomenon consistent with a higher demand concentration. Interpreting the coefficients, we find that increasing product variety by 1,000 titles may increase the market share of the top 1% of DVDs by 1.96%, the market share of the top 10% of DVDs by 0.58%. However, increasing product variety by 1,000 titles may reduce the market share of the bottom 1% of DVDs by 21.29% and the market share of the bottom 10% of DVDs by 5.28%.

Besides the 1% and 10% cutoffs, we also examine the ten deciles of the entire distribution for each month. Only for the top decile the coefficient of $Variety_t$ is significantly positive (0.0058), while the coefficients for the rest of the deciles are all significantly negative, supporting the results in Table 5. Additionally, we analyze the effect on the total number of rental turns in each popularity segment, namely, the "Absolute Long Tail" measure. The results are still qualitatively consistent. For example, increasing product variety by 1,000 titles may increase the total number of rentals of the top 1% of DVDs by 3.91%, but it may reduce the total number of rentals of the bottom 1% of DVDs by 18.53%. We further replicate the analyses using the regression discontinuity design proposed in Subsection 4.1.3 and find congruent results.

Admittedly, this "Relative Long Tail" measure may tautologically create the result of less significant "long tail" if a more significant number of unpopular DVDs were introduced to the product variety than popular ones. To address the concern, which we additionally alleviate in Subsection 4.3.2, and to show that our findings are specific to the popularity-segment of the movies[5], we randomly select 1% and 10% of the movies every month and replicate our analysis of market share with these movies. If indeed more extremely niche DVDs than popular ones were added to the product variety every month, we would expect the coefficients of *Variety* to be negative. However, it turned out that the coefficients of *Variety* are all insignificant, which suggests that product variety has no effect on the market share of any randomly selected

[5]We thank one anonymous reviewer for this helpful suggestion.

Table 5: Regression of Hit and Niche Movie Market Shares

	Top 1%	Top 10%	Bottom 1%	Bottom 10%
$Variety_t$	0.0196***	0.0058***	-0.2129***	-0.0528***
	(0.0054)	(0.0017)	(0.0202)	(0.0077)
$\log(\sum_{j \in l} Share_{jt-1})$	0.3120*	0.2586*	-0.0019	0.0788
	(0.1216)	(0.1263)	(0.0914)	(0.1113)
$Trend_t$	-0.0019	0.0023***	0.0254***	0.0090***
	(0.0017)	(0.0007)	(0.0057)	(0.0027)
$Trend_t^2$	-0.0000	-0.0001***	0.0002	0.0000
	(0.0000)	(0.0000)	(0.0001)	(0.0001)
$Month_t$	Included	Included	Included	Included
Observations	54	54	54	54
R^2	0.836	0.878	0.977	0.937

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

2) Standard errors are shown in the parentheses.

group of movies because the demand gain for hit DVDs may cancel out the demand loss for niche DVDs. Instead, these results by falsification reinforce our empirical findings that product variety affects the market share of specific popularity segments.

Log-Linear Relationship between Product Rank and Market Share Another often used metric of the Long Tail is the slope of the log-linear relationship between rank of the product and its demand (Brynjolfsson et al., 2009, 2010; Hinz et al., 2011). This metric assumes that the demand follows a power law distribution and the slope of the log-linear relationship therefore captures the degree of concentration. We visually check the monthly demand distribution in our data and the distribution patterns seem to have power law-like shapes. In addition, the parsimonious log-linear models consistently yield high R^2s (all are greater than 0.8), suggesting a sufficient "goodness-of-fit". Specifically, every month we estimate the following log-linear model:

$$\log(Share_{it}) = \beta_{0t} + \beta_{1t} \log(Rank_{it}) + \varepsilon_{it} \quad \forall i \in 55 \text{ months.} \tag{6}$$

In these models, $Rank_{it}$ is the monthly popularity rank of DVD i during month t. A low rank indicates a higher market share. Coefficients β_{1t}, which we refer to as $PowerCoeff_t$ in Table 2, should be negative and they are the coefficients of interest. If $PowerCoeff_1 > PowerCoeff_2$, demand concentration during month 2 is higher than that during month 1 because the gap between movies having different ranks is widened during the second month.

