

## Association for Information Systems AIS Electronic Library (AISeL)

---

ICIS 2008 Proceedings

International Conference on Information Systems  
(ICIS)

---

2008

# The Hitchhiker's Guide to the Long Tail: The Influence of Online-Reviews and Product Recommendations on Book Sales - Evidence from German Online Retailing

Thomas Ehrmann  
*University of Muenster*

Hendrik Schmale  
*University of Muenster*

Follow this and additional works at: <http://aisel.aisnet.org/icis2008>

---

### Recommended Citation

Ehrmann, Thomas and Schmale, Hendrik, "The Hitchhiker's Guide to the Long Tail: The Influence of Online-Reviews and Product Recommendations on Book Sales - Evidence from German Online Retailing" (2008). *ICIS 2008 Proceedings*. 157.  
<http://aisel.aisnet.org/icis2008/157>

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2008 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# **THE HITCHHIKER'S GUIDE TO THE LONG TAIL: THE INFLUENCE OF ONLINE-REVIEWS AND PRODUCT RECOMMENDATIONS ON BOOK SALES – EVIDENCE FROM GERMAN ONLINE RETAILING**

*Le guide du voyageur de la longue traîne : l'influence des commentaires en ligne  
et des recommandations des produits sur les ventes de livres –  
résultats issus du commerce en ligne allemand*

*Completed Research Paper*

**Thomas Ehrmann**

University of Muenster  
Department of Strategic Management  
Leonardo-Campus 18  
48149 Muenster, Germany  
ehrmann@ism.uni-muenster.de

**Hendrik Schmale**

University of Muenster  
Department of Strategic Management  
Leonardo-Campus 18  
48149 Muenster, Germany  
schmale@ism.uni-muenster.de

## **Abstract**

*Exploring the long tail phenomenon, we empirically analyze whether online reviews, discussion forums, and product recommendations help to reduce search costs and actually alter the sales distribution in online book retailing. We have collected a data set containing 320,248 observations for 40,031 different books at Amazon.de, each assigned to one of 111 different product categories in our sample. By adopting an innovative approach, we provide the first long tail conversion model for the German online market, based on publicly available sales data. Our results indicate that online reviews and automated product recommendations reduce search costs by facilitating the identification of adequate books and the assessment of their quality. This highlights the relevance of information technology implementation as vital part of the marketing strategy.*

**Keywords:** Long tail, ecommerce, search costs, online reviews, entertainment markets, books

## Résumé

En analysant un jeu de données contenant 320 248 observations relatives à 40 031 livres différents sur Amazon.de, nous montrons que les évaluations en ligne, les forums de discussion et les recommandations de produits aident à réduire les coûts de recherche du consommateur et modifient effectivement la répartition des ventes de livre, conformément aux dires de la théorie de la longue traîne. Adoptant une approche innovante, nous proposons le premier modèle de conversion de la longue traîne à l'e-commerce en Allemagne.

## Introduction

A fast growing body of literature deals with the widely recognized phenomenon of the long tail of ecommerce demand. Being omnipresent throughout the whole entertainment industry, especially in the book sector, this phenomenon is based on the following insight:

While big brick-and-mortar book retailers' assortments comprise a maximum of approximately 200,000 books, online retailers like Amazon or Barnes & Noble generate more than a quarter of their sales volume with titles which are positioned beyond the Top200,000 in their sales ranking lists. Hence, it could be argued that the real existing book market exceeds the conventional brick-and-mortar market by one-third in sales volume. The phenomenon of a high total sales volume of products which individually sell a very small number of copies is known as *the long tail*. Chris Anderson, chief editor of the Wired-Magazine, was first to analyze the far reaching economic impact of this insight on the existing entertainment market landscape (Anderson 2006).

Traditionally, the world of retailing followed the Pareto principle. The Pareto principle, also known as 80/20-rule, states that 80 percent of total sales are generated by 20 percent of the offered products. Hence, in a Pareto world economic success emanates from the efficiency of attracting a high number of consumers for single hit products. In the motion picture industry, for example, the studios' chase for the next blockbuster results in steadily increasing production and marketing budgets aiming at occupying the first position in the potential audience's cognitive space. Consequently, entertainment markets were typically dominated by a winner-take-all mentality.

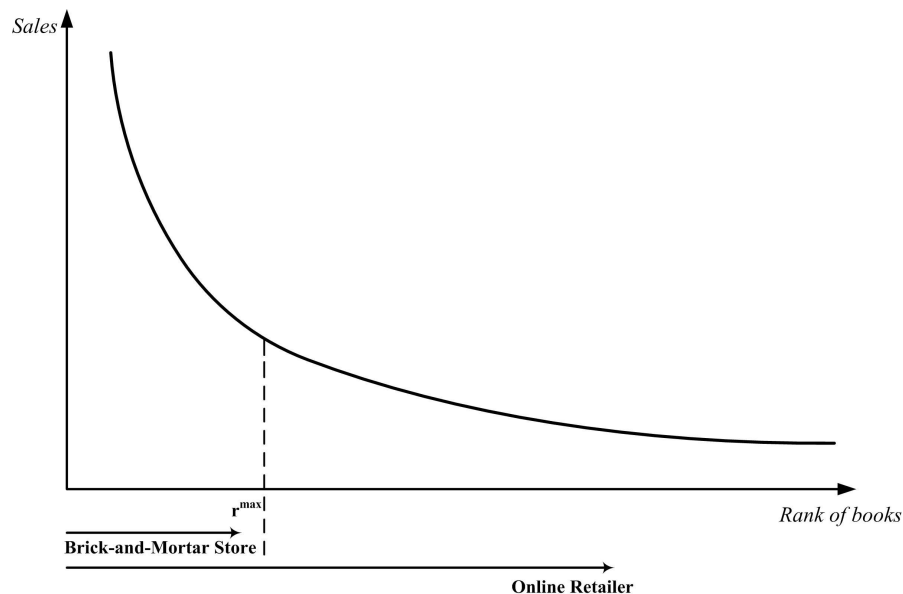


Figure 1: The Long Tail of Books

However, the long tail phenomenon suggests that those products which sell too little to be accommodated in brick-and-mortar stores are responsible for a much higher fraction of total sales than predicted by the Pareto principle. Consequently, this fact could alter the supply-sided orientation moving away from hits, towards niche products. Put

differently, in online markets, it could be lucrative to broaden the product range with respect to less popular and more obscure products.

Theory predicts that both supply-side and demand-side drives foster the profitability of a long tail strategy. On the one hand, online markets feature – as opposed to conventional stores – nearly unlimited shelf space. From this follows that the opportunity costs of listing obscure products converge to zero, eventually making it profitable to list niche products. Online retailing also allows to aggregate geographically dispersed demand, which potentially has positive effects on the overall sales volumes of low-selling products.

