

Is the Niche Culture Rising? An Empirical Study on the Internet Software Market¹

Wenqi Zhou

Department of Information Systems & Technology Management, School of Business,

The George Washington University, Washington D.C. 20052,

wenqi@gwu.edu

Sep. 30th 2008

First Year Summer paper

Presented to GWSB Doctoral Committee

¹ I thank Prof. Wenjing Duan for kindly sharing the data. I am also very grateful to her valuable guidance and suggestions.

ABSTRACT

The prevalent prediction of online consumer preferences is represented as “superstar” effect, an extension from the research on traditional market, which illustrates a scenario in which the Internet makes it easier for consumers to discover and follow other consumers’ choices, thus making them flock to the most popular products. However, the lower search costs and higher variety of products on the Internet can also enable consumers to find unpopular products as their favorites, referred to the idea of “long tail”. This study examines the shift of consumer preferences in the context of online software adoptions. We first identify the coexistence of a steeper head and a longer but flatter tail by comparing download distributions over two periods through various descriptive statistics and Pareto distribution analysis. We then take a further step to empirically investigate the causes underlying this shift via quantile regression model. We find significant interaction effect between user reviews and product variety on users’ adoptions. In terms of product variety, the increase in product variety would lead to increase in adoption of products with neutral ratings. This impact is stronger with respect to unpopular products than popular ones, which is consistent with the emerging long tail literature. We also find that products with extreme user ratings would obtain more increase in adoptions than those with neutral ratings, resulted from the corresponding increase of product variety. In terms of online user review, its impact on user adoptions is not uniform across all levels of ratings in markets with high product variety. Changes in neutral user ratings lead to more changes in online users’ software adoption of popular products regardless of level of product variety. As to extreme reviews, we find there exists a threshold of weekly product variety for each quantile. The results suggest that when all product varieties exceeding their corresponding thresholds of that quantile, in contrast to neutral ratings, the increase of positive user ratings has more positive impact on user choices whereas the positive impact of negative user ratings is less significant. The interaction effect suggests that consumers can better match their preferences with software products by referring to extreme user reviews, especially when facing a large pool of potential choices. Our findings offer an explanation for the mixed results in prior long tail studies with regard to the influence of online word of mouth and product variety on consumer preferences.

Keywords: Long tail, online user review, product variety, word of mouth, online software market

1. INTRODUCTION

Many researchers have discussed such “superstar” effect through both theoretical and empirical analysis in various contexts of experience goods [20, 31], which also has been illustrated on the Internet market [15]. A prominent promotion in brick-and-mortar stores is to highlight the hits, which is supported by the well known Pareto principle applied in sales distribution, namely 80/20 rule. This rule illustrates that the cumulative sales from the most popular products (less than 20 percentiles) account for approximately more than 80% of total. Many researches have discussed this concentration of consumer preferences on a very few top products, which is also termed as “superstar” effect, through both theoretical and empirical analysis in various contexts of experience goods [20]. Rosen [31] pointed out that the pattern of consumption will be more homogenized to a very small number of popular products because of their lower distribution and transaction costs. Moreover, Frank & Cook [20] showed that there exists a winner-take-all market in the context of traditional retailer. In essence, the Superstar effect represents the scenario that the very popular products would attract more and more adoptions over time and thus have the richer-get-richer effect.

Naturally this homogenized prediction of user choices is extended to apply in the online market and gets support from some empirical works [15]. Yet the Internet and other digital channels also enable more underdog products to be discovered. Anderson [2] firstly developed the idea of “long tail” as those unpopular products sold only online but not in brick-and-mortar stores, which is also referred to the tail part in the sales distribution. It is predicted that the cumulative sales from tail products online will overweigh the sales in blockbuster’s market. Anderson [2] provides the empirical

supports by conducting a descriptive analysis on Amazon book sales data. The sales from products unavailable in offline retail stores account for more than 50% of all Amazon book sales. This long tail concept has been extended by Elberse & Oberholzer-Gee [16] to include all the underdogs in both online & offline markets. The “long tail” phenomenon has attracted attentions from both practitioners and academics. On the one hand, some corporations begin to consider capturing the niche market. Eric Schmidt [34], Google's CEO, referenced long tail in his statement about Google’s mission as serving “*individual contributor, the small business, the company where Joe or Bob is the CEO, the CIO, the CFO and the worker and the support person--a one person company, a two-person company, a three-person company*” as well as very large customers. On the other hand, there is doubt on whether it is practical to build the marketing strategy according to the long tail prediction and how to predict and interpret the shift of online consumer preferences. For example, Anderson and Elberse are arguing that it is the heterogeneity of consumption pattern that has been exaggerated as long tail [3, 17, 18]. The traditional superstar proponents believed that customers will stick to the bestsellers if available without considering the poor substitute-unpopular products [20, 32]. McPhee [30] also points out consumers are even less loyal to their unpopular cultural purchases. Being consistent with superstar, Zhao *et al.* [37] point out online word of mouth (WOM) restrains the formation of long tail in the context of online book market. In contrast to the above conclusions, Brynjolfsson *et al.* [8] find that the internet consumers are more likely to purchase niche products in the book industry in support of long tail prediction. To shed light on whether online consumer preferences changes according to the long tail prediction, this research is designed to empirically examine the shift of consumer

preferences in the context of online software market. The Internet has been hailed to provide more choices of products and reduce consumer search cost of finding price and product information. On the supply side, the product variety is enhanced greatly in the online market due to the lower inventory cost and less constraint to physical shelves compared to the brick-and-mortar stores. For example, number of book titles sold in Amazon.com is 23 times larger than those on shelves of a typical brick-and-mortar store-Barnes & Nobel [7]. Also in Online software market, the number of competing products in each product category has increased greatly as a result of software programs' digitalization and thus has enabled virtually unlimited shelf space of online retailers. On the demand side, the search costs for both price information and product information are much lower facilitated by search engines, i.e. Froogle, and online recommendation system. For instance, many e-commerce websites provide online feedback systems to allow users to share their evaluations in both detailed comments and overall ratings. Amazon.com provides a score distribution to summarize the user ratings as well as posting all user reviews for each products. Beside users' evaluations, some of software products listed at CNETD (download.com) also received expert opinions as CNET ratings by CNET professional editors. Anderson [2] predicts that increased product variety and online WOM information would make consumer choices better match their preferences. Thus consumers are more likely to purchase hard-to-find products which are not usually available in brick-and-mortar stores. This trend of consumers' choices would result in sales distribution shifting to tail part, as shown in Figure 1.

Insert Figure 1 about here

Most of extant research on long tail phenomenon either focused on search costs or product variety and obtaining mixed results. Our study, nevertheless, is particularly interested in investigating the impact of both online product information and product variety on consumer preferences. For example, Brynjolfsson *et al.* [7], Gouville & Soman [23] discuss how the product variety would change consumer choices while more researcher put their focus on examining the influences of WOM or recommendation system on user adoptions [9, 14, 19, 25, 26, 35, 37]. This study models consumer behavior involved with factors from both sides by mainly examining how an interaction term between product variety and online user reviews plays a role in the shift of user choices. The significance of this term suggests that the relationship between product variety and software adoptions differs by level of online user reviews. In addition, in the markets with different levels of product varieties, the impact from user reviews is expected to be correspondingly different.

This study also differentiates the impacts of both product variety and online user reviews towards user choices of products with different popularities levels. Via adopting quantile regression method to better gauge to our skewed sample distribution, our model is able to provide an individual marginal effect of either product variety or online user reviews for each quantile. Quantile regression model is not widely applied in information system fields, which is firstly extended from OLS regression by Koeneker & Bassett [28] to model the effects of covariates on the conditional quantiles. Recently Elberse & Oberholzer-Gee [16] successfully applied this model in the context of U.S. movie sales and identified a coexistence of long tail and superstar. Our study takes a step further to look into how product variety and online user reviews influence user adoptions of

products for each quantile. Thus the results provide a broader picture to cover products with all levels of popularities other than only focus on top products and underdogs, which is the way the most extant research conducted in [9, 35].

The descriptive results and Pareto distribution analysis consistently demonstrate the co-existence of the long tail and superstar effect. This result is consistent with some extant researchs. For example, Elberse & Oberholzer-Gee [16] found that ever-smaller number of titles account for the hit sales besides observing a longer tail over time in the context of U.S. home video industry during the period from 2000 to 2005. Also Tucker & Zhang [35] conducted a field experiment on a website listing wedding service vendors to show that superstar products complement instead of competing the tail part by attracting new customers who would choose nothing at all. In addition, the empirical results from quantile regression model show that impact from either online user reviews or product variety on software adoptions varies with respect to different levels of user ratings. Moreover, the influence from online user reviews depends on the level of product variety. Increase in neutral user ratings leads to the increase of popular product adoptions regardless of level of product variety. But extreme user reviews' impact is more intricate. In a market with very high product varieties for all quantiles, extreme user reviews have a positive influence on user choices. The impact from positive reviews is stronger than neutral ratings'. In other words, if online consumers face only a few choices, extreme user reviews may even exhibit a negative influence on their choices. This implies that extreme user reviews would exhibit very different influences on user adoptions in markets with different levels of product variety. As to the impact of product variety, for products with neutral ratings, we find increase in product variety leads to the significant

increase of product adoption. The impact is stronger for unpopular products, which is consistent with the long tail prediction. Such impact pattern, however, does not hold for product with extreme user ratings. Increase in product variety indeed leads to more adoptions and such impact is even stronger than that those with neural ratings. Nonetheless, it's not clear how such changes influence the consumer adoption distribution.