After we estimate the 55 $PowerCoeff_t$s, we specify the following time series regression model:

$$PowerCoeff_t = \alpha_0 + \alpha_1 Variety_t + \alpha_2 PowerCoeff_{t-1} + \alpha_3 Month_t + \alpha_4 Trend_t + \alpha_5 Trend_t^2 + \xi_t \qquad (7)$$

We use the same instrumental variable and the 2SLS procedure to estimate the model because it may suffer from the same aforementioned endogeneity issues. Both OLS and 2SLS estimates of α_1 are significant and negative (-0.0059 and -0.0056, respectively), suggesting that product variety is again likely to increase demand concentration, all else being equal. We also use the regression discontinuity design analysis, which yields consistent results.

In estimating Model 6, we choose not to fit all data across 55 months in one log-linear power law model and not to estimate the moderating effect of *Variety* to measure the effect of product variety on demand concentration. Fitting all the data in one log-linear relationship model assumes that the total number of products remains the same across markets/time periods. For example, Brynjolfsson et al. (2011) control for product assortment sizes both online and offline and estimate the moderating effect of the Internet channel to examine the effect of the Internet channel on demand concentration. Since the product variety of DVDs changes over time, large values of $Rank_{it}$ do not happen equally during early months. For example, a $Rank_{i55}$ equaling 20,000 is unavailable during month 1 because there are only about 3,000 DVDs during the first month. Hence, fitting all the data in one model to estimate the effect of *Variety* may cause selection bias. Instead, we choose to estimate 55 coefficients separately.

4.3.2 New Movie Composition

Brynjolfsson et al. (2010) provide a hypothetical example in which introducing a large number of new niche products, each of which has very low sales, will trivially cause increasing demand for hits by the "Relative Long Tail" measure. Since in practice we do not a priori know that newly added movies are necessarily niche products, we test whether or not this is the case in our data. We first define a newly added movie as a movie that was first observed in the current month or the previous month. For example, if movie *i* first appeared in our data set in January 2012, it is coded as a newly added movie in January and February of 2012. To verify whether the newly added movies are necessarily comprised of niche movies or not, we develop two approaches.

For the first approach, we examine the correlations between product variety and $ShareNew_{pt}$, the proportion of new movies that rank in the top 1% (hits) or in the bottom 1% (niches) in month *t*, i.e., the number of newly added movies that are hits or niches divided by the total number of hits or niches (old and new movies). For example, if there are 1,000 movies ranking in the top 1% in a certain month and 200 of them are newly added movies, then $ShareNew_{top1\%,t} = 200/1,000 = 0.2$ for the hits in that month. We restrict our analysis from April 2001, the fourth month in our data, to avoid left censoring issues. We specify the models as follows:

$$ShareNew_{pt} = \beta_0 + \beta_1 Variety_t + \zeta_t \quad p \in \text{top } 1\% \text{ or bottom } 1\%. \tag{8}$$

In these parsimonious models, when $p \in$bottom 1%, a positive β_1 would imply that increasing product variety is associated with more *newly-added* niche movies, an evidence for the hypothetical example that would trivialize the “Relative Long Tail” measure.

For the second approach, we compute how the average “age” of a hit (top 1%) or a niche (bottom 1%) movie changes as product variety changes. In particular, we first define the *Age* of a movie as the number of months that a movie has been in the data set since its first appearance. We also restrict the analysis from April 2001 to avoid the left censoring issues. We specify the model as follows:

$$Average(Age_{pt}) = \gamma_0 + \gamma_1 Variety_t + \xi_t \quad p \in \text{top } 1\% \text{ or bottom } 1\% \text{ movies.} \tag{9}$$

In these models, when $p \in$bottom 1%, a negative γ_1 would suggest that the expanded product variety is associated with *newer* movies in the niche segment, another evidence for the aforementioned hypothetical example against the “Relative Long Tail” measure.