On the other hand, the long tail phenomenon is driven by changes in consumer behavior. The implementation of powerful search tools and automated recommendation systems which allow “hitchhiking” on popular books to the unknown sphere of the niche comes along with declining consumer search costs in terms of opportunity costs of time. Thus, it becomes more advantageous for consumers not to buy the well known and prominently positioned hit products, but to expand search effort to find a product which better fits personal requirements and preferences than the mainstream product. The theoretical prediction is that declining search costs will even out demand between products, thereby leading to reduced demand inequity, and to a further flattening of the sales distribution (Brynjolfsson et al. 2003; 2006).

Given that first empirical evidence indicates substantial changes in the distribution of sales across products (e.g. Brynjolfsson et al. 2006; Elberse et al. 2006; Oestreicher-Singer et al. 2006), researchers have begun to reveal determinants of the long tail formation in more detail. In particular, and most interesting to our research, scholars and practitioners emphasized the role of user-generated online reviews and discussion forums for driving demand down the tail by reducing product quality uncertainty (Kawasaki, 2006). However, despite this widespread belief there is a lack of empirical evidence on the *effectiveness* of e-Word-of-Mouth instruments for reducing search costs, and relatedly, on their role in changing the sales distribution.

In our paper, we empirically test the effects of online reviews, discussion forums, and automated product recommendations on the demand for individual products and the overall distribution of sales of the related product categories. Thereby, we elaborate on complementarities between the sales-enforcing instruments. Using weekly data of over 40,000 books from Germany's biggest online book retailer Amazon.de, we compare the demand distributions of more than 100 product categories with respect to differences in their *Gini* coefficients, measuring inequality in sales distribution. Our main prediction is that user-generated online reviews reduce search costs and thus alter the categories' sales distributions towards equality.

One major challenge in analyzing the determinants of long tail formation is to estimate actual demand levels from corresponding sales ranks, since online retailers typically are very hesitant in providing real sales data. However, using an innovative approach based on publicly available sales data, we provide the first estimation of a long tail conversion model inferring actual demand levels for obscure products from corresponding sales ranks for the German book market.

The remainder of this paper is structured as follows. First, we will give an overview of the theoretical background and introduce our research questions. Second, we describe the data. Third, we will give an introduction into our research methodology. Fourth, we present the results of our empirical test. Finally, we conclude.

## **Theoretical Background and Hypotheses**

### ***Search Costs and the Long Tail***

Analyzing the U.S. VHS and DVD market Elberse et al. (2006) show that a supply-sided broadening of the product range indeed comes along with an increasing popularity of niche products. Nevertheless, such broadening also implies a very high number of titles which never sell at all. Hence, this might constitute a non-negligible expense factor for retailers pursuing a long tail strategy. Accordingly, Anderson (2006) formulates another fundamental prerequisite for the successful implementation of a long tail strategy other than broadening the assortment to a maximum diversified product range: Reduction of demand-side search costs.

Obviously, only those retailers who manage to bring together supply and demand cost effectively will profit from the tail in the long run. First empirical evidence indicates that search costs decline in the internet channel compared

to conventional channels like catalogue retailing. This highlights the relevance of information technology implementation as vital part of the marketing strategy (Brynjolfsson et al. 2006).

Interpreting product search as a two-staged process of (1) identification of products and (2) assessment of fit between the products characteristics and a consumer's quality and functional requirements (Stiglitz 1989), there are two main classes of setting levers for reducing search costs.

First, search costs emanating from attempts to identify potential products can be reduced by developing specialized search filters and automated recommendation systems. For instance, Amazon provides on most of the books' web pages information on other books that have been co-purchased by other customers. Following these product recommendation links, customers possibly locate formerly unknown but adequate products with a higher probability than by random search, thereby incurring lower search costs due to less total time spent searching. Oestreicher-Singer et al. (2006) provide empirical evidence for the influence of Amazon's co-purchase network on product sales concentration. They show that product categories which are exposed to higher levels of network influence exhibit a flatter sales distribution which implies search cost reducing effects of recommendation systems and an increasing demand for niche products. Relatedly, Chen et al. (2004) find that product recommendations improve sales better for less popular books than for more popular books at Amazon.com.

Second, search costs can be reduced by implementing instruments which are targeted toward reducing quality uncertainty. As described above, product recommendations foster demand for niche products. However, a prominent position in the co-purchase network necessitates that a high number of customers previously bought the product in a bundle with a more prominent good. Actually, most of the niche products are per definition not rated by the purchase behavior of the broad customer mass. As Guy Kawasaki (2006) puts it: "this is where two cool concepts butt head: long tail vs. wisdom-of-the-crowds." To the contrary, online customer reviews can give additional support for the assessment of obscure products by providing more detailed information on the products characteristics even though very few copies have previously been sold. That is to say, one antecedent purchase and review post alone can potentially help to reduce the perceived risk of wasting time and money on inadequate products (Anderson 2006; Bakos 1998; Kawasaki 2006; Oestreicher-Singer et al. 2006). With regard to a long tail strategy, reviews are of high economic relevance due to their cost efficiency, since this form of community content requires relatively low maintenance compared to retailer self-provided product information.

Among long tail scholars there is widespread belief that, with respect to the demand side of the market, decreasing transaction costs come along with an increasing heterogeneity in consumption patterns (e.g. Brynjolfsson et al. 2006, Hervas-Drane 2007). Theory argues that consumers benefit from increasing product variety in being able to buy products which better match personal preferences than conventional mainstream products. As the number of different offerings increases it becomes more likely that adequate products will be found (e.g. Chernev 2003). In this respect, conventional brick-and-mortar markets' Pareto sales distributions reflect to some extent – the long tail view holds – suboptimal choices from an individual perspective. Hence, it is the "consumers' appetite for niche products" (Elberse et al. 2006, p.4) which constitutes a prerequisite for a lucrative long tail to emerge.

Based on long tail theory's basic assumptions Brynjolfsson et al. (2006) develop a theoretical model which illustrates the effects of a demand-side search cost reduction on sales concentration. Key to the model is the fact that products differ *ex ante* in search costs. On the one hand, "key" products which are heavily promoted by expensive marketing campaigns generate lower search costs since these products entail a high awareness among consumers. On the other hand, purchasing niche products involves high search costs because these products are not visible per se. The assumption of heterogeneous search costs implies a change in sales distribution as a result of search cost reduction. In contrast, if products were assumed to be homogeneous a reduction in search costs should affect all products symmetrically, implying a stable sales concentration.

In our research setting, we allow for the fact that products might differ *ex ante* in search costs. Consider for example products that stand out of the broad mass due to their appearance on the most visible web pages of an online retailer or due to the fact that they have been reviewed in TV shows and so on. In contrast, for niche products, there are few other sources of information available since obscure products are not prominently positioned due to heavy advertising or promotion. Consequently, user generated online reviews – or word-of-mouth generally – should be especially influential on the sales of these obscure products because it is the only source of information available (Chevalier et al. 2006).