This paper contributes to the emerging long tail literature by offering an examination on shift of online user adoptions by comparing various descriptive statistics and Pareto shape indices over times, and conducting rigorous quantile regression analysis. To our knowledge, it's the first empirical research that examines the impact of the interaction between online WOM and product variety on long tail formation. Moreover, this study contributes to the field of information system (IS) and marketing research by offering a new perspective on the impacts of WOM effect and product variety on user choices in online market, which are widely discussed as the important influencing factors of consumer behaviors and received mixed results. Our results suggest that these two factors interact with each other to exhibit the influence on consumer choices in a more complicated way than what has been revealed before. How the different level of user ratings influences use choices depends on the corresponding level of product variety; and the magnitude of impact from increased product variety varies along with the different levels of user ratings. The existence of thresholds provides an explanation to reconcile the mixed results about the impact of online user reviews. In a market with all product varieties exceeding corresponding thresholds, user reviews with all levels have positive impact on user adoptions. Otherwise, their impacts would depend on the level of ratings.

Moreover, the level of threshold may vary from market to market, which has important implications in terms of building a corresponding marketing strategy in various industries.

This paper also contributes to the guideline of assortment strategy by pointing out the necessity to complement inventory level with online store's marketing strategy.

Traditionally, more emphasis has been put on top sellers, such as highlighting Top-Ten products and more storage space; however, there is not enough attention on unpopular products. Until recently, some firms begin to consider whether and how to adjust their strategies in order to capture this potential emerging niche market and a new paradigm of making profit. For example, Google has already been working on building "*a whole bunch of small, self-service tools*" which allowed their tail consumers to "*almost automatically use this service*". It tries to "*capture very large and historically underserved businesses as well as a whole new area that never had access to these kinds of online services*" [34]. However, our results suggest that the online retailers should be very careful to determine product variety level with respect to their marketing strategy. For example, in an online store where most of user ratings are around middle level, if the store is intended to target on niche market then a high level of inventory is recommended. It may turn out to be inappropriate to capture niche market in the case of extreme user ratings. Specifically how many products should be kept depends on the situation of online user reviews. In sum, each store should choose its target market by exploring current situation of online user reviews and correspondingly determine its assortment strategy. The rest of paper proceeds as follows. Theoretical background and research hypotheses are described in section 2 proceeded by data description and methods in section 3.

Section 4 discusses our empirical model and results, which is followed by the conclusion of this paper and areas of future research.

2. THEORETICAL BACKGROUND AND RESERACH HYPOTHESES

2.1 Research Context

We choose to examine the change of customer preferences in the context of online software industry, specifically using data collected from CNET Download.com (CNETD) for several reasons. First, CNETD, as a part of CNET network, is a leading and representative online platform for software downloads without any charge. CNETD naturally provides all the information we needed in this study. In CNETD, there are over 300,000 free or free-to-try software programs for Windows, Macintosh, and Handheld devices available. For each software post on this website, CNETD updates the cumulative number of downloads and the number of downloads for the most recent week every day. The available information about user choices make our study capable to specify the discussion of consumer preference regarding different popularity levels of products, which is the essence of the debate between long tail and superstar. The Figure 2 depicts what user encounter on CNETD. Since some of the software programs posted are free-to-try, it's reasonable to extend our conclusion carefully to some commercial software retailers.

Insert Figure 2 about here

Second, CNET solicits reliable user reviews for all programs posted, representing the realizations of online WOM effects. Moreover, online reviews as the true evaluations from prior users at CNET are more reliable than the recommendations from online commercial software program retailers. Some researcher has already pointed out that the

online reviews may get biased and manipulated by commercial online stores for the sake of profits. The retailer may intentionally employ experts to copy the good comments for the tail goods or to relate them with popular products in reviews. This may result in a temporarily fake 'long tail', which is totally far away from our research question. Yet neither CNET nor any software corporation directly gets any profit from the increasing product adoptions. In this case, CNETD data become a good source to figure out the true consumer preferences behind their choices in an online market by their decision of downloading.

Finally, our specific context can naturally fix price effects by providing free downloading without any charge, which makes us only need to focus on the low search cost caused by WOM communication. More potential external factors excluded, more reliable conclusions we can obtain. In CNET all customers can download whatever software they are interested in without being charged any fees. Therefore the price impact is controlled by default and there is no need to pay attention on search cost about price information. As discussed in more details in the following hypotheses part, fixed price effect narrows down this study to simply focus on the change in search cost of product information indicated by WOM effects.

2.2 Literatures and Research Hypotheses

The two basic presumptions for the emerging long tail phenomenon are increased product variety and the lower search cost, which are the major differences between e-commerce shopping environment and brick-and-mortar stores. Brynjolfsson *et al.* [7, 9] constructed two related studies to show how these characteristics of online market significantly benefit consumers. On supply side, the increased product variety greatly enhance

consumer surplus, even much more than consumer surplus obtained from increased competition and lower prices; On the demand side, the lower search cost of both price information and product information in online channel makes book sales distribution become less concentrated. Both results imply that consumers are more likely to adopt underdogs in online market, which is supportive to long tail prediction.

In our study, as mentioned above, the price impact is controlled by default because CNETD provides downloading without any charge, which makes us only need to focus on the lower search cost caused by WOM communication. To be specific, it mainly includes online user reviews. Thus our hypotheses are structured via these two aspects: online user reviews and product variety with the objective to examine the mechanism underlying the change of consumer preferences.

2.2.1 Online user reviews

Intuitively, the more product information online retailers provide, the better consumers could match choices with their tastes, which is also confirmed by some researchers [11, 33]. For example, in the context of online book market, Chevalier & Mayzlin [11] find the significant and positive relationship between online reviews and book sales.

However, there are different opinions about how these online reviews influence consumer preferences and to what degree. Van den Bulte and Lilien [36] show that the WOM influence on tetracycline adoption was overestimated. Moreover, some previous studies also show that users are more sensitive to the extreme user reviews, namely positive and negative ratings. Clemons *et al.* [12] confirm the resonance marketing theory and point out that new products would grow fastest in the marketplace with the help from variance of ratings and the strength of the most positive quartile of reviews. Chevalier and

Mayzlin [10] point out that the 5-star reviews improve sales whereas 1-star reviews hurt sales. The impact from 1-star reviews on user choices is even greater than the impact of 5-star reviews. Their results imply that it's reasonable to expect different impacts on user choices from ratings with different levels. The positive scores, which are higher than the middle level 3-star reviews, are more powerful in persuading users to try the relative products than the neutral 3-star ones; users would also get strongly impressed by the extreme low scores and make a non-negative prediction about product quality. In other words, this indicates that the influence of user ratings on user choices should have better-get-better effect.

Besides, according to Anderson's [2] long tail opinion, the benefits from WOM for experience goods are not uniform among products with different popularities. Bakos [5] also shows that WOM recommendations would help consumers to match their preferences and find more unpopular goods. Recently, Duan *et al.* [15] applied informational cascade theory to show that online user reviews have different impacts on products with different popularities in the context of software program, which is consistent with the emerging phenomenon of long tail consumption pattern. Their results point out that online user ratings have positive influence on products except the most popular one. Hervas-Drane [26] shows that the recommendation system would function as a tasting matching mechanism to help consumers get product information from others with similar preferences and thus reduce sales concentration. Clemons [13] points out that the change in demand side would make firms earn more than expected in the "new created resonance marketing sweet spots" whereas earn less in the "traditional mass-market fat spots", which seems supportive for the positive influence from WOM on user

adoptions of tail products. However, there are also different voices regarding the influence of online user reviews on user adoptions of top products or underdogs. Fleder & Hosanagar [19] point out that some common recommenders lead to a net reduction in average sales diversity. Dellarocas & Narayan [14] employed Dicher's four main motivational categories of offline WOM communication and combined with the Internet unique properties to prove that propensity to review is positively correlated with a movie's online popularity, which seems supportive to superstar idea. Ghose & Gu [25] also empirically shows the search cost are lower for popular products compared with underdogs.

Therefore the current inconsistent results might be caused by the failure to differentiate the impacts from the different level of user scores and different products popularities. For example, the influences of positive reviews should vary from negative ones towards products with different popularities. Zhao *et al.* [37] take a micro approach to show that the influence from WOM not only differs significantly on the sales of hit and tail products but also depend on whether the review is positive and negative, in support of superstar theories. In our study, software program is one typical format of experience goods with the attribute that users have to deal with significant uncertainty before trying out. In this case, the useful online user reviews provided by recommendation system would greatly reduce their risks. Yet users are hard to get any strong impression and much product information about their potential choices from their neutral ratings. Therefore, the increase of user ratings could change users' impressions to positive ones. Users receive this positive signal and are more likely to be interested in trying it out, which cumulatively results in more adoptions of corresponding products. According to

our above analysis, this positive influence should vary towards popular products and underdogs. Generally, the popular products have already obtained better reputations denoted by high rankings and a much larger volume of user reviews than unpopular products. While a larger volume of user reviews would amplify this positive signal, it also seems make users more confident to rely on picking one from popular groups given the same increase in neutral ratings. Thus, overall it's expected that the popular product could get a better-get-better effect resulted by an increase of neutral user ratings.

Therefore, we propose that:

H1: Given all products with neutral ratings, an increase in user rating would result in more product adoptions. This impact is stronger with respect to popular products than less popular ones.

Following the similar argument, if the user ratings are above the middle level, namely positive reviews, the increase would not only provide consumers a positive impression about the product quality but also strengthen the already existed positive signal by the original ratings. Thus the positive scores combined with its increase provide users a stronger impression about product than neutral scores do.

H2: Given all products with positive ratings, an increase in user rating would result in more product adoptions. For products with the same popularity, this influence is significantly stronger than influence from an increase in neutral user ratings.

Regarding to the negative reviews, its one unit increase still makes the final user rating a non-positive one, which would probably give users some better feelings of the product characteristics. Yet this benefit is not as strong as the positive impact from the one unit increase in non-negative reviews in terms of believing them as favorites. In other words,

the potential impact directly from the increase can not totally overweighs the non-positive impression caused by the original negative rating, which resulted in the less increases in product adoptions as compare with above two cases. Therefore, we propose that:

H3: Given all products with negative ratings, an increase in user rating would result in more product adoptions, but this influence is less powerful than non-negative rating's.