Table 6 shows the results of the new movie composition analysis. As can be seen, both the coefficients of *Variety* in the *ShareNew* model (Model 8) are significant and negative (-0.002 for hits and -0.0125 for niches). The negative signs suggest that the share of newly added movies among either hits or niches drops as product variety expands, which does not support the hypothetical example above in our data. In addition, for both hits and niches, the coefficients of *Variety* in the average age model (Model 9) are significant and positive (0.9628 for hits, 0.758 for niches), implying that both the hits and the niches are comprised of more and more “older” movies as product variety increases, which does not support the hypothetical example in our data, either. These two models seem to suggest that newly added movies do not necessarily become hits or niches as soon as more and more movies are added into product offering.

Table 6: New Movie Composition Analysis

	Model 8 (Top 1%)	Model 8 (Bottom 1%)	Model 9 (Top 1%)	Model 9 (Bottom 1%)
Variety	-0.0020*	-0.0125***	0.9628***	0.7580***
	(0.0009)	(0.0025)	(0.0646)	(0.0698)
Constant	0.2018***	0.3706***	5.8902***	3.4187***
	(0.0118)	(0.0317)	(0.8240)	(0.8900)
Observations	52	52	52	52
R^2	0.087	0.335	0.816	0.702

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

2) Standard errors are shown in the parentheses.

Besides the composition of the newly added movies, we also examine the subsample of the movies that were likely to be independent movies, i.e., not released by the seven largest studios in our data[6]. Focusing on these independent movies, which are dissimilar from those released by big studios for mass appeal, could generate additional insights of the impact of product variety on demand concentration in a horizontally differentiated market. However, we find that the subsample results are consistent with the entire sample results.

4.3.3 Alternative Format/Channel Interactions

During our study period, although home movie format VHS was dwindling, it still had a significant market share, especially during the early months. If consumers tended to rent disproportionately more popular movies than niche movies in the VHS format in the early months, they may have cannibalized the demand for the popular DVDs. In addition, some of the most popular movies may not have been licensed to the DVD format in the early months. As VHS industry kept shrinking, the aforementioned cannibalization may have declined. Meanwhile, more and more movies were licensed to DVDs. These two trends can provide alternative explanations to the observed increasing demand concentration. Furthermore, during the same study period, online DVD rental channel, represented by Netflix, grew quickly. Research has shown that online channel tends to reduce demand concentration (Brynjolfsson et al., 2011; Zentner et al., 2013). If an increasingly significant online channel cannibalizes the demand for niche movies in the physical stores in later months, it will provide an alternative explanation to the higher demand concentration. Although

[6]We thank one anonymous reviewer for this helpful suggestion.

we made an attempt to control for industry trend-related issues using our time series control variables in the main models, in what follows we collect additional data on VHS rentals and online movie rental data to alleviate these important identification concerns. We gather VHS rental data from the same Company that provides us with the DVD rental data. In addition, we collect Netflix Prize competition data as a proxy for online rental. Netflix, a major U.S. online movie/TV series rental service made public a large data set during the competition to elicit the best algorithm to improve its movie recommendation system. The data set consists of all the consumer-level movie ratings through the Netflix website from 2000 to 2005, encompassing our study period.

It is important to stipulate here that using rating data as a proxy for actual rental demand at Netflix can be inaccurate. Despite this limitation, the rating data can provide insights into what movies consumers are aware of and are interested in. We believe that the number of ratings can be a reasonable proxy for online demand for the following reasons. First, Netflix incentivizes and facilitates its consumers to rate every time they watch a movie to improve their personalized recommendations. It is true that the propensity to rate a product can be due to the social influence (e.g., Moe and Trusov, 2011; Susarla et al., 2012). Nevertheless, the social influence factor may be rather limited in our setting. Unlike on Netflix today, when users can see what movies are popular among his/her friends in the social networks, during our time frame (2001 - 2005), Netflix focused on eliciting individual ratings without showing other raters' identities. There was barely any interaction among the users except that users could see the average rating of the movie so far. Even though this social influence on propensity to rate exists, we have no reason to believe that such an influence changes over our study period, so it should not affect our main analysis. Muchnik et al. (2013) additionally find that the number of ratings is immune from social influence, giving us some further assurance about using it as a proxy for demand. Second, Netflix' 10Ks in 2002 and 2003 reported that the company shipped out 5 million and 9 million DVDs correspondingly (they did not go public until 2002 and stopped reporting the total shipment after 2003), while our data included about 4.3 million ratings and 10 million ratings in those two years, respectively. Third, we find a strong correlation between the monthly number of rentals in the DVD rental data and the monthly number of ratings in the Netflix data among the matched movies in the two data sets, controlling for time fixed effects. This strong correlation is consistent with previous literature (Chen et al., 2004), which suggests a strong connection between product demand and the number of consumer reviews. For all these reasons, we proceed with utilizing Netflix ratings as a proxy for online movie demand.