## ***Research Questions and Hypotheses***

We analyze whether recommendations and user generated online reviews can help to reduce search cost and whether these instruments in fact foster a shift in consumption patterns away from hits to the niche.

If this is the case, we should observe asymmetric effects – in the sense of strength or direction of influence – of those instruments on sales across books with respect to their ranking. To find this out we focus on two perspectives in the remainder of the paper: (1) The relevance of online reviews and product recommendations for demand from a book perspective, and (2) the influence of product recommendations and online reviews on the distribution of sales from a category perspective. In particular we pose following global research questions:

- (1) Do customer reviews, discussion forums and product recommendations affect sales for hit and long tail products equally?
- (2) Do customer reviews and product recommendations complement each other and do they affect the sales distribution?

Some empirical evidence has been provided by scholars on the basic influence of online customer reviews, or generally, e-Word-of-Mouth (eWOM), on sales of books (e.g. Chevalier et al. 2006). On the one hand, eWOM can create customer awareness and on the other hand it may be one of the only sources of information about the quality of experience goods (Li et al. 2004), i.e. products which cannot be assessed prior to consumption (Akerlof 1970). Hence, as customers face high risk to waste money on inadequate products they might rely on other customers' prior experiences in order to reduce quality uncertainty.

However, customer ratings as a measure of product quality have been controversially discussed in the literature. For example, Chevalier et al. (2003a; 2003b; 2006), Li et al. (2004), and Dellarocas et al. (2004) show that the average rating – Amazon's review system allows for customer ratings on a five-point scale – significantly affects product sales. In contrast, Chen et al. (2004) and Liu (2006) prove positive influence by the number of reviews posted rather than the reviews' valences. Hu et al. (2007) ascribe this inconsistency of findings with regard to review valence to the fact that the average rating might be a skewed measure for quality. They show that reviews are not normally distributed over the five-point scale but j-shaped, implying a very high number of positive reviews, a smaller number of negative reviews, and an even smaller number of medium reviews. One possible explanation for the low number of medium ratings is the under-reporting bias, which states that review posting requires a certain involvement by the customer. The authors suggest that this involvement is higher when a product is either really appreciated or disliked. The high number of positive reviews is attributed to a purchasing bias. Most reviewers ex-ante seem to have a positive attitude towards the product they review, because they have obviously purchased it. As a consequence, customers might display a negativity bias in assigning higher credibility to negative rather than positive reviews because of their rare occurrence (Sen et al. 2007). With regard to the literature, we try to disentangle the effects of review valence by explicitly accounting for differences in influence of negative and positive reviews. Hence, with respect to review valence we pose following hypotheses:

H1a: The higher the fraction of positive reviews, the higher the title's sales.

H1b: The higher the fraction of negative reviews, the lower the title's sales.

Because of its experience good character buying a book can constitute a rather complex situation. As noted above, people might index their expectations regarding fit and quality of the book to consumer reviews. However, the interpretation of feedback and the assessment of its usefulness in many cases require a certain amount of prior knowledge on the topic or the author, for instance. Moreover, as the consumption of many books is also characterized by an affective and sensory experience of aesthetic or sensual pleasure (Hirschman & Holbrook 1982) – i.e. the products not least satisfy emotional wants – reviews are subjective and thus depend on the individual emotional constitution of the reviewer. As this constitution can hardly be assessed by potential buyers it is difficult to appraise whether the evaluation applies to oneself (Sen et al. 2007). Consequently, people might try to cope with this situation by transforming the complex context into a rather simple decision rule. In particular, they might orientate themselves to the number of review posts as an indicator for reviewer involvement and popularity of the book.

H2: The higher the number of customer reviews, the higher the title's sales.

Another possibility to cope with uncertainty in terms of feedback credibility and applicability is to rely on opinion leaders. Since top reviewers possess an above-average history of helpful reviews customers might impute certain

knowledge to them which comes along with higher credibility. Moreover, books often feature a symbolic character, i.e. people read books in order to appeal cultured or literate. Hence, people face high risk to read the “wrong” books and consequently might prefer books which top reviewers have found worth reading:

H3: Titles which have been discussed by top reviewers have higher sales.

Due to the social nature of people books play an important role as means of social interaction. That is, people seem to be devoted to hear the same music, watch the same movies and read the same magazines and books (Elberse et al. 2006). One theoretical explanation for this fact is that utility derived from reading – at least to some extent – depends on the ability to appreciate it. This ability in turn improves with the accumulation of author or topic specific ‘consumption capital’. Put differently: The more you know, the more you appreciate it (Stigler et al. 1977). Consumption capital thereby can be accumulated either by consuming books of the same author or of related topics or by discussions with others. Since it is costly to search for someone to interact people might prefer reading books with larger ‘user communities’ because it becomes more likely to find interaction partners. Furthermore, large user communities indicate high involvement and thus might convince that the book is worth reading. Hence, with respect to the relationship between the number of discussions and sales we pose the following research hypothesis:

H4: The higher the number of discussions, the higher the title’s sales.

Providing community content is one strategy to reduce customer uncertainty involved in buying experience goods. Another more and more frequently pursued strategy is to enable free trial. Specifically, retailers provide reading excerpts (e.g. the first 20 pages of a book). The excerpts allow for unbiased quality assessment by the customer and thus might contribute to search cost reduction:

H5: Titles with reading excerpts have higher sales.

As already noted, automated recommendation systems can reduce search costs by simplifying the identification of adequate books. In this respect, a book’s network position on a retailer’s e-commerce site influences the amount of traffic which is being directed to the product’s detail page. The network position thereby is determined by the books it links to, and those that link to it. However, books that are either linked to by one or more popular books or by a high number of moderate-sellers are likely to enjoy an increase in sales on account of improved customer traffic. Put differently, the book should “enter” more potential buyers’ individual choice sets. Besides that, product links do not only entail higher traffic levels but they might also provide content information because they allow customers to relate unknown books to something they possibly know. Hence with respect to automated recommendation links we deduce following hypothesis:

H6: The higher the network influence, the higher the title’s sales.

Books benefit from enhanced ease of finding as a result of a favorable network position. Furthermore, reviews, discussions and excerpts reduce search costs by providing quality information. We argue that both classes of search cost reducing instruments, namely recommendations on the one hand and reviews, discussion forums and excerpts on the other hand reinforce each other. To be more specific, recommendations should intensify the sales enforcing effects of reviews, discussions and excerpts because they increase the amount of customers who notice the quality information in the first place, i.e. a review that is never found cannot impact sales. Secondly, reviews, discussions and excerpts might intensify the positive impact of recommendations because they provide additional information from a quality perspective, i.e. product links limit the relevant choice set, however, following those links might not help much further if none of the recommended books has ever been reviewed.

H7: Higher network influence in association with reviews, discussions or excerpts improves the title’s sales.