2.2.2 Product variety

Internet retailers have virtual inventory instead of the physical space constraints and the holding cost. It is much less costly for the Internet retailers to add more variety of goods and to manage the product library more effectively than the traditional sale channels [7]. Especially for the digitized products, like software program in our study, adding new copies results in almost zero inventory fee, which makes the online retailers capable to keep a much higher product variety than ever before.

However, there exist mixed results about the influence of product variety on consumer choices. While the proponents of long tail argues that the increased product variety can provide a larger pool of candidates for online users and thus make them better choosing products to match their diversified preferences, some researchers has questioned this predicition. Gourville & Soman [23] use “overchoice” to explain that overload product variety can make consumers too confused to choose nothing at all.

Either argument might turn true under certain situation. The previous studies don't put enough emphasis to differentiate the impacts from this factor towards products with various levels of popularities. The long tail theory argues that increased product variety would give users more chances to find their favorites other than simply following the prior blockbuster's choices of popular products, which in turn exhibits a trend to have

more adoptions of unpopular products. The potential explanation to converge the above related but mixed results is to consider the interactive impact between product variety and user reviews. Imagine following two cases for online users both facing a great number of products: 1) users with only little product information provided, especially other consumers' opinions; 2) users are able to get enough product information from the useful online user reviews. Obviously, users in the second situation can make their decisions much more easily to find out their favorites. The users in first situation are easily to get lost during the process to choose one from these too many potential candidates. The worst situation is that consumers may even give up trying and end up with choosing nothing at all according to the results of Gourville & Soman [23]. Then what kind of user reviews can be considered as useful? As mentioned above, we believe that users are more sensitive to the extreme scores of user reviews. In other words, the extreme ratings are more useful than neutral ones given that the extreme reviews would provide more product information by directly emphasizing the sparkling points or shortcomings of product qualities. The above reasoning points out that the influence from product variety depends on the different levels of user ratings.

Thus to discuss this interaction between product variety and user reviews, we firstly look into the products with neutral ratings. The long tail theory demonstrates that increased product variety would give users more chances to notice and then find their favorites other than simply following the prior blockbuster's choices of popular products. In this case, increasing product variety functions as a long tail factor to make the adoption mass shift to the tail part. Specifically, the impact from increase in product variety brings more adoptions to less popular products than to more popular ones. Thus we propose that:

H4: Given all products with neutral user ratings, the increase in product variety leads to more product adoptions. This impact is stronger with respect to unpopular products than more popular ones.

Second, if considering the products with the same popularity, as indicated in discussion about user reviews part, the increase from extreme user ratings makes users more impressed about product quality than the increase from neutral ones. In other words, the extreme user ratings help users to better find their favorites facing a large number of potential choices and then reinforce the positive influence from product variety. Therefore, we propose that:

H5: Given all product with the same popularity, the increase in product variety leads to more product adoptions. This impact is stronger with respect to products with extreme user ratings than products with neutral user ratings.

3. DATA

3.1 Data Collection Methodology

We collect data from CNET Download.com (CNETD: www.download.com), part of the CNET networks. CNETD lists more than 20 groups of software program for users to download free of charge while each group has 8~20 categories. For each software program posted, besides the product descriptions and operating system requirements, CNETD also provides user reviews, rankings, license to show free or free to try and cumulative downloads. The user review system includes detailed comments and an overall evaluation indicated by a five-star user rating system, with *one* being the worst and *five* the best.

Our sample consists of three software categories which provide a diversified coverage with different application purposes. The categories we choose are listed below: Digital Media Player, Download Manager and MP3 Finder. We collect data at the end of every week during two periods: from Dec. 2004~Jul.2005 (period 1), from Aug. 2007 to Feb.2008 (period 2). Everyday we extracted following data on every software program from CNETD as software names, updated dates, total download, last week download, categories, average user ratings, CNET ratings, identifications of free or free-to-try and software characteristics including operating system requirements, file size, publisher, license and price if its license is free-to-try.

The CNETD data has several advantages in terms of methodology. Usually it's not easy to obtain sales data from the online commercial retailers. Thus in order to capture user adoptions, one alternative is to use estimated sales quantity inferred from rankings. This technique was reported to have difficulty with getting robust estimator for the mapping function either from expert opinions or experiments. In software industry the expert opinions about this mapping estimators hardly help in the sense that only a very few studies has worked on this mapping inference [24, 8]. Similarly, the experiment method seems unsuitable to our research question either. Only one estimator can be provided by one experiment at one time. Thus to compare the change of estimators over time, we will need a series of estimators by either previous studies or experiments replicated for each time lag in the examined period. The other alternative is to use the public U.S. government data, which also has the disadvantage to underestimate the sales from small sellers [4]. Instead, CNETD naturally provides the consumers' downloading decisions in public, which makes it a good sample to be analyzed. In CNETD, there are over 300,000

free or free-to-try software programs for Windows, Macintosh, and Handheld devices available. For each software post on this website, CNETD updates the cumulative number of downloads and the number of downloads for the most recent week every day. The available information about user choices make our study capable to specify the discussion of consumer preference regarding different popularity levels of products, which is the essence of the debate between long tail and superstar.

3.2 Key Variable

In our data collection process, weekly download counts of each software program are captured to indicate the customer choices, which is the key dependent variable in our study. One of the advantages to use weekly data is to prohibit weekend effects. The other advantage is to rule out the random effects often caused by using daily data.

Our major independent variables include online average user ratings, CNET ratings, cumulative downloads and weekly product variety. We capture the user scores at the end of every week to treat as average weekly ratings. For CNET ratings, the editorial staff reviews a fraction of the software programs (less than 20% shown in Table 1) with an emphasis on popular software programs and gives the score by the similar five-star rating system as user ratings. Total downloads for each software program is also recorded to fix the net work effect, which is explained in more detail in section 4. In addition, we calculate the number of all different software programs posted each week as the weekly product variety. More details on variables and model specification will be discussed in section 4.

Insert Table 1 about here

3.3 Descriptive Analysis of User Choices' Shift

3.3.1 Market definition

In our particular context, unlike some other forms of experience goods (i.e. book & video), each category refers to a different market, which makes it necessary to make separate analysis in each category. Users with different seeking purposes would end up choosing software programs in totally different categories via different approaches. For example, users, who need to find one MP3 Finder software program, may give more weights on the existed user base than the information provided by previous users. Yet users interested in Digital Media Player category would be more likely to benefit a lot from recommendation system. Thus we start our analysis by generating various descriptive statistics in each of three categories.

3.3.2 Descriptive Analysis of User Choices' Shift

We choose the month period as time lag so that our measures would be possibly sensitive enough to examine the accumulated changes of consumers' choices and the fluctuation resulted by random effects will not be mistaken as the true change. The traditional examination for 80/20 rule has been extended to calculate percentages of monthly downloads accounted by a series of quantiles for total monthly titles, namely from 1% to 99%. Also, we use several forms of position indicators – i.e., $Q\alpha$, to reveal more about the shift with respect to head and tail, like scale, skewness and kurtosis. $Q\alpha$ denotes the number of software programs in the lower α quantile of monthly total downloads. For example, Q_{10} means how many software programs constitute the lower 10% of the total monthly downloads. We use $(Q_{0.5}-Q_{0.25})/Q_{0.5}$ and $(Q_{0.75}-Q_{0.5})/Q_{0.5}$ to capture how left tail and right tail shift with respect to the median.

After that we choose to make comparisons over two collection periods in order to allow for a large variation detectable to extract the exact shift of consumer pattern. We don't find any clear trend of user choices over months for each category², which may be caused by the mixed seasonal factor and some other unmeasured sources. The two periods are comparable partly because they includes similar time lag: period 1 has eight months while period 2 has seven month. And also, during these two periods, there is no fundamental change about the design of CNETD's website including downloading interface, user ratings, CNET ratings, search options, promotions and so on. As shown in Table 2, almost each monthly average percentage of software downloads accounted by each level of quantile is decreased in period 2 compared with period 1. For example, for category of Digital Media Player, the percentage of downloads accounted by 99 quantile of total titles is decreased by 16.82 % during more recent period than the old period; the percentage for 50 quantile is decreased by 82.32%; the measure for 25 quantile is also decreased up to 82.14%. These results imply the coexistence of a steeper head and a longer but flatter tail during more recent period. This pattern also holds for other two categories, which is consistent with the studies conducted by Tucker & Zhang [35] and Elberse & Oberholzer-Gee [16].

Insert Table 2 about here

The average monthly results shown in Table 3 about inter-quantile positions also confirm this pattern. For each category, left tail, right tail, skewness, scale, kurtosis all decrease during the period 2, which is consistent with the above statement of a steeper head and a

² The results about distribution of user choices and summaries about inter-quantile positions for each month are available upon request.

flatter tail. Especially for category of Mp3 Finder, the degree how its head turns steep to is even much larger than the other two categories during both two periods in the sense that during some months in both two periods³, the quantile 75% position doesn't exist because 90% quantile position is only one title more than 50% quantile position. Yet in terms of the degree how the steep head changed, category of Download Manager changes the most. The head titles even decreases along with the increased total titles, which indicates a much less percentage of total products constituting the higher fifty quantile of user choices than the other two products. Besides, the tail becomes much flatter during period 2 in category of Digital Media player than the other two categories. The inter-quantile index is decreased along with the increased total titles, which indicates more weight on the tail part. Meanwhile the existence of above slightly different pattern in each category also confirms that each category should be considered as an individual market.

Insert Table 3 about here

In addition, we note that the variety of software programs posted during the period 2 in category of MP3 Finder is less than the variety during period 1. For example, number of software programs in category of MP3 Finder during period1 has 37.15% times more than period 2. Yet generally, it almost becomes a widely accepted trend that consumers are provided more potential choices over time just like the cases in other two categories because of the Internet's reach in online shopping websites. Many studies also consider this improvement as one of the major reasons to make consumption pattern changed, like the debate between "overchoice" and "long tail". However, our specific context shows

³ The report about position summary for each month is available upon request.

that, CNETD didn't always follow this routine. The old period even have more software programs posted in some category. When mostly websites are obsessed with providing consumers more choices, why CNET is still providing less for this category? The strategy of CNET to control the quality of posted software programs by several requirements⁴ may result in this decreasing trend in product variety. Taking a step further, this result makes our study more interesting to justify CNET's assortment strategy by examining the impact from product variety.