We first replicate Model 1 on a set of common movies between DVD and VHS formats to alleviate the interaction effect between the two formats. Particularly, we regress the monthly coefficients of the aggregated demand over the two formats on the monthly size of the common set of the movies. Aggregating demand across formats and channels can alleviate the potential interaction effects. Furthermore, we replicate this model on a set of common movies among DVD, VHS formats and online channel to account for the interaction effect of the formats and channels. Since all these models may be susceptible to the same endogeneity issues discussed in Subsection 4.1.2, we use the instrumental variable 2SLS approach to estimate the models.

Table 7 presents the estimates of alternative format/channel interactions. The estimates of the monthly variety in the aggregated demand models are consistently significant and positive (0.0019 and 0.0172, respective), which suggests that product variety may increase demand concentration across formats and channels. These congruent signs with our main result in Table 3 further reassure us that our results are not driven by the alternative explanations of the format/channel interactions.

Not all the movies were licensed to DVD in early days when the total DVD product variety was small, and niche movies were probably more likely to remain unlicensed than popular movies because distributors probably expedited DVD adoption by popular titles. In fact, in our data, we calculate that close to 100% of niche VHS (bottom 1% of VHS titles) were not licensed to DVDs in early 2001, while at the same time only 6% of popular VHS (top 1% of VHS titles) were unlicensed to DVDs. In contrast, over time, more niche VHS titles were licensed to DVDs, which not only alleviates the alternative explanation of the limited availability of popular titles at the beginning of the study period, but also makes our findings more conservative. Furthermore, DVD and VHS media may be perceived as two different products because of their differences in picture and sound quality, and in players that are needed, which may further limit the above concern of media format interactions. Finally, the aforementioned alternative explanations assume that other format/channel may exert cannibalization effect on the popular DVDs. However, Kumar et al. (2014) find an information spillover effect that consumers may sample some movies in one channel at a low cost which then stimulates the demand for those products in another channel. It is unclear a priori whether the information spillover effect or the cannibalization effect dominates in our setting. Kumar et al. (2014) suggest that the net effect of information spillover and cannibalization is positive for niche products because they find that the demand concentration for DVD sales decreases after consumers sample movies in the cable channel. Nevertheless, their finding may not apply in our setting because in their setting, watching a movie on the scheduled cable channel is arguably cheaper than buying a DVD (especially in early days) and can provide an information spillover source, while the cost difference between renting a DVD or a VHS from a physical store or renting a DVD from Netflix is not clear. Hence, we have no definite direction about where the information is spilled over from. If online DVD channel was a source of information spillover for niche movies because Netflix marketed on a wide selection of niche movies in those days, their growing trend would make our findings that there is a stronger demand for hits even more conservative. Regardless, Table 7 provides evidence that product variety may increase demand concentration notwithstanding movie format or channel format.

Table 7: Alternative Format/Channel Interactions

	DVD+VHS	DVD+VHS+Online
Monthly Variety	0.0019***	0.0172***
	(0.0004)	(0.0046)
$Gini_{t-1}$	0.2339*	0.1911
	(0.1130)	(0.1295)
$Trend_t$	0.0005*	-0.0008*
	(0.0002)	(0.0004)
$Trend_t^2$	0.0000	0.0000*
	(0.0000)	(0.0000)
$Month_t$	Included	Included
Observations	54	54
R^2	0.919	0.906

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

2) Standard errors are shown in the parentheses.