As discussed in the previous section, long tail theory argues that reducing search costs fosters a shift in sales concentration away from hits to the niche. The rationale behind this is the fact that books are not homogeneous with respect to search cost. In particular, hits feature a higher awareness among customers due to the general availability of various information sources such as bestseller lists, newspaper reviews, advertising and so on. Consequently, community content (i.e. reviews and discussions), recommendations and excerpts should play a much more important role for sales success in the long tail where search costs are high. Or the other way round:

H8: Community content (i.e. reviews and discussions), recommendations and reading excerpts do not affect sales in the front of the sales distribution.

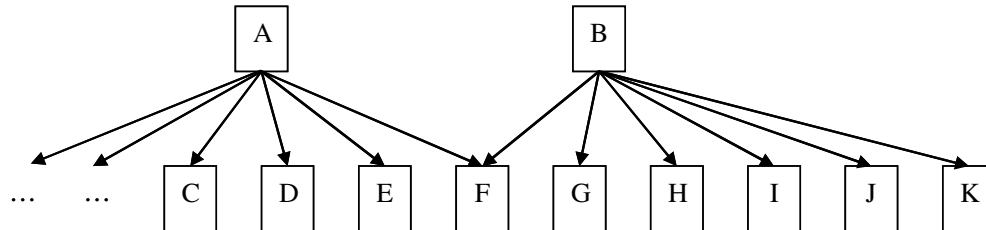
And, relatedly:

- H9: Community content (i.e. reviews and discussions), recommendations and reading excerpts alter the sales distribution towards equality.

## Data

Using Perl-based scripts, we collected individual characteristics, review data, and network information of initially 56,744 books at the public website of Amazon.de, Germany. Since we did not have access to Amazon.de's real sales data, we tried to select a representative sample as follows:

On Amazon.de, you can find different listings of bestselling books, all with a maximum of 4,800 titles – either including books related to one of the main categories, or including books across all categories. We took the list across all categories as a starting point for our data collection. Given that these books are Amazon.de's overall bestsellers, this sample is not sufficient because of two reasons. First, this sample only consists of very successful titles, which is obviously a major downside with respect to our main intention: testing assumptions on the long tail of book sales. Second, although this list possibly contains books from all existing categories, books from categories with a high average demand, like e.g. “Mystery & Thrillers > Thrillers”, will surely be overrepresented. This in turn is obviously a major downside for collecting a maximum variety of categories. Consequently, we added another 4,800 books collected from all 22 main categories. In order to dig deeper into the long tail, we furthermore added a random generator based sample of 10,000 books from vlb-directory (Verzeichnis lieferbarer Buecher) which is the German Books in Print directory. The vlb-directory consisted of over 1,100,000 different titles in January 2008. Hence, the probability of extracting niche titles was very high. However, we also wanted to collect information on product recommendations, but unfortunately the co-purchase link network is a directed graph, i.e., it is possible to collect all the subsequent books one title refers to but it is not possible to identify all precedent books which refer to a title. Thus, we took a backward induction approach to draw our main sample. Amazon.de provides up to six co-purchase links on every book's main page (see for example book A and B in Figure 2). We collected all co-purchase links of our initial sample of 19,600 books thereby broadening the sample size to 56,744 different titles with co-purchase information (books C to K in Figure 2).



**Figure 2: Examples of Co-Purchases for Books A and B**

In addition to this we collected following data for each of the books in our sample: isbn (unique serial number), title, author, price, shipping time, release date, format, publisher, number of pages, a set of different review variables and category affiliation.

Amazon.de classifies its books by a hierarchy of categories. This hierarchy consists of main categories, sub-categories and eventually sub-sub-categories and so on. For example, “The Hitchhiker’s Guide to the Galaxy” by Douglas Adams is classified as follows: “Books > Science Fiction & Fantasy > Science Fiction > Adventure”. Hence, the majority of titles are assigned to one or more of these classification hierarchies. In our sample, we concentrate on the third level from the top of the hierarchy in order to assure a maximum number of categories, while maintaining a minimum category size of 40 books.

The review variable set includes: total number of reviews, number of 5-star (highest rating) and 1-star ratings (lowest rating), and number and valence of reviews by “Top-Reviewers”. Furthermore, we included a dummy variable indicating whether Amazon.de provides an excerpt (e.g., a pdf-file containing the first 20 pages).

Finally, we collected the sales rank of each book. The top-selling book at Amazon.de has a sales rank of one, the lower sellers are assigned higher sequential ranks. Amazon.de states on its website that ranks 1-10,000 are updated hourly, ranks 10,001-100,000 daily and ranks beyond 100,001 every month (Amazon.de 2008). However, our data



sample contains a considerable number of books whose ranks are far beyond 100,000 and changing at least every two hours, indicating that all ranks seem to be updated daily, if not hourly. For the data collection this poses issues. Hourly changing ranks imply significant variations in sales ranks within one day. For example, a book which is ranked around 150,000 can move up to rank 20,000 as a result of a sale of one. Thus, collecting sales ranks resembles shooting at moving targets. Consequently, we collected data on each of our books twice a week in a four week period between March 10 and April 4, 2008. We computed an average rank for each title based on eight observations which should represent a sufficient approximation of weekly sales ranks (Rosenthal 2008). Since some of the books fell out of the co-purchase graph during our data collection period we excluded all books which were not represented by eight observations in our database. Our final sample consists of 320,248 observations for 40,031 different books at Amazon.de, each assigned to one of the 111 different categories in our database.

## Method and Variables

As mentioned above, our empirical test will focus on two perspectives: (1) The relevance of online reviews and product recommendations for demand from a book perspective, and (2) the influence of product recommendations and online reviews on the distribution of sales from a category perspective. Hence, we will give an introduction into our research method and the main variables of interest.

### *Long Tail Conversion Model for Estimating Sales*

Brynjolfsson et al. (2003) and Chevalier et al. (2003a) find that the relationship between sales ranks and sales at Amazon.com is log-linear constituting a power-law distribution:  $\ln[\text{sales}_i] = \alpha + \beta \ln[\text{rank}_i]$ . Using proprietary sales data from a publisher Brynjolfsson et al. (2003) calibrate the relationship between sales and sales ranks by estimating  $\alpha = 10.525$  and  $\beta = -0.871$  for Amazon.com. Most empirical work using publicly available data resorts to these parameters for estimating sales. This seems to be appropriate for studies based on the American market; however, these parameters are less suitable when analyzing consumer behavior in other countries. This model would obviously overestimate sales of books, taking into account that the American market exceeds the German approximately by a factor of four. Consequently, we estimated our own conversion model, which should be more suitable for the data from Amazon.de. Since we did not have access to proprietary data we took another approach: During our data collection we realized, that Amazon.de provides information on remainder of stock for books which they have five or less copies left. This information is updated several times a day. Hence, we extracted all the books within our database which had this information displayed and checked sales rank variations and stock variations five times a day for a seven-day period in order to compute weekly sales and sales ranks.