3.4 Descriptive Analysis of WOM Effects

We further generate a monthly average summary of WOM effects in terms of user ratings, volume and CNET ratings with respect to different quantiles of total listings. As shown in Table 4 below, for all three categories, the lower quantile products don't necessarily have lower average user ratings than the higher quantile products do.

Obviously this phenomenon is consistent with the fact that there exist so many mixed results about the relationship from user ratings and user choices, which is one of our interests of this study.

The analysis results in section 3.3 seem as a reconciliation of long tail phenomena and superstar effect during period 2, consistent with the previous studies. However, when we look into the detailed summary of product variety and online user review, it turns to bring a question—what factors result in this trend of user choices? Which combination of product variety and WOM would be able to contribute to this change? The long tail trend is uncovered in all three categories with different situations of changes in both product

⁴ The detailed software policy is available in the following link:
http://www.upload.com/1200-21_5-750430.html

variety and WOM, which indicates that the underlying reasons for user choices' shift in online software market is way more complicated than what has been explored.

Insert Table 4 about here

4. EMPIRICAL METHODOLOGY AND RESULTS

4.1 Analysis on Shift Pattern

Beside descriptive analysis, we also examine the shifts of head part in consumption pattern statistically by comparing the shaping index of Pareto distribution applied on relationship between download counts and product ranks over two periods. Pareto distribution is firstly used to describe the location of wealth and has been extended to more applications [10, 24, 29]. Ghose & Sundararajan [24] inferred sales of software program at Amazon from their ranks with the underlying Pareto distribution. It's natural to consider that the relationship between download counts and ranks would also follow Pareto definition. Specifically, in our study, the user adoptions are public and captured by download counts for each software program. Therefore, we don't need to obtain the Pareto shape index from extra experiments or previous estimators in order to map from ranks to approximate magnitude of user behaviors, which is widely adopted by previous studies [10, 22, 24, 29]. Instead we can utilized and compare the true shape estimators from applying Pareto distribution on known download counts and rankings to indicate if the distribution changed over time. The following equation, which denotes the probability of a software program with download counts D more than counts d , can be obtained after applying the Pareto definition in our case.

$$\Pr (D>d) = (dm/d)^k \tag{1}$$

, where dm and k are constants

From statistic characteristics of Pareto distribution, it's easy to know that K is the shape index. In other words, the greater K denotes the steeper PDF curve as show in Figure 3, where the horizontal axis is download counts and the vertical axis is rank/total titles. If we exchange the horizontal axis with vertical axis and vice versa, the smaller K would denote the steeper PDF curve. In addition, if the software with download counts equal to D locates at ranking denoted by R , the number of software programs whose download counts exceed D is just $R-1$. Thus the probability of a software with download counts D more than counts d is $(R-1)/\text{total titles}$. The for software i at time t we can get

$$R_{it}/(\text{total titles})_t^{=}=(R_{it}-1)/(\text{total titles})_t=\Pr (D_{it}>d_t)=(dm_t/d_t)^{kt} \tag{2}$$

 Insert Figure 3 about here

After using natural log on both sides of equation (2), we get our regression equation as

$$\ln D_{it} = C_t - \alpha_t \cdot \ln(\text{RANK}_{it}) \tag{3}$$

,where $C_t = \ln(dm_t) + 1/k_t \cdot \ln(\text{total titles})_t$ is a constant, $\alpha_t = 1/k_t$.

Thus, another alternative to capture the head shifts of download counts other than descriptive summary is to test whether the shape index α_t of Pareto distribution in above equation (3) changes over time. If the estimator α_t increase over times, the distribution curve of download counts becomes steeper. We run the regression model on our weekly CNETD data in each category for each month correspondingly during two periods in order to reflect more about the monthly changes statistically.

It's not surprising to note that the error terms are heteroskedastic in the sense that the observations of downloads with low ranks (high value of ranks) would fluctuate more

than the observations with high ranks (low value of ranks). In other words, the unpopular products represented by low rankings would have a large variance than the superstar products, which is also confirmed by White's test conducted⁵. Therefore these OLS estimators are biased and inconsistent. Because the format of error terms is unknown, as recommended in Econometric theory, Generalized Method of Moments (GMM) estimation is applied to get consistent estimator for the monthly shape index α_t in each category. The R-Square measures for each category show that the aggregated data fit log-linear regression model very well.

From the results⁶ of monthly estimators for each category, we can't see clear trends in both periods. That's because monthly time lag is too short to extract a complicated and fluctuational pattern of user choices over time. However, as shown in Table 5 when we look into the average monthly estimators of two periods, for each category, the coefficients have larger absolute values during period 2. For example, for category of MP3 Finder, the absolute value of average estimator is higher by 25.5% during more recent period than the ones in period 1. This implies that PDF curve is steeper in period 2, which provides a statistical evidence to support our descriptive results that period 2 have a steeper head on average. In other words, the consumer preference of popular products is more homogenized over time.

Insert Table 5 about here

4.2 Quantile Analysis

⁵ The report about White's test is available to upon request.

⁶ The report of estimators for each month is available upon request.

Although the above analysis provides some insights about the shift direction of download counts, the underlying reasons about this change are still unknown. Thus the following more rigorous model is conducted to test our hypotheses.

Our major objective is to analyze the underlying causes for the different shifts in user choices of products with regard to different popularities in terms of online user reviews and product variety. Our focus is not how online user reviews and product variety influence the average effect of user choices in online market generally tested by the widely used Ordinary least-squares (OLS) regression models. OLS is only able to model the relationship between one or more covariates X and the mean of the response variable Y given $X = x$. However, in our case, this traditional method to simply get an estimation of mean effect from each influencing factor doesn't fit any more. Koeneker & Bassett [28] firstly extend OLS regression to model the effects of covariates on the conditional quantiles. After that, quantile regression has achieved success in ecology as a type of regression analysis; however, it has not received many attentions yet in fields of information system and e-commerce. Recently, Elberse & Oberholzer-Gee [16] applied this method in the context of U.S. movie sales to find out the coexistence of both long tail and steeper head, and demonstrated its advantages on analysis over skewed population compared with OLS regression.

Thus the quantile regression methodology fits our research question well to obtain differentiated impacts on adoptions of popular products and unpopular ones. Quantile regression model is particularly useful in the situation that the change rate in response variable, expressed by the regression coefficients, depends on the quantile. Hence we choose to conduct an in-depth exploration by a series of quantile regression models. By

comparing the estimated coefficients across all quantiles, we are able to reveal the trend of influences from the explanatory factors on user choices of products with different popularities. The general form of conditional quantile regression models is present as below

$$Q_{\beta}(y/x) = x' \gamma(\beta) \quad (4)$$

, where $Q_{\beta}(y/x)$ denotes the β th quantile of the distribution of y and x denotes the independent variables.

Specifically, we apply the natural log transformation on response variable-weekly downloads for the sake of reducing non-constant variance and also transforming its value to the comparable magnitude level as all the independent variables. Hence our quantile regression model is as following for i software at time t in each market.

$$\log(Q_{\beta_{it}}) = \gamma_{1t} + \sum \gamma_{jt} * x_{ijt} + \varepsilon_t \quad (5)$$

By aggregating the CNETD data together in each category during period 2 as mentioned above, we use a set of aggregated data to estimate equation (5) using quantile model defined as 5th, 10th, 20th, 30th, 40th, 50th, 60th, 70th, 80th, 90th, 95th and 99th.

Specifically, we use weekly download counts as dependent variable to capture the user choices unobserved directly. For different levels of quantiles, $Q_{\beta_{it}}$ denotes the weekly downloads of software i in the β th quantile of total titles at time t and γ_{1t} is a constant term. In addition, x_{ijt} denotes the value of j -th explanatory variables for software i at time t and ε_t is error term.

We also follow the previous research to use the total number of each software program's existed base to control for network effects notated as $TOTALDOWNLOAD_{i,t}$

(Brynjolfsson & Kemerer, 1996; Gallaughier & Wang, 2002; Duan et al., 2008). Network

effect refers to the idea that a product becomes more valuable as base expands (Katz & Shapiro 1994). In our context, it refers to more adoptions as a result of the increased user base. For example, peoples using MSN are better off when more customers choose to adopt this online chat tool. Although only online chat category is considered to be more vulnerable to network effects in context of software market [15], we also need to control it for all three categories in our study since network effect is the most common way to explain the herding behavior. Figure 4 shows that the most common popular options used for consumers to search any software program at CNETD is total downloads, which also implies network effect plays an important role in user adoptions at CNETD. Moreover, by adding two class variables (software-specific var. and time-specific var.) into the quantile model, we are able to control for the influences from intrinsic individual software characteristic and time differences and thus to correct the heterogeneity. Our model also includes product age and the quadratic term of product age used to control for product diffusion. Table 6 lists all the exploratory variables x_{ijt} with the detailed explanations.

Insert Figure 4 about here

Insert Table 7 about here

Our hypotheses predict that the impact from extreme user ratings on consumer choices is different from the neutral scores'. In our specific context, neutral ratings refers to 3-star and we define the extreme user rating as all the other rating levels other than 3-star. We use $USERRATINGR_{i,t}$ as the difference of user rating software i received at time t divided by the middle level point 3. This minor transformation of user rating can make

the relevant coefficients more intuitive to explain the influences from different levels of user reviews.

Moreover, in order to differentiate the marginal impacts from different levels of user ratings, we also include a square term of $USERRATINGR_{i,t}$ noted by

$USERRATINGRSQ_{i,t}$. Also in order to examine the interaction influence between product variety and extreme user rating, an interaction term is introduced as $WEEKLYVARIETY_{i,t} * USERRATINGRSQ_{i,t}$ where $WEEKLYVARIETY_{i,t}$ indicates total number of software program the category, which includes software i , has at time t .