4.3.4 Heterogeneous Consumer Migration from VHS to DVD

Between 2001 and 2005, DVD rental was adopted by more and more consumers. Some of them turned from physical rental stores to online channels. Heterogeneous consumers may have entered or exited the offline DVD rental market (our main setting) at different times. For example, consumers who like movies having a "mass" appeal, i.e., popular movies may be late to shift to DVD format from VHS, which would create another mechanism for causing higher demand concentration. Although we do not know a priori what type of consumers tends to migrate in our data sample, we can track a balanced cohort of consumers to alleviate this alternative explanation (Zentner et al., 2013). Ideally, we would like to track such consumers in our physical stores. But since the offline rental data are not available at the consumer level, we turn to online Netflix data to construct a balanced panel throughout the entire study period.

We examine how product variety affects the proportion of most popular, less popular and least popular movies that an average consumer rated under the classic ABC classification, according to which (Monczka et al., 2008), we call the top 20% of the movies every month A movies, we call middle 30% B movies, and we call the bottom 50% C movies. Then we create variables $AShare_{it}$, $BShare_{it}$, and $CShare_{it}$ that represent the shares of A, B and C movies that a consumer i rated in month t. Besides these dependent variables, to control for individual consumer characteristics we include variables $Frequency_{it}$ and $AvgRating_{it}$, i.e., the number of movies and the average rating that consumer i gave in month t. A large $Frequency_{it}$ may indicate that consumer i is likely to be a "movie buff", a heavy user. A low $AvgRating_{it}$ may suggest that consumer i tends to be critical. In addition, we use categorical variable $Month_t$ to adjust for the seasonality of movie rentals. We also introduce μ_i, a time-invariant preference for each consumer's movies through the panel data analysis to difference out time-invariant consumer heterogeneity 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 select the consumers who appeared in all 55 months in our study period, i.e., a balanced cohort of consumers and we employ the following fixed-effect panel regressions to predict the effect of product variety on the proportions of rating movies of different popularity levels:

$$\log(AShare_{it}) = \beta_0 + \beta_1 Variety_t + \beta_2 Frequency_{it} + \beta_3 AvgRating_{it} + \beta_4 Month_t + \mu_i + \varepsilon_{it}, \qquad (10)$$

where $Variety_t$ is the monthly number of different DVDs that had a least one rating in the Netflix data. We subsequently substitute the dependent variable with *BShare* and *CShare,* respectively.

Table 8 shows the results of the consumer-level ABC Classification analysis. The coefficient of *Variety* in the log(*AShare*) model is significant and positive (coefficient = 0.0041), suggesting that product variety may increase the share of A movies (top 20%) that an average user rated in a month. However, product variety may decrease the shares of B movies (middle 30%) and C movies (bottom 50%) because the coefficients of *Variety* in the other two models are both significant and negative (coefficients are -0.0041 and -0.0016). These consumer-level ABC Classification results are consistent with our movie level analysis results that product variety may increase the demand for hit movies and reduce the demand for niche movies. Since we conduct the analysis on the same cohort of consumers and find congruent results, we alleviate the alternative explanation of the heterogeneous consumers migration.

In addition, the significant and negative coefficient of *Frequency* for log(*Share*) and the positive coefficient for log(*BShare*) indicate that heavier consumers may rent more of the less popular movies, but they may not venture into extremely niche movies because the coefficient for log(*CShare*) is insignificant. Furthermore, the coefficient of *AvgRating* in the log(*AShare*) model is significant and positive, suggesting that higher ratings are associated with more A movies. In contrast, the coefficients of *AvgRating* in the log(*BShare*) and log(*CShare*) models are both significant and negative, indicating that lower ratings are associated with more B and C movies. Since user ratings may suffer from various social influence biases (e.g.,Li and Hitt, 2008; Moe and Trusov, 2011; Muchnik et al., 2013), we refrain from making causal inference of the user rating and movie quality.