Our initial sample contained 14,089 books of which about 6,000 exposed variations in the remainder of stock. We excluded all titles which had positive variations in stock within the collection period and all those which had less than one copy left at the end of day seven. We did this to ensure that the variation in sales ranks can be related to variations in stock, since Amazon.de does not stop selling a product which is not in stock. Another approach to minimize potential effects of unobserved replenishment was to exclude all titles from the sample that displayed declining sales ranks but either no variation in stock or positive variation (i.e. sale and replenishment within the same 2½ hours). Furthermore, we excluded all titles with increasing ranks (i.e. no sale) and positive variation in stock (i.e. replenishment).

Our final sample contained 540 books with weekly sales between one and four books and average rankings between 53,193 and 959,888. The mean number of sales was 1.44 and the mean average rank was 163,322. Table 1 shows our regression results. We estimated  $\alpha = 8.114$  and  $\beta = -0.656$ . The model is highly significant, although the  $R^2$  is not as high as expected. Rosenthal (2008) reports that Amazon.com's new ranking system includes own sales as well as third party sales, which could be an explanation for this value. A limitation of this approach is the fact that we relied on long tail data. In particular, small errors with respect to the log-normal parameters which were estimated in the tail of the curve might be magnified significantly near the head of the curve, possibly leading to wrong estimates for the top sellers. Yet, anecdotal evidence suggests that our estimates fit quite well: Welt-Online (2008) reports that the book "Feuchtgebiete" (i.e. the No.1-seller in our sample) should have sold 3,100 units per week at Amazon.de within our collection period. Our model estimates weekly sales of 3,341 which imply an accuracy of 93% for the top selling book. However, without having proprietary data this should nevertheless be an approximate estimation of Amazon.de's sales which will be the dependent variable in the book perspective.

**Table 1: Long Tail Conversion Function**

Method: OLS	coefficients	
<i>constant</i>	8.114***	(0.516)
<i>Ln[avg_rank]</i>	-0.656***	(0.043)
n	540	
F	235.518***	
Adjusted R <sup>2</sup>	0.30	
Dependent variable: Ln[sales]. Standard errors in parentheses.		
<b>Significance levels (2-tailed):</b> *** p < 0.01; ** p < 0.05; *p < 0.1.		

### Measuring Sales Distribution

In line with the existing literature we characterize the sales distribution of a category by computing the Gini coefficient which is based on the categories' Lorenz curves. The Gini coefficient will constitute the independent variable in our category research perspective. The Gini coefficient is a measure of distributional inequality ranging from 0 to 1. A value of 0 thereby corresponds to perfect equality which means that all books within a category have identical sales. A value of 1 corresponds to perfect inequality, meaning that the most popular book of a category generates all sales within a category and all other books have no sales.

A category's Lorenz curve ranks all the products within the category in increasing order of demand and then plots the cumulative fraction of sales  $L(p)$  associated with each ascending rank percentile  $p$ , where  $0 < p < 1$ . We follow Oestreicher-Singer et al. (2006) in the definition of the Lorenz curve: Let  $N = \{1, 2, 3, \dots, n\}$  be the set of all books in a category of size  $n$ , having  $q(i)$  as the demand for book  $i$ . To compute the Lorenz curve, the size of the set of all products whose demand is equal or less to that of  $i$ ,  $R(i)$ , has to be defined. Thus,  $R(i)$  is simply the inverse rank of a product within its category, with the product with the lowest demand having the lowest rank. Next, we define the set of product indices whose rank is less than or equal to  $r$ :  $S(r) = \{y \in N, R(y) \leq r\}$ . Then, for each percentile  $p$  (which corresponds to the books ranked  $np$  or lower) the Lorenz curve is defined by:

$$L(p) = \frac{\sum_{y \in S(np)} q(y)}{\sum_{y \in N} q(y)}.$$

In other words,  $L(0.05)$ , for instance, depicts the cumulative fraction of sales of those 5% of a category's books which have the lowest ranks within the category. Think of a category of size 40;  $L(0.05)$  then depicts the cumulative fraction of sales of the books with the ranks 40 and 39. Consequently,  $L(0.1)$  depicts the cumulative fraction of sales of the books with the ranks 40, 39, 38, and 37. Notice that the Lorenz curve is increasing and convex.

The Gini coefficient is computed as twice the area between the Lorenz curve  $L(p)$  and the 45-degree line between the origin and (1,1). Note that the 45-degree line corresponds to perfect equality. This means, when the Lorenz curve has the form of the 45-degree line, the sales distribution is perfectly equal. The Gini coefficient is calculated by first computing the entire area above the Lorenz curve, the Lorenz upper area:

$$LU = \sum_{y=1}^n \left[ 1 - L\left(\frac{y}{n}\right) \right],$$

and then using identity:

$$Gini = 2(LU) - 1.$$

## Empirical Test

### *Model Specification and Descriptive Statistics: Book Level*

To estimate the influence of online reviews, discussion forums, and product recommendations on sales we specified the following model:

$$sales_i = b_1X_i + b_2Y_i + b_3Z_i + b_4XY_i + \varepsilon_i$$

Here  $sales_i$  denotes the actual demand of book  $i$  according to our previously introduced conversion model.  $X_i$  describes a set of review variables including: number of reviews, fraction of 1-star reviews, fraction of 5-star reviews, a dummy variable indicating whether a top-reviewer occurs on the books main page, and number of discussion posts. Furthermore, we included a dummy for reading excerpts.

$Y_i$  measures the network's immediate influence on a book due to recommendations. This variable, linkvalue, is a score which depends on both the number of recommendations (i.e. links) a book receives and the quality of these recommendations in terms of demand of the preceding books (Oestreicher-Singer et al. 2006).

$Z_i$  is a set of control variables including book characteristics: price, shipping days, format (paperback dummy), and number of days since the book was published. Furthermore, we tried to control for two sources of "visibility" which could explain the variation in sales: a dummy indicating whether a book appeared in one of the three leading German literature TV-shows (Lesen!, ZDF; Druckfrisch, ARD; Literaturclub, 3sat) in the month before data collection and a dummy indicating whether a book was displayed on the categories' main pages on Amazon.de.

$XY_i$  is a set of interaction terms which intend to capture complementarities between the sales enforcing instruments. In particular, we included: one variable capturing whether product recommendations intensify the positive influence of online reviews on demand, one variable capturing complementarities between product recommendations and top-reviews, one variable indicating complementarities between product recommendations and discussions, and one variable capturing complementarities between recommendations and reading excerpts.  $\varepsilon_i$  captures the random error.

**Table 2: Descriptive Statistics Book Level**

	Mean	S.D.	Min.	Max
0. sales	5.441	27.991	0.200	3340.910
1. number_rev	6.390	29.906	0	2569
2. review_dummy	0.690	0.462	0	1
3. linkvalue	18.471	88.773	0	3884.360
4. frc_1star	0.053	0.143	0	1
5. frc_5star	0.618	0.357	0	1
6. top_dummy	0.180	0.387	0	1
7. excerpt	0.150	0.358	0	1
8. number discuss	0.030	3.394	0	676
9. price	15.794	14.275	2	980
10. ship_time	0.440	1.519	0	35
11. paperback_dummy	0.650	0.477	0	1
12. days_published	1314.430	1418.951	0	39475
13. categorypage_dummy	0.010	0.083	0	1
14. tvapperance_dummy	0.000	0.011	0	1

N varies between 22,706 and 40,031.