Therefore the impact of user ratings on user choices can be expressed as $\beta_1 * USERRATINGR_{i,t} + \beta_2 * USERRATINGRSQ_{i,t} + \beta_3 * WEEKLYVARIETY_{i,t} * USERRATINGRSQ_{i,t}$. The first derivative for above expression shows that the marginal influence of change in user rating is $\beta_1 + (2\beta_2 + 2\beta_3 * WEEKLYVARIETY_{i,t}) * USERRATINGR_{i,t}$.

If the user ratings are equal to/above/below point 3, $USERRATINGR_{i,t-1}$ is zero/positive/negative. Therefore, to test H1, we expect β_1 to be significantly positive and getting a larger value along the increasing quantile. To test H2 & H3, we expect $2\beta_2 + 2\beta_3 * WEEKLYVARIETY_{i,t-1}$ is significantly positive.

We include the term $\beta_4 * WEEKLYVARIETY_{i,t} + \beta_3 * WEEKLYVARIETY_{i,t-1} * USERRATINGRSQ_{i,t}$ to represent the influence of product variety on user choices. The first derivative for this expression shows that the marginal influence of change in product variety is $\beta_4 + \beta_3 * USERRATINGRSQ_{i,t}$. To test H4, we expect that β_4 is significant and has the larger absolute value in lower quantile, which indicate a downward sloping trend across quantils. To test H5, we expect β_3 is significantly positive.

Our model also includes another dummy variable $\text{FREEPRICED}_{i,t}$ to examine if the option of free-to-try would have different impacts on user choices from the option of free. The product age $\text{AGE}_{i,t}$ and the quadratic term of product age $\text{AGESQ}_{i,t}$ are used to control for product diffusion.

4.3 Quantile Analysis Results

We estimate our empirical model using the aggregated weekly observations for each of three categories only during period 2 for two reasons. First, all hypotheses are predicted in the context of current market with high product variety and active WOM communication. Period 2 is the one period more recent and containing these market characteristics. Second, this study is designed to explain the coexistence of long tail and superstar, which exists in period 2 as summarized above. The quantile regression results are shown in process plot Figure 5a, 5b, 5c respectively for each category. The changes of influences across different quantiles are clearly portrayed. For each category, most of all estimators are significant across quantiles besides only a few exceptions whose confidence interval (grey area) include zero-value horizontal line. Overall our theoretical model fits the data well.

Insert Figure 5a, 5b, 5c about here

The coefficients on $\text{USERRATINGR}_{i,t}$ for each category, are significantly positive and keep increasing along the increasing quantiles. This implies that H1 is supported.

Adoptions of products received neutral scores are significantly influenced by user ratings and this influence is getting stronger with respect to more popular products.

The coefficients on product variety are all significant and have larger value on the lower quantiles and most of them are positive. The positive influence turns more powerful on

adoptions of unpopular product. Thus H4 is supported. This result implies that tail products receive more adoptions resulted by the increase of product variety than popular products do. This benefit for unpopular product may even happen at the expense of popular products in categories of Download Manager and MP3 Finder because their estimated values of β_4 even turn negative in the highest quantile, which is consistent with the “overchoice” claim from Gourville & Soman (2005). For these two categories, the popular products not only get less increased adoptions from the higher variety but also, would lose consumers as shown by the negative sign of the estimators in the highest quantile (99%). Moreover, the coefficients on interactive term for each category are significantly positive across all quantiles, which indicates H5 is supported. The products received extreme user ratings would benefit more from increased product variety.

In addition, H2&H3 are supported partially. The coefficients on $USERRATINGRSQ_{i,t}$ are significantly negative across all the quantiles, which, to be honest, is not exactly expected initially and actually turns to provide more insights to our research question. Given the different signs of β_2 and β_3 , there exists a threshold of weekly variety for each quantile as shown in table 7 to guarantee that the value of $(2\beta_2 + 2\beta_3 * WEEKLYVARIETY_{i,t})$ is positive. In other words, if the weekly variety exceeds the corresponding threshold for each quantile, which implies a market with high product variety, then H2&H3 are supported by the positive sign of $(2\beta_2 + 2\beta_3 * WEEKLYVARIETY_{i,t})$. In period 2, all the thresholds for all categories are beyond the current average product variety, which is in support of H2&H3 in current CNETD context. Meanwhile, given product variety is very high, $\beta_1 + (2\beta_2 + 2\beta_3 * WEEKLYVARIETY_{i,t}) * USERRATINGR_{i,t}$ has the possibility to be negative in case of

negative ratings, indicating that the increase of negative user review may even not result in more user choices until they reach neutral or positive rating. For example, in traditional market, when product variety is relatively low due to the limited storage space, the conclusion referred to extreme user review would be totally different. Specifically, the influence from increase of positive user reviews on user adoption becomes less powerful than the influence from neutral one whereas the increase of negative user reviews has more influence than neutral one's. Let's imagine the case that there are only 10 products available in the market, consumers are most likely to read all the reviews. Those with negative ratings would provide more useful information and thus are more likely to make consumers confident about matching these products with their preferences [11]. The existence of variety thresholds is able to justify the mixed result about online user reviews. We will examine in-depth about this special switching point-threshold in our future research.

Finally, we find the coefficients on ranking for each category are almost significantly negative and exhibit a slightly increasing trend across all quantiles. Consistently with previous studies [15], the drop in ranking (increase the value of ranking) will result in the decrease of user adoptions. And this influence almost keeps no big difference to products with various levels of popularities.

5. CONCLUSION AND DISSCUSSION

5.1 Conclusion

The objective of this paper is to demonstrate how the typical features in online shopping environment of experience goods influence user choices and thus contribute to changes of consumption pattern cumulatively. Our empirical analysis shows that the complicated

impacts from increased product variety and online user reviews may result in the different observed consumption patterns in different contexts, namely long tail, superstar or even both.

This paper contributes to extant literature about these two attributes by differentiating extreme scores from neutral user ratings and recognizing the interaction effect between user reviews and product variety. The impacts from user ratings are not uniform for all levels of ratings. The relationship between online user ratings and adoptions differs by the level of product variety. In an online market with product varieties exceeding thresholds for each quantile, products with original positive user ratings would exhibit better-by-better effect although the increase in neutral/negative reviews would also lead to more adoptions. And also, for all products with neutral user ratings, the increase in product variety makes more changes on product adoptions, which is stronger with respect to unpopular products than more popular ones, in support of long tail. In addition, this influence is more powerful for products with extreme user reviews than those with neutral user reviews. These results provide an explanation to reconcile the mixed results from previous studies.

Our results also shed light on the discussion about the emerging long tail prediction and superstar effect in e-commerce context. The various descriptive results and Pareto regression models shows a steeper head and a longer but flatter tail over time, which is consistent with some previous empirical studies [16, 35]. In addition, our findings point out that this shift of consumer preferences may not hold in other context with different levels of product variety and online user reviews. Also this study provides an explanation of the reasons underlying this coexistence of two consumption patterns, which has

received many mixed conclusions in extant literatures. The impact from either user reviews or product variety depends on whether the corresponding rating is positive, negative or neutral. Firstly, when the increase of neutral user reviews has positive impacts on adoptions of relative popular products more greatly, whether the increase of extreme user reviews exhibits more influence on popular products than tails is of uncertainty, depending on the discrepancy from actual product variety to threshold for each quantile. Secondly, on average, given all products with neutral ratings, the positive influence from product variety turns more powerful on unpopular products than on the more popular ones. Hence, the neutral user ratings contribute to superstar in contrast to the impact from product variety on adoptions of products with neutral ratings. Given different levels of user reviews, products with different conditions of product varieties would be possible to get increased to form the similar trends. That's also why it's possible to have both long tail and superstar in all three category although the variety of software programs posted during the period 2 in category of MP3 Finder is less than the variety during period 1 whereas other two categories has more product varieties.

To identify the trend of consumer preferences is a crucial component of the online business model. Our findings could provide some suggestions for the online market about how to better serve their consumers' demands. It shows that there is not a unique optimal level of product variety for all different product categories and even for products in one category with different popularities. Its interaction with different levels of user reviews should also be taken into account. This implies that a better understanding of assortment strategy should be related with the design of online WOM recommendation system in order to capture the targeted market or the more profitable market, i.e. more concentrated

superstar products, the niche market or both. At the same time, online retailers should also keep in mind that the recommendation system may also bring the negative impact on sales if the negative reviews are exposed too many.

5.2 Discussion

This study provides a very interesting idea—threshold of variety, on the way to explain the interactive influence from product variety and WOM effect. To our best knowledge, there is rarely existing research talking about this idea not mention delving into its meaning and implication. Due to the limited time, we didn't investigate how these thresholds comes and what they would influence the assortment strategy of online retailers. Thus as a further step of the following research, more focuses will be put on examining a legitimate approach for sellers to estimate the thresholds given their products' popularities and the corresponding marketing strategy.

Secondly, this study can't thoroughly answer the question about whether the online retailers get the commercial benefits from this coexistence of long tail and super star. Does this emerging market combined with the traditional blockbuster's market would bring the sellers more profits? To answer this question, one issue to be done is about testing if the niche market is rising at the expense of recession in the popular products. One of our current results about the product variety shows that give all products with original neutral reviews in categories of Download Manager and MP3 Finder, the increase in adoption of unpopular products is coming along with the decrease in very popular products'. Yet, the more detailed examination is needed to extend this implication to a general conclusion and also to the demand side of the market.

Finally, our current results show that the existence of expert opinions can have either positive or negative influences on user choices. We should go step further to figure out what's the threshold for expert opinions to impose positive influences by including the value of CNET reviews into our empirical model.