We also used alternative dependent variables of demand for different popularity segment, such as the relative median rank of the movies that consumer *i* rated every month, i.e., median rank divided by product variety, the relative top 10% and the relative bottom 10% of the ranks. We further replicated Model 10 to analyze the entire population in the Netflix data. All of these additional robustness checks show qualitatively consistent results that product variety may increase consumers' demand for hit DVDs rather than niche ones.

Table 8: Consumer-level ABC Classification Analysis

	log(*AShare*)	log(*BShare*)	log(*CShare*)
$Variety_t$	0.0041***	-0.0041***	-0.0016***
	(0.0002)	(0.0002)	(0.0001)
$Frequency_{it}$	-0.0002***	0.0004***	-0.0001
	(0.0000)	(0.0000)	(0.0000)
$AvgRating_{it}$	0.0180***	-0.0127***	-0.0131***
	(0.0021)	(0.0024)	(0.0014)
$Month_t$	Included	Included	Included
Observations	14,905	14,905	14,05
Overall R^2	0.046	0.038	0.020

1) ***p-value<0.001.

2) Standard errors are in parentheses.

5 Conclusion and Discussion

The reasoning behind the Long Tail effect presumes that expanding product variety due to the adoption of information technology will satisfy consumers' increasingly heterogeneous tastes, thus causing the demand for niche products to rise. However, there is a general lack of empirical research that tests this presumption by analyzing the effect of product variety on demand concentration at both product and consumer level. In this paper, we analyze large data sets (that became available due to the implementation of information technologies in the movie rental industry) to identify the causal effect of product variety on the Long Tail phenomenon, taking potential endogeneity and alternative explanations into consideration. After controlling for industry trends, seasonality and potential serial correlations, we find that product variety is likely to increase demand concentration, i.e., boosting the demand for hits and reducing the demand for niche products, which contradicts the Long Tail effect. In our sample, we estimate that increasing product variety by 1,000 movies may increase the Gini coefficient by 0.0029 or increasing the market share of top 1% movies by 1.96% and reducing the demand for bottom 1% movies by 21.29%. Our finding is not mechanistically caused by our "Relative Long Tail" measure, which would artificially inflate the demand for hits if the newly added movies are primarily low-selling niche products; we provide evidence that the newly added movies are not predominantly niches. Instead, we explain that the increased product variety dilutes the demand for each individual movie, but this demand dilution is less significant for hits than for niche movies, thus contributing to our counter-Long Tail finding that product variety increases the demand concentration. We further conduct various robustness checks to alleviate alternative explanations: we use alternative "Long Tail" measures, we study media formats/channel interaction and we analyze possible heterogeneous consumer migration from VHS to DVDs. All of these robustness checks lead to qualitatively consistent results.

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 addition, focusing on niche products can lower economies of scale in ordering (Gallino et al., 2015) and increase supply-demand mismatch costs, such as discounts and inventory (Rajagopalan, 2013; Moreno and Terwiesch, 2015). 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) and more than 50% of the total home video rental disc market in 2013, while Netflix lost 470,000 subscribers to its DVD-by-mail business during the same time (Variety, 2013). 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). If a company decides to leverage its expansive product variety offering, it has to pay more attention to recommendation systems to help consumers discover the niche products (Oestreicher-Singer and Sundararajan, 2012; Adomavicius et al., 2013) and new omni-channel strategies, such as "ship-to-store", to ease the purchase process of niche items (Gallino et al., 2015). Finally, instead of considering product variety as the number of products offered, companies can use an attribute-based measure of product line breadth (e.g., range of fuel efficiency in automobiles or star power in movies) to hedge against demand uncertainty without necessarily increasing the number of products offered (Moreno and Terwiesch, 2015).

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 channels (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. We particularly focus on the brick-and-mortar stores to control for the similar business model and industry trend, and include VHS format and Netflix channel as a robustness check. In addition, during our study period between 2001 and 2005, digitalization of movie consumption (e.g., online streaming) was limited. Of course, movies are distributed in multiple channels nowadays, such as video-on-demand, theaters, BitTorrent, Tugg. In the age of this digital economy, we resonate with Brynjolfsson and McAfee (2014), who conjecture a stronger demand concentration because of the wider accessibility of popular products and the increased importance of demand side economies of scale. It is therefore more important than before to measure the total demand across channels for future research, although collecting data across channels is actually very difficult (Laporte, 2016).