Table 2 displays the descriptive statistics of our sample. Sales vary between 0.2 and 3340.91 which corresponds to average sales ranks of 2,712,656 and 1, respectively. The review dummy indicates that 69% of the books had at least

one review posted, whereas 18% of titles had a rating by a top reviewer. The average star-rating is 4.286 (s.d. 0.79, not included in table 2) indicating that reviews are very positive on average on Amazon.de, which is in line with previous findings (e.g. Chevalier et al. 2006; Hu et al. 2007).

### ***Empirical Results: Book Level***

Table 3 displays the results of our first two model specifications. Model 1 includes the main variables of our analysis. In order to achieve linearity between the dependent variable and both number of reviews and linkvalue, we log-transformed these two independent variables (Hair et al. 1998). Overall, the results indicate that the fraction of 5-star reviews, the number of reviews, and the number of discussions posted, have significantly positive effects on demand, thereby corroborating hypotheses H1a, H2, and H4. This indicates that the implementation of e-Word-of-Mouth instruments is of relevance for demand. Furthermore, the results support H6: books which receive more and qualitatively better recommendations by other books in the network have higher sales.

However, we do not find any support for H1b and H5 as the fraction of 1-star reviews and reading excerpts have no significant effects on sales overall. With respect to negative reviews this is indeed surprising, since previous findings and theoretical considerations predict a stronger influence of negative reviews compared to positive reviews. Furthermore there is no support for H3. Although feedback by top reviewers affects sales significantly, the variable has a negative sign. We will refer to this fact in the next section in order to provide possible explanations for this finding.

**Table 3: Empirical Results Full Sample Book Level**

Method: OLS	Model 1			Model 2		
	coefficients	stand. coefficients	st. errors	coefficients	stand. coefficients	st. errors
<i>constant</i>	-2.324***		(0.652)	1.588**		(0.666)
Ln[number_rev]	3.385***	0.111	(0.198)	1.766***	0.058	(0.205)
frc_1star	-1.333	-0.005	(1.526)	-0.499	-0.002	(1.505)
frc_5star	1.808***	0.018	(0.614)	1.996***	0.020	(0.605)
Ln[linkvalue]	3.377***	0.149	(0.142)	1.699***	0.075	(0.160)
top_rev (dummy)	-0.952*	-0.012	(0.489)	-0.276	-0.003	(0.499)
excerpt (dummy)	0.609	0.006	(0.568)	0.299	0.003	(0.571)
number_discuss	0.367***	0.045	(0.048)	1.657***	0.205	(0.155)
Price	-0.021	-0.007	(0.018)	-0.040**	-0.014	(0.018)
ship_time	0.027	0.001	(0.161)	-0.094	-0.003	(0.159)
paperback (dummy)	-0.458	-0.006	(0.467)	-0.676	-0.009	(0.461)
days_published	-0.001***	-0.026	(0.000)	-0.001***	-0.029	(0.000)
categorypage (dummy)	108.384***	0.303	(2.156)	102.776***	0.287	(2.138)
tvappearance (dummy)	17.312	0.007	(14.579)	26.681*	0.011	(14.394)
Ln[linkvalue]*Ln[number rev]				2.526***	0.181	(0.106)
Ln[linkvalue]*top rev				-1.380***	-0.031	(0.296)
Ln[linkvalue]*excerpt				0.939***	0.016	(0.358)
Ln[linkvalue]*discuss				0.932***	0.163	(0.110)
n	24,351			24,351		
F	362.687***			326.925***		
Adjusted R <sup>2</sup>	0.16			0.19		

Dependent variable: sales. Standard errors related to unstandardized coefficients in parentheses.

**Significance levels (2-tailed):** \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1

Note that books without review are treated as missing value here, since our main models include both the fractions of positive and negative reviews. We fitted another model including a dummy variable indicating the presence of at least one review and the results were stable. Of course, the dummy was highly significant and positive indicating that books profit from being reviewed (not reported in Table 3 due to space constraints).

Turning to the control variables: Table 3 shows that the number of days since publishing has a highly significant but weak negative impact on sales. Surprisingly, a book’s TV appearance has no significant effect on sales although one might expect a positive promotion impact at least. A possible explanation for this fact might be compensating effects with respect to the TV critics’ review valences since tvappearance is a dummy variable. However, analyzing the sales enforcing effects of TV critics Clement et al. 2006 show that neither positive nor negative feedbacks influence demand, which basically is in line with this result. Nevertheless, a higher visibility on the categories’ main pages had positive impact on sales.

However, with respect to the question whether the sales enforcing instruments interact, we included the interaction terms in Model 2. As can be seen, all interactions are significant. Hence, recommendations have greater impact on demand when there is a reading excerpt available indicating that detailed and unbiased quality information complements the influence of recommendations. However, H7 is only partly corroborated since linkvalue\*top\_rev displays a negative sign thereby indicating an alleviating effect of top reviewers’ feedbacks on the positive influence of recommendations. The interaction terms containing number of reviews and number of discussions are both significant and positive. Hence, we see an intensifying effect.

**Table 4: Empirical Results Long Tail vs. “Short Tail” Books**

Method: OLS	Ranks < 50,000			Ranks ≥ 50,000		
	coefficients	stand. coefficients	st. errors	coefficients	stand. coefficients	st. errors
constant	-5.366***		(1.488)	1.593***		(0.013)
Ln[number_rev]	5.020***	0.126	(0.372)	0.119***	0.197	(0.005)
frc_1star	0.526	0.001	(4.359)	-0.273***	-0.089	(0.026)
frc_5star	4.565***	0.028	(1.462)	0.127***	0.094	(0.011)
Ln[linkvalue]	4.328***	0.141	(0.269)	0.181***	0.338	(0.004)
top_rev (dummy)	-1.743*	-0.016	(0.946)	0.018*	0.014	(0.011)
excerpt (dummy)	1.415	0.011	(1.136)	0.046***	0.032	(0.012)
number_discuss	0.360***	0.045	(0.068)	0.163***	0.023	(0.056)
Price	-0.067	-0.013	(0.047)	-0.004***	-0.095	(0.000)
ship_time	-0.141	-0.003	(0.414)	-0.028***	-0.078	(0.003)
paperback (dummy)	-0.904	-0.008	(0.964)	-0.010	-0.009	(0.010)
days_published	-0.001***	-0.039	(0.000)	-0.000***	-0.056	(0.000)
categorypage (dummy)	106.019***	0.297	(3.097)	0.486	0.012	(0.337)
tvappearance (dummy) <sup>a</sup>	16.257	0.007	(20.766)			
n	11,851			12,499		
F	164.706***			252.853***		
Adjusted R <sup>2</sup>	0.15			0.20		

Dependent variable: sales. Standard errors related to unstandardized coefficients in parentheses.