REFERENCES

- [1] Anderson, C. 2005. *The real meaning of 80/20*. Chris Anderson's Blog: http://www.thelongtail.com/the_long_tail/2005/03/the_real_meanin.html
- [2] Anderson, C. 2006. *The long tail: why the future of business is selling less of more*. Hyperion Press, New York, NY.
- [3] Anderson, C. 2008. *Excellent HBR piece challenging the Long Tail*. Chris Anderson's Blog: http://www.longtail.com/the_long_tail/2008/06/excellent-hbr-p.html
- [4] Bailey, D., Gao, G., Jank, W., Lin, M., Lucas, H. and Viswanathan, S. 2008. *The long tail is longer than you think: The surprisingly large extent of online sales by small volume sellers*. Working Paper, The Robert H. Smith School of Business, University of Maryland.
- [5] Bakos, Y. 1997. Reducing buyer search costs: implications for electronic marketplaces. *Management Science*, 43(12): 1676-1692.
- [6] Brynjolfsson, E., and Kemerer, C.F. 1996. Network externalities in microcomputer software: an econometric analysis of the spreadsheet market. *Management Science*, 42(12): 1627-1647.

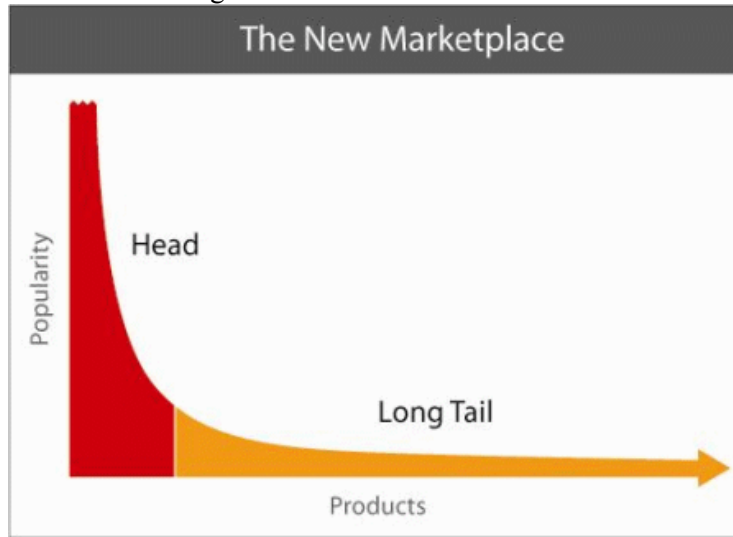
- [7] Brynjolfsson, E., Hu, Y., and Smith, M. 2003. Consumer surplus in the digital economy; estimating the values of increased product variety. *Management Science*, 49(11): 1580-1596.
- [8] Brynjolfsson, E., Hu, Y., and Smith, M. 2006. From niches to riches: The anatomy of the long tail. *Sloan Management Review*, 47(4):67-71.
- [9] Brynjolfsson, E., Hu, Y., and Simester D. 2007. *Goodbye pareto principle, hello long tail: the effect of search costs on the concentration of product sales*. Working paper.
- [10] Chevalier, J.A. and Goolsbee, A. 2003. *Valuing Internet retailers: Amazon and Barnes and Noble*. Working Papers ysm416, Yale School of Management.
- [11] Chevalier, J.A., and Mayzlin, D. 2006. The effect of Word of Mouth on sales: online book reviews. *Journal of Marketing Science*, 5: 345-354.
- [12] Clemons, E. , Gao, G. and Hitt, L.2006. When Online Reviews Meet Hyperdifferentiation: A Study of the Craft Beer Industry. *Journal of Management Information Systems*, 23 (2): 149 – 171.
- [13] Clemons, E. 2008. How Information Changes Consumer Behavior and How Consumer Behavior Determines Corporate Strategy. *Journal of Management Information Systems*, 25(2):13 – 40.
- [14] Dellarocas, C. N., and Narayan, R. 2007. *Tall heads vs. long tails: Do consumer reviews increase the informational inequality between hit and niche products?* Working paper. R.U. Smith School of Business, University of Maryland.

- [15] Duan, W., Gu, B. and Whinston, A. 2009. Informational cascades and software adoption on the Internet: an Empirical investigation, Forthcoming in *MIS Quarterly*, 33(1):23-48.
- [16] Elberse, A. and Oberholzer-Gee, F. 2007. *Superstars and underdogs: an examination of the long tail phenomenon in video sales*. Working Paper No.07-015, Harvard Business School.
- [17] Elberse, A. 2008a. Should you invest in the Long Tail? *Harvard Business Review*, 86: 88-96.
- [18] Elberse, A. 2008b. *The Long Tail Debate: A Response to Chris Anderson*. Harvard Business Online's Conversation Starter:
http://conversationstarter.hbsp.com/2008/07/the_long_tail_debate_a_respons.html
- [19] Fleder, D. and Hosanagar, K. 2007. *Blockbuster culture's next rise or fall: the impact of recommender systems on sales diversity*. Working Paper No.07-10, The Wharton School, University of Pennsylvania.
- [20] Frank, R. H. and Philip J. C. 1995. *The winner-take all society*. The Free Press, New York, NY.
- [21] Gallagher, J.M., and Wang, Y. M. 2002. Understanding network effects in software markets: evidence from web server pricing. *MIS Quarterly*, 26(4): 303-327.
- [22] Godes, D. and Mayzlin, D. 2004. Using online conversations to study Word of Mouth communication. *Marketing Science*, 23(4): 545-560.
- [23] Gourville, John T. and Soman, D. 2005. Overchoice and assortment type: when and why variety backfires. *Marketing Science*, 24(3):382-395.

- [24] Ghose, A. and Sundararajan A. 2005. *Software versioning and quality degradation? An exploratory study of the evidence*. Working Paper #ceDer-05-20, Center for Digital Economy Research, Leonard N. Stern School of Business.
- [25] Ghose, A. and Gu, B. 2007. *Search costs, demand structure and long tail in electronic markets: theory and evidence*. Working Paper No. 06-19, Stern School of Business, New York University.
- [26] Hervas-Drane, A. 2009. *Word of mouth and tasting matching: a theory of long tail*. Working paper #07-14. Columbia University and Universitat Pompeu Fabra.
- [27] Katz, M.L., and Shapiro, C. 1985. Network externalities, competition, and compatibility. *American Economic Review*, 75(3): 424-44.
- [28] Koenker R. and Gilbert B. 1978. Regression quantiles. *Econometrica*, 46(1): 33-50.
- [29] Lorenz, M. O. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, 9: 209–219.
- [30] McPhee, W. N. 1963. *Formal theories of mass behaviour*. The Free Press of Glencoe, New York, NY.
- [31] Rosen, S. 1981. The economics of superstars. *American Economic Review*, 71(5): 845-858.
- [32] Rosen, S. 1990. Selling versus producing in market fluctuations: a comment. *Carnegie-Rochester Conference Series on Public Policy*, 33(1): 89-91.
- [33] Senecal, S. and Nantel, J. 2004. The influence of online product recommendations on consumers' online choices. *Journal of Retailing*, 80(2): 159-169.
- [34] Schmidt, E. 2005. *Annual Stockholders' Meeting, Google Inc*, 12 May.

- [35] Tucker, C. and Zhang, J.J. 2007. *Long tail or steep tail? A field investigation into how online popularity information affects the distribution of customer choices*. Working Paper NO.4655-07, MIT Sloan School of Management.
- [36] Van den Bulte, C. and Gary L. L. 2001. Medical innovation revisited: social contagion versus marketing effort. *American Journal of Sociology*, 106(5): 1409-35.
- [37] Zhao, X., Gu, B. and Whinston, A. 2008. The influence of online word-of-mouth long tail formation: an empirical analysis. *Proceeding of CIST 2008*.

FIGURE 1
Long Tail on Statistical Distribution



Adapted from Chris Anderson's Blog

FIGURE 2
Screen Snapshot of Software Listing Page

MP3 Finders

The MP3 Finders directory consists of software programs created to help users search for and find MP3 files online, on local networks, or on their local computers. Programs in this category may use P2P technology, a form of file sharing, or may facilitate commercial music outlets.

Narrow search

by License type:

- Free (68)
- Free to try (24)

by Operating system:

- Windows XP (91)
- Windows 2000 (83)
- Windows Vista (62)
- Windows 98 (69)
- Windows Me (57)
- Windows NT (49)
- Windows 2003 Server (44)
- Windows 95 (32)
- Windows 3.x (2)
- Windows MCE (1)
- See all

▼ Ad Feedback
Advanced Registry Optimizer 5

Show 10 results per page Sort by: Downloads last week

sponsored

Free MP3 Downloads
Get new music, free and legal. Download new MP3s at fuse.tv!
fuse.tv

Download Music - \$.99/ea
Listen to 5,000,000+ FULL Songs Download Your Favorites - Try Free!
www.rhapsodymusicervice.com

LimeWire 4.18 Share files online. Read editor's review... Added on June 01, 2008	Editors' rating: ★★★★☆ User rating: ★★★★☆	159,931,996 total downloads 573,309 this week	Download Now
FrostWire 4.17 Download and share files on Gnutella and Bittorrent network. Read editor's review...	Editors' rating: ★★★★☆ User rating: ★★★★☆	10,659,521 total downloads 278,319 this week	Download Now