Second, our online channel 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 online data as robustness checks. Admittedly, the rating data 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. We also find evidence that the actual DVD shipments on Netflix were similar to the number of ratings, which provides us some assurance of using the rating data as a reasonable proxy for online demand. 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 (e.g., Milkman et al., 2009)

Third, our data are at the market level and do not observe retailer-specific assortment and promotion decisions, which should affect demand concentration. If retailers acquire a larger number of and promote popular movies when facing quickly expanding product variety, this tendency creates the endogeneity issue, which we use instrumental variable 2SLS approach to correct. However, if the retailers stock more and promote niche items, this tendency will actually make our estimates even more conservative. Still, future research may consider the moderating effects of retail structure of product variety on demand concentration to shed light on the complicated mechanisms of the empirical results.

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) and inventory pooling (Bimpikis and Markakis, 2015) is warranted.

References

Adomavicius, G., J.C. Bockstedt, S.P. Curley, J Zhang. 2013. Do recommender systems manipulate consumer preferences? a study of anchoring effects. *Information Systems Research* **24**(4) 956–975.

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.

Association, Entertainment Merchants. 2015. A history of home video and video game retailing. URL `http://entmerch.org/press-room/industry-history`.

Baumol, W., E. Ide. 1956. Variety in retailing. *Management Science* **3**(1) 93–101.

Berger, J., M. Draganska, I. Simonson. 2007. The influence of product variety on brand perception and choice. *Marketing Science* **26**(4) 460–472.

Bimpikis, K., M. G. Markakis. 2015. Inventory pooling under heavy-tailed demand. *Management Science* .

Brynjolfsson, E., Y.J. 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.J. Hu, M.D. Smith. 2009. The longer tail: The changing shape of Amazon's sales distribution curve. MIT Working Paper.

Brynjolfsson, E., Y.J. 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.

Brynjolfsson, E., A. McAfee. 2014. *The second machine age: work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

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.

Caves, R.E. 2000. *Creative industries: Contracts between art and commerce*. Harvard Univ Press.

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

Chernev, A. 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**(2) 170–183.

Dewan, S, J Ramaprasad. 2012. Music blogging, online sampling, and the long tail. *Information Systems Research* **23**(3) 1056–1067.

Fasolo, B., R. Hertwig, M. Huber, M. Ludwig. 2009. Size, entropy, and density: What is the difference that makes the difference between small and large real-world assortments? *Psychology & Marketing* **26**(3) 254–279.

Fisher, M., J. Hammond, W. 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.

Gallino, S., A. Moreno, I. Stamatopoulos. 2015. Channel integration, sales dispersion, and inventory management. *Management Science* Forthcoming.

Gourville, J.T., D. Soman. 2005. Overchoice and assortment type: When and why variety backfires. *Marketing Science* **24**(3) 382–395.

Hauser, J.R., B. Wernerfelt. 1990. An evaluation cost model of consideration sets. *Journal of consumer research* **16**(4) 393.

Hervás-Drane, A. 2013. Word of mouth and sales concentration. 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.

Hosanagar, K., D. Fleder, D. Lee, A. Buja. 2014. Will the global village fracture into tribes? recommender systems and their effects on consumer fragmentation. *Management Science* **60**(4) 805–823.

Huffman, C., B.E. Kahn. 1998. Variety for sale: Mass customization or mass confusion? *Journal of Retailing* **74**(4) 491–513.

Imbens, G. W., T. Lemieux. 2008. Regression discontinuity designs: A guide to practice. *Journal of econometrics* **142**(2) 615–635.

Inman, J.J. 2001. The role of sensory-specific satiety in attribute-level variety seeking. *Journal of Consumer Research* **28**(1) 105–120.

Iyengar, S.S., M.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.

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

Kahn, B.E. 1995. Consumer variety-seeking among goods and services: An integrative review. *Journal of Retailing and Consumer Services* **2**(3) 139–148.