**Significance levels (2-tailed):** \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1

<sup>a</sup> The TV-appearance dummy was excluded from the analysis in the long tail sample due to missing variance.

Table 4 displays the results of our main model specification for the book level, yet we ran the regressions comparing different groups of books in order to capture asymmetric effects of our predictors with respect to sales ranks: more popular books and long tail books. The capacity of brick-and-mortar stores has frequently been used in the literature

to (theoretically) approximate where the long tail in online retailing begins, i.e. the group of books which individually sell too little in order to be stocked in conventional book stores (e.g. Anderson 2006, Brynjolfsson et al. 2003). In the literature, figures vary between 40,000 and 250,000 books depending on store size. However, Rosenthal (2005) states that Barnes and Noble's core stocking (i.e. the group of the most popular books which are available in all stores) comprise 50,000 titles. Hence we chose this rather conservative demarcation line of 50,000 to divide long tail from, let say, "short tail" books.

The results for the "short tail" books (rank < 50,000) remain stable compared to the full sample. Interestingly, the fraction of 1-star reviews displays no significant negative influence whereas it does in the long tail sample, as we will see later. Hence, looking at rank one of Amazon.de's hit list one might be tempted to think: "There is no bad publicity – there is just publicity". The book "Feuchtgebiete" by Charlotte Roche, a German (ex-) TV-star, has been controversially discussed within the last weeks. Although the average customer rating of this book is 2½ stars with a fraction of one-star reviews of 45.4% (total of 447 reviews) the book sticks to rank 1 for weeks now. However, the publicity argument should only explain why the total number of negative reviews might not have negative influence. As long as bad publicity is worse than good publicity the fraction of 1-star reviews should be unambiguously bad.

Turning to the long tail (rank ≥ 50,000) we see that negative reviews have negative impact on sales, which was expected. This indicates that quality information is valued more for books which are not "hot" anymore.

Feedback by top-reviewers significantly fosters sales in the long tail which contrasts our results for the "short tail" sample. A further investigation intended to reveal possible differences in the functionality of word-of-mouth with respect to popular and unpopular books. A test, which we have conducted, indicates that in the "short tail" consumers seem to respond to negative deviations of the top-reviewers' ratings relative to the average customer rating (i.e. negative deviations suppress sales) but they do not respond to positive deviations. Interestingly, in the long tail things are completely reversed implying that positive deviations foster sales and negative have no effect. This might indicate that when it comes to the niche, where the number of possibly adequate substitutes decreases, critical opinion leader reviews lose their bite and positive ones are welcome confirmations of quality.

It appears that in this sample, overall, all sources of additional information are highly valued by the customers. This accounts for reading excerpts as well as for top-reviews. The review of an opinion leader can apparently be a fixed star within the long tail which helps to identify adequate books, when there is no crowd to follow. Price has a significantly negative impact here in contrast to the "short tail" sample. This might be due to the following: the German book market is characterized by a book price control law which prohibits price reductions for newly published books. However, this fixing can be released for older books, implying little variance in the sample of popular books, which mostly consists of new books.

In sum and with respect to hypotheses H8 we indeed see systematic differences of the influence of online-reviews on the two data samples, yet recommendations have positive influence on both the "short tail" and long tail. Hence, to further analyze the effects of these instruments on the sales distribution we turn to the category perspective.

### ***Model Specification and Descriptive Statistics: Category Level***

To estimate the influence of online reviews, discussion forums and product recommendations on the sales distribution we specified the following model:

$$Gini_i = b_1X_i + b_2Y_i + b_3Z_i + \varepsilon_i$$

Here  $Gini_i$  denotes the actual level of inequality in sales of category  $i$ .  $X_i$  describes a set of review variables with respect to the category level, including: fraction of titles being reviewed, average number of reviews per category, fraction of titles being discussed, and fraction of titles with reading excerpts.

$Y_i$  is a set of recommendation variables: average link value of the category, and variance in link values.

$Z_i$  is a set of control variables. We control for category size and for the fact whether books are fictional or nonfictional.  $\varepsilon_i$  captures the random error.

Table 5 displays the descriptive statistics of our sample.  $Gini$  varies between 0.003 and 0.280 which indicates that the average sales distribution is relatively flat. This stems from the fact that we included all books with ranks greater 50,000 into our category sample, since we are especially interested in the long tail. With respect to the full sample, our results indicate that reviews actually do have an impact on both popular and niche books. Consequently, even if

an online review has greater impact on sales in the long tail because there is less alternative information available (e.g. Chevalier et al. 2006), the average higher number of reviews in the “short tail” might overcompensate this effect. A similar finding is reported by Elberse et al. (2006), who found that the long tail of DVD-sales indeed gets flatter in the tail, but steeper in the front at the same time. In order to separate possible effects on the long tail we took this approach.

**Table 5: Descriptive Statistics Category Level**

	Mean	S.D.	Min.	Max
0. GINI	0.147	0.471	0.004	0.279
1. frac_reviewed_titles	0.622	0.151	0.183	0.967
2. avg_linkvalue	5.524	2.534	2.353	18.057
3. frac_titles_excerpt	0.170	0.150	0	0.975
4. frac_discussed_titles	0.004	0.007	0	0.038
5. category size	190.810	269.947	40	2240
6. nonfictional	0.800	0.400	0	1
7. var_linkvalue	124.434	262.667	4.847	2495.777
8. avg_number_rev	3.680	1.878	1.182	11.109
N is 111				

The mean fraction of discussed titles is very small in our sample. This is plausible, since this tool is one of the newer features on Amazon’s website and therefore books in the tail, which are likely to be dated, are less discussed. The mean fraction of titles which are reviewed is 0.622, indicating that in each category more than a half of titles are reviewed at least once. On average, every category has a mean number of reviews of 3.680 varying from 1.182 to 11.109.