FIGURE 3
Different Shape Indexes in Pareto Distribution

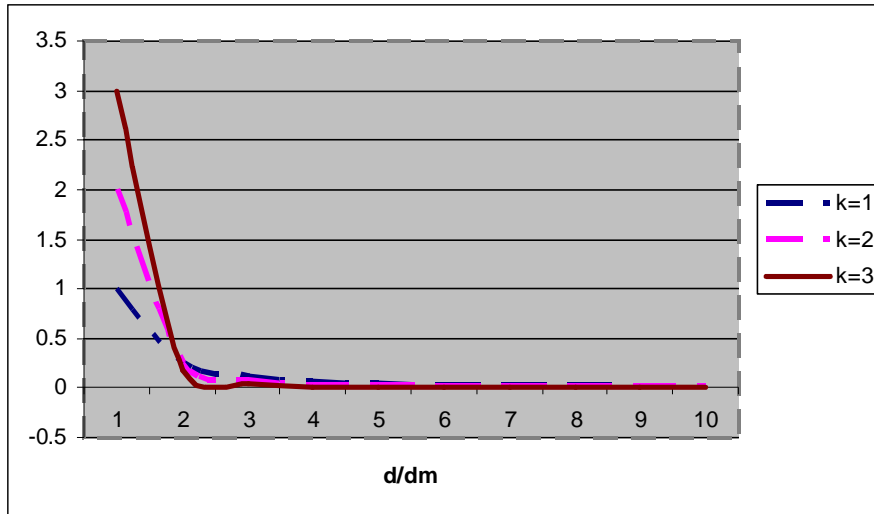


TABLE 1
Distribution of CNET Ratings during Period 2

	Monthly average number of software programs	Monthly average number of software programs with CNET reviews	Monthly percentage of software programs with CNET reviews
Digital Media Players	489	41	8.38%
Download Managers	254	49	19.44%
MP3 Finders	106	11	10.38%

TABLE 2
Distribution of User Choices

%Rank	mean		Max		Min		Sd. Deviation	
	Period1 ⁷	Period2 ⁸	Period1	Period2	Period1	Period2	Period1	Period2
Digital Media Player								
1	0.001	0.0001	0.003	0.0003	0.00	0	0.0001	0
5	0.01	0.002	0.02	0.003	0.01	0.001	0.0004	0.0001
10	0.05	0.01	0.06	0.01	0.03	0.01	0.001	0.0002
25	0.28	0.05	0.34	0.06	0.21	0.03	0.005	0.001
50	1.64	0.29	1.87	0.36	1.31	0.23	0.02	0.01
75	6.69	1.28	8.40	1.54	5.19	1.08	0.12	0.02
90	16.22	4.56	20.37	4.99	13.26	4.15	0.23	0.03
95	24.62	10.55	31.14	12.37	19.58	9.48	0.37	0.09
99	57.07	47.31	70.679	61.48	42.68	36.13	1.06	0.94
Download Manager								
1	0.001	0.0004	0.002	0.001	0.00	0	0.00	0
5	0.03	0.01	0.05	0.01	0.02	0.001	0.001	0.0002
10	0.10	0.02	0.14	0.03	0.07	0.01	0.002	0.001
25	0.45	0.14	0.63	0.17	0.31	0.07	0.01	0.004
50	1.78	0.77	2.31	0.96	1.39	0.51	0.03	0.02
75	5.85	3.87	7.45	5.01	4.80	2.97	0.10	0.08
90	17.51	12.14	20.05	15.65	14.94	9.68	0.18	0.21
95	31.47	22.54	36.28	29.35	26.12	16.45	0.39	0.41
99	72.75	68.40	80.30	79.12	68.87	63.33	0.36	0.58
MP3 Finder								
1	0.0005	0.0001	0.002	0.0002	0.00	0	0.00	0
5	0.01	0.001	0.01	0.002	0.003	0.0002	0.0004	0.0001
10	0.02	0.003	0.04	0.01	0.01	0.002	0.001	0.0002
25	0.14	0.02	0.19	0.04	0.07	0.01	0.004	0.001
50	0.86	0.17	1.17	0.28	0.48	0.09	0.02	0.007
75	3.30	1.03	5.17	1.60	1.99	0.56	0.10	0.04
90	8.80	4.33	15.41	5.60	6.30	2.97	0.32	0.11
95	18.46	11.45	33.68	18.04	11.35	7.82	0.74	0.38
99	64.94	61.87	81.87	77.43	47.27	37.59	1.49	1.46

⁷ Period 1 refers to period from Dec. 2004~Jul.2005.

⁸ Period 2 refers to period from Aug. 2007 to Feb.2008

TABLE 3
Inter-quantile Positions Summary

		mean	
		Period1	Period2
Digital Media Player			
Location	Q0.5	203.75	484
Scale	$(Q0.75-Q0.25)/(Q0.75+Q0.25)$	0.02	0.01
Skewness	$(Q0.75+Q0.25-2Q0.5)/(Q0.75-Q0.25)$	-0.73	-0.38
Kurtosis	$(Q0.90-Q0.10)/(Q0.75-Q0.25)$	4.35	3.06
Inter-quantile	Q0.75-Q0.25	8.50	8.14
Left tail	$(Q0.5-Q0.25)/Q0.5$	0.04	0.01
Right tail	$(Q0.75-Q0.5)/Q0.5$	0.006	0.01
Total titles	Q100:Monthly average product variety	205.88	489
Head titles	Total titles- Q0.5	2.13	5
Download Manager			
Location	Q0.5	140.5	252
Scale	$(Q0.75-Q0.25)/(Q0.75+Q0.25)$	0.03	0.02
Skewness	$(Q0.75+Q0.25-2Q0.5)/(Q0.75-Q0.25)$	-0.59	-0.37
Kurtosis	$(Q0.90-Q0.10)/(Q0.75-Q0.25)$	2.90	3.21
Inter-quantile	Q0.75-Q0.25	8	10.17
Left tail	$(Q0.5-Q0.25)/Q0.5$	0.04	0.03
Right tail	$(Q0.75-Q0.5)/Q0.5$	0.01	0.01
Total titles	Q100:Monthly average product variety	143.38	254
Head titles	Total titles- Q0.5	2.88	2
MP3 Finder			
Location	Q0.5	145.38	106
Scale	$(Q0.75-Q0.25)/(Q0.75+Q0.25)$	0.018	0.012
Skewness	$(Q0.75+Q0.25-2Q0.5)/(Q0.75-Q0.25)$	-0.34	-0.22
Kurtosis	$(Q0.90-Q0.10)/(Q0.75-Q0.25)$	2.59	2.50
Inter-quantile	Q0.75-Q0.25	5.33	2.67
Left tail	$(Q0.5-Q0.25)/Q0.5$	0.02	0.01
Right tail	$(Q0.75-Q0.5)/Q0.5$	0.01	0.009
Total titles	Q100:Monthly average product variety	147.38	108
Head titles	Total titles- Q0.5	2	2

TABLE 4
Descriptive Results of User Ratings

%Rank	mean		Max		Min		Sd. Deviation	
	Period1	Period2	Period1	Period2	Period1	Period2	Period1	Period2
Digital Media Player								
1	3.46	4.00	5.00	4.00	0.00	4.00	1.86	0.00
5	4.09	4.00	4.64	4.00	3.70	4.00	0.32	0.00
10	4.05	4.00	4.38	4.00	3.81	4.00	0.20	0.00
25	3.87	2.91	4.01	3.25	3.73	2.63	0.11	0.21
50	3.75	3.08	3.84	3.19	3.63	2.94	0.08	0.10
75	3.69	3.12	3.78	3.19	3.58	3.00	0.07	0.07
90	3.67	3.19	3.75	3.24	3.61	3.16	0.05	0.03
95	3.64	3.22	3.72	3.27	3.57	3.20	0.06	0.02
99	3.64	3.24	3.71	3.28	3.57	3.21	0.05	0.02
Download Manager								
1	0.50	0.00	3.00	0.00	0.00	0.00	1.22	0.00
5	2.47	2.50	3.00	2.50	2.00	2.50	0.43	0.00
10	3.20	2.17	3.38	2.50	2.92	1.50	0.19	0.52
25	3.23	2.53	3.53	3.00	3.00	2.00	0.18	0.33
50	3.34	2.96	3.51	3.08	3.24	2.833	0.11	0.08
75	3.40	3.28	3.56	3.44	3.30	3.17	0.09	0.09
90	3.45	3.32	3.51	3.51	3.38	3.21	0.05	0.10
95	3.47	3.34	3.52	3.48	3.41	3.24	0.04	0.08
99	3.47	3.38	3.52	3.51	3.42	3.29	0.05	0.07
MP3 Finder								
1	3.08	0.00	4.50	0.00	1.00	0.00	1.21	0.00
5	3.22	4.00	3.57	4.00	3.00	4.00	0.20	0.00
10	3.17	3.75	3.44	4.00	2.92	3.00	0.22	0.50
25	3.32	3.81	3.36	4.00	3.31	3.50	0.02	0.24
50	3.38	3.30	3.42	3.37	3.34	3.20	0.03	0.07
75	3.45	3.38	3.57	3.41	3.29	3.33	0.09	0.03
90	3.44	3.40	3.55	3.43	3.34	3.35	0.07	0.03
95	3.43	3.40	3.52	3.44	3.33	3.36	0.06	0.03
99	3.43	3.42	3.50	3.45	3.34	3.38	0.05	0.02

TABLE 5
Estimation of Pareto Regression Model

	mean		Max		Min		Sd. Deviation	
	Period1	Period2	Period1	Period2	Period1	Period2	Period1	Period2
Digital Media Player	-0.85 *** ⁹	- 1.01***	-0.80 ***	- 0.97***	-0.89 ***	- 1.04***	0.03	0.02
Download Manager	-0.85 ***	- 0.96***	-0.79 ***	- 0.91***	-0.87 ***	- 1.04***	0.02	0.04
MP3 Finder	-0.96 ***	- 1.24***	-0.89 ***	- 1.09***	-1.05 ***	- 1.34***	0.04	0.10