Kahn, B.E. 1998. Dynamic relationships with customers: high-variety strategies. *Journal of the Academy of Marketing Science* **26**(1) 45.

Kennedy, P. 2003. *A guide to econometrics*. The MIT Press.

Kumar, A., M.D. Smith, R. Telang. 2014. Information discovery and the long tail of motion picture content. *Management Information Systems Quarterly* **38**(4) 1057–1078.

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

Lancaster, K. 1990. The economics of product variety: A survey. *Marketing Science* **9**(3) 189–206.

Laporte, N. 2016. How neielsen plans to solve the audience puzzle. *Fast Company* .

Lehmann, D.R. 1991. Modeling choice among assortments. *Journal of retailing* **67**(3) 274–299.

Li, Xinxin, Lorin M Hitt. 2008. Self-selection and information role of online product reviews. *Information Systems Research* **19**(4) 456–474.

Makridakis, S., S. C. Wheelwright, R. J. Hyndman. 2008. *Forecasting methods and applications*. John Wiley & Sons.

McAlister, L. 1982. A dynamic attribute satiation model of variety-seeking behavior. *Journal of Consumer Research* 141–150.

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.

Moe, W.W., M. Trusov. 2011. The value of social dynamics in online product ratings forums. *Journal of Marketing Research* **48**(3) 444–456.

Monczka, R., R. Handfield, L. Giunipero, J. Patterson. 2008. *Purchasing and supply chain management*. Cengage Learning.

Moreno, A., C. Terwiesch. 2015. The effects of product line breadth: Evidence from the automotive industry. *Manufacturing Service Operations Management* Forthcoming.

Muchnik, L., S. Aral, S. J. Taylor. 2013. Social influence bias: A randomized experiment. *Science* **341**(6146) 647–651.

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.

Oestreicher-Singer, G., A. Sundararajan. 2012. Recommendation networks and the long tail of electronic commerce. *MIS Quarterly* **36**(1) 65–83.

Payne, J.W., J.R. Bettman, E.J. Johnson. 1993. *The adaptive decision maker*. Cambridge University Press.

Rajagopalan, S. 2013. Impact of variety and distribution system characteristics on inventory levels at us retailers. *Manufacturing & Service Operations Management* **15**(2) 191–204.

Raju, P.S. 1980. Optimum stimulation level: Its relationship to personality, demographics, and exploratory behavior. *Journal of Consumer Research* **7** 272–282.

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.

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

Schwartz, B. 2004. *The paradox of choice: Why More is Less*. Harper Collins.

Sela, A., J. Berger, W. Liu. 2009. Variety, vice, and virtue: How assortment size influences option choice. *Journal of Consumer Research* **35**(6) 941–951.

Simonson, I. 1990. The effect of purchase quantity and timing on variety-seeking behavior. *Journal of Marketing Research* **27**(2).

Staiger, D., J.H. Stock. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**(3) 557–586.

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

Susarla, A., J.H. Oh, Y. Tan. 2012. Social networks and the diffusion of user-generated content: Evidence from youtube. *Information Systems Research* **23**(1) 23–41.

Trivedi, M., F.M. Bass, R.C. Rao. 1994. A model of stochastic variety-seeking. *Marketing Science* **13**(3) 274–297.

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

Variety. 2013. Redbox now controls more than 50Tech. rep.

Waldfogel, J. 2012. Copyright protection, technological change, and the quality of new products: Evidence from recorded music since napster. *Journal of Law and Economics,* **55**(4) 715–740.

Wooldridge, J.M. 2002. *Econometric analysis of cross section and panel data*. The MIT press.

Xu, L., J. Chen, A. Whinston. 2012. Effects of the presence of organic listing in search advertising. *Information Systems Research* **23**(4) 1284–1302.

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

Zentner, A., M.D. Smith, C. Kaya. 2013. How video rental patterns change as consumers move online. *Management Science* **59**(11) 2622–2634.

Zhou, W., W. Duan. 2012. Online user reviews, product variety, and the long tail: An empirical investigation on online software downloads. *Electronic Commerce Research and Applications* **11**(3) 275–289.