FIGURE 4¹⁰
Popularity of Sort options on Download.com Listing Pages

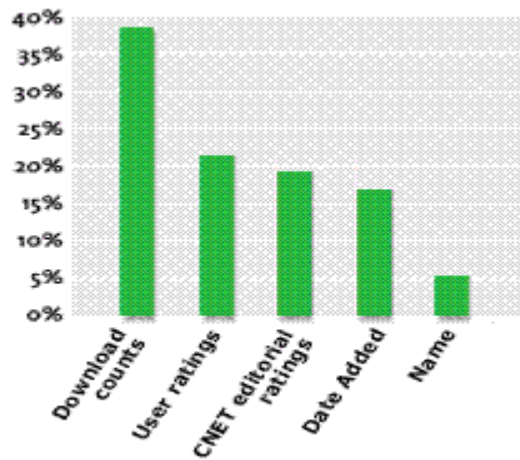


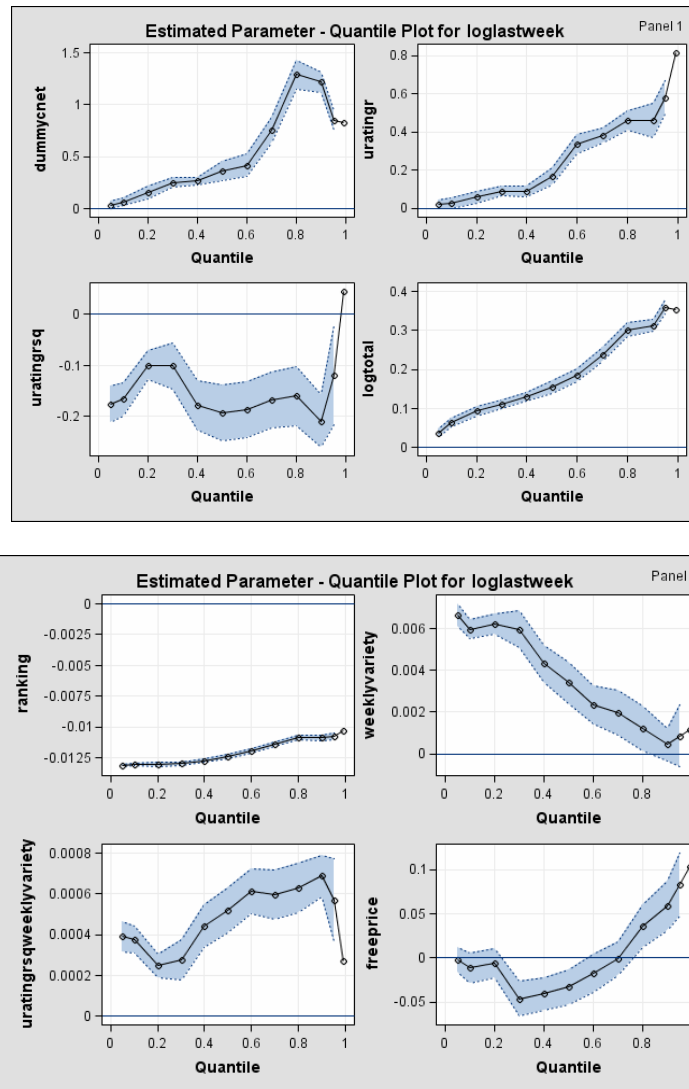
TABLE 6
Description of Independent Variables

Variable	Description and Measure
CNETRATINGD _{i,t}	A dummy variable measures if software i has CNET rating at time t
USERRATINGR _{i,t}	Difference between user rating for software i at time t and point “3”
USERRATINGRSQ _{i,t}	Square term of USERRATINGR _{i,t}
LOG(TOTAL DOWNLOAD) _{i,t}	Log transformation of cumulative total number of downloads of software i at time t
WEEKLYRANK _{i,t}	The rank of weekly downloads of software i at time t
WEEKLYVARIETY _{i,t}	The total number of weekly software programs in which category software i posted at time t
WEEKLYVARIETY _{i,t} * USERRATINGRSQ _{i,t}	Interaction Term
FREE PRICED _{i,t}	A dummy variable measures if software i is free-to try at time t
AGE _{i,t}	Days since software i has been posted.
AGESQ _{i,t}	Square term of AGE _{i,t}

⁹ *** for p < .0001.

¹⁰ Adapted from CNETD at http://www.upload.com/1200-21_5-5139592.html?tag=fd

FIGURE 5a
 Quantile Plot for Regression of Digital Media Player during Period2 ¹¹



¹¹ The solid circle denotes at this quantile the estimator is insignificant at 5%. The shadow denotes the 95% confidence level. The estimation results for variable “Age” and “Agesq” are not shown in above figures.

FIGURE 5b
 Quantile Plot for Regression of Download Manager during Period 2

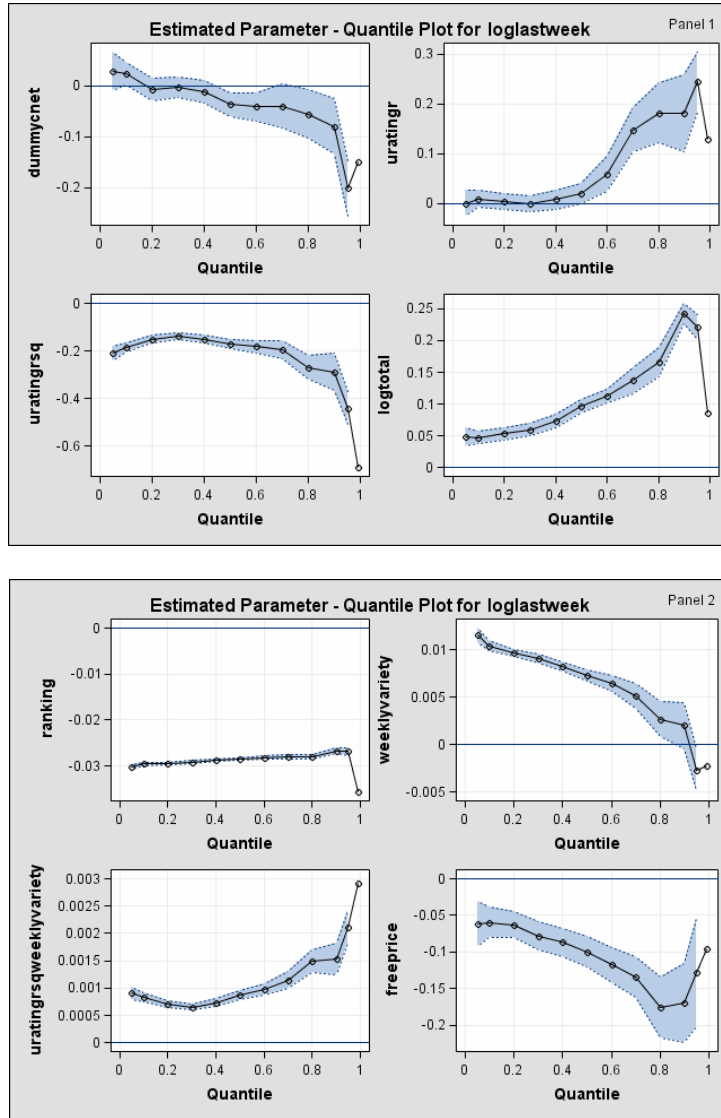


FIGURE 5c
Quantile Plot for Regression of MP3 Finder during Period 2

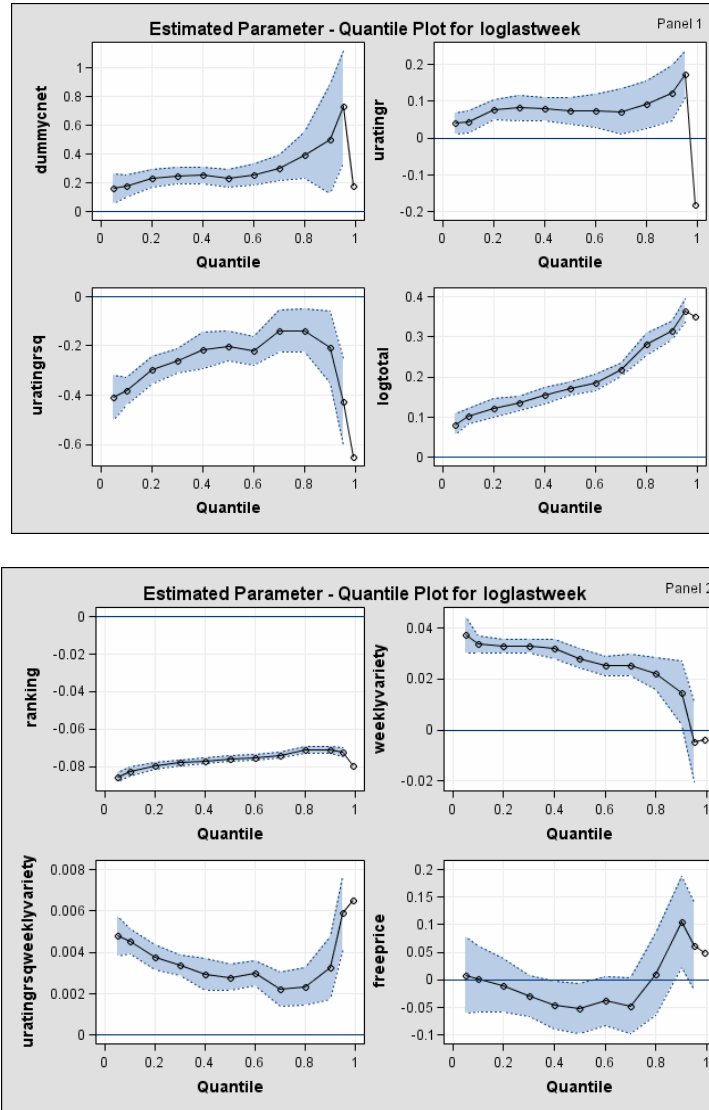


TABLE 7
Threshold for Product Variety

	Digital Media Player	Download Manager	Mp3 Finder
0.05	440.75	234.56	85.63
0.10	415.00	232.25	85.18
0.20	498.00	215.00	78.55
0.30	337.00	197.00	76.91
0.40	445.00	215.57	75.03
0.50	384.80	191.22	71.64
0.60	309.83	183.30	73.20
0.70	279.33	162.42	64.05
0.80	265.67	179.60	59.83
0.90	299.00	193.33	64.22
0.95	197.83	210.14	72.61

0.99	Insig. coefficient	238.48	100.43
Average weekly variety	454.23	222.92	93.19
Sd. Deviation	68.26	35.99	12.80