**Empirical Results: Category Level**

**Table 6: Empirical Results Category Level**

Method: OLS	Category Level		
	coefficients	stand. coefficients	st. errors
<i>constant</i>	0.201***		(0.022)
frac reviewed titles	-0.111***	-0.355	(0.035)
avg linkvalue	-0.007***	-0.373	(0.002)
frac titles excerpt	-0.070**	-0.224	(0.027)
frac discussed titles	0.772	0.115	(0.558)
category size	0.000*	0.165	(0.000)
nonfictional	0.045***	0.378	(0.011)
var linkvalue	0.000*	0.241	(0.000)
avg number rev	0.004	0.169	(0.003)
n	111		
F	6.670***		
Adjusted R <sup>2</sup>	0.29		
Dependent variable: GINI. Standard errors related to unstandardized coefficients in parentheses.			
<b>Significance levels (2-tailed):</b> *** p < 0.01; ** p < 0.05; * p < 0.1			

Table 6 shows the results of our estimation on the category level. As can be seen, the fraction of reviewed titles, the average level of recommendations (*avg\_linkvalue*) and the fraction of titles with reading excerpt are negative and highly significant. Note that a smaller Gini coefficient implies a more equal sales distribution. Hence, we find support for H9 in the sense that online reviews, recommendations and excerpts reduce search costs, thereby altering the sales distribution of products in the long tail. This indicates that these instruments are indeed of high relevance for promoting niche products because products which ex-ante face higher search costs actually profit more from being reviewed, at least in the tail. For product recommendations this effect remains stable while we control for the variance of product recommendations in a category (*var\_linkvalue*). This coefficient is weakly positive and significant on the 10% level, indicating that a high variance in the number and the quality of recommendations within a category fosters sales concentration. Category size is significant, too, and weakly positive. The fraction of discussed titles shows a positive coefficient, however, it is not significant. The dummy variable controlling for structural differences between fictional and nonfictional categories is highly significant, and positive. This implies that nonfictional categories have a steeper sales distribution than fictional ones.

## **Conclusion**

Exploring the long tail phenomenon, we have analyzed whether online reviews, discussion forums, and product recommendations help to reduce search costs and actually alter the sales distribution in online book retailing. We have collected a data set containing 320,248 observations for 40,031 different books at Amazon.de, each assigned to one of 111 different product categories in our sample. By adopting an innovative approach we were the first to develop a long tail conversion model for the German online market, based on publicly available sales data.

Examining variation in sales of individual books we found out that these e-WOM based instruments and automated recommendation systems in fact influence sales asymmetrically. In particular, our results indicate that negative reviews have no impact on sales of popular books in the “short tail”, whereas they depress sales in the long tail. Furthermore, positive reviews from opinion leaders are valued more in the long tail, which also accounts for reading excerpts. Positive reviews, the number of reviews, and product recommendations foster sales in the “short tail” and the long tail. Generally, e-WOM instruments and product recommendations interact in their influence. Specifically, reading excerpts intensify the positive impact of product recommendations on demand. The number of reviews and discussion posts interact with product recommendations in terms of mutual intensification of their impact on demand.

With respect to our main research intention, our empirical test from the category perspective proves that the fraction of reviewed titles, the average level of recommendations and the fraction of titles with reading excerpt within a category account for differences in the categories’ long tail sales distributions. Hence, online reviews and automated product recommendation systems obviously reduce search costs by facilitating the identification of obscure books and the assessment of their quality.

Although our results indicate that there might be opposing effects of reviews on the full length of the short and long tail, e-WOM and product links nevertheless function as customer leading fixed stars in the endless space of the long tail of books, thereby fostering the profitability of a long tail strategy.

Our paper seeks to contribute to the emerging literature on the long tail of e-commerce. Yet, we see several opportunities for further research. For instance, the analysis of panel data could reveal yet unobserved category specific properties that might drive sales concentration. Furthermore, the consideration of competitor data could minimize the effects of possible endogeneities in the sense that higher sales also trigger review volume and link formation. Finally, more research is needed from other industries (e.g. the music industry) in order to better understand the new rules in long tail markets.

## **Acknowledgements**

We thank two anonymous reviewers, the associate editor and the track chair for valuable insights and comments that greatly improved the paper.



## References

- Akerlof, G.A. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics* (84:3), August 1970, pp. 488-500.
- Anderson, C. *The Long Tail*, Hyperion, New York, June 2006.
- Bakos, Y. "The emerging role of electronic marketplaces," *Communications of the ACM* (41:8), August 1998, pp. 35-42.
- Brynjolfsson, E., Hu, Y.J., and Smith, M. D. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science* (49:11), November 2003, pp. 1580-1596.
- Brynjolfsson, E., Hu, Y.J., and Simester, D. "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales" *Working Paper*, December 2006.
- Chen, P.-Y., Wu, S.-Y., and Yoon, J. "The Impact of Online Recommendations and Consumer Feedback on Sales" *Proceedings of the International Conference on Information Systems (ICIS 2004)*, Washington, D.C., December 2004.
- Chernev, A. "When More is Less and Less is More: The Role of Ideal Point Availability and Assortment in Consumer Choice," *Journal of Consumer Research* (30), September 2003, pp. 170-183.
- Chevalier, J. and Goolsbee, A. "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com," *Quantitative Marketing and Economics* (1:2), June 2003a, pp. 203-222.
- Chevalier, J. and Goolsbee, A. "Valuing Internet Retailers: Amazon and Barnes and Noble," *Advances in Applied Microeconomics: Organizing the New Industrial Economy* (12), Elsevier Science, Michael Baye, Amsterdam, 2003b.
- Chevalier, J.A. and Mayzlin, D. "The effect of word of mouth on sales: online book reviews," *Journal of Marketing Research* (43:3), August 2006, pp. 345-354.
- Clement, M., Proppe, D. and Sambeth, F. "Der Einfluss von Meinungsführern auf den Erfolg von hedonischen Produkten" *Zeitschrift für Betriebswirtschaft* (76), July 2006, pp. 797-824.
- Dellarocas, C., Awad, N. and Zhang, X. "Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning," *Proceedings of the 24<sup>th</sup> International Conference on Systems*, Washington, D.C., May 2004.
- Elberse, A. and Oberholzer-Gee, F. "Superstars and Underdogs: An Examination of the Long Tail Phenomenon in Video Sales" *Working Paper Series No. 07-15*, Harvard Business School, August 2006.
- Godes, D. and Mayzlin, D. "Using online conversations to study word-of-mouth communication," *Marketing Science* (23:4), Fall 2004, pp. 545-560.
- Hair, J., Anderson, R., Tatham, R., and Black, W. 1998. *Multivariate data analysis – Fifth edition*, Prentice-Hall, Upper Saddle River, NJ, 1998.
- Hervas-Drane, A. "Word of Mouth and Recommender Systems: A Theory of the Long Tail" *NET Institute Working Paper No. 07-41*, November 2007.
- Hirschman, E.C. and Holbrook, M.B. "Hedonic Consumption: Emerging Concepts, Methods, and Propositions," *Journal of Marketing* (46:3), 1982, pp. 92-101.
- Hu, N., Pavou, P.A., Zhang, J. "Why do Online Product Reviews have a J-shaped Distribution? Overcoming Biases in Online Word-of-Mouth Communication," *Working Paper*, February 2007.
- Kawasaki, G. [www.blog.kawasaki.com](http://www.blog.kawasaki.com), 2006.
- Li, X. and Hitt, L.M. "Self selection and information role of online product reviews" *Workshop on Information Systems and Economics (WISE 2004)*, Washington, D.C., 2004.
- Liu, Y. "Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing* (70:3), July 2006, pp. 74-89.
- Oestreicher-Singer, G. and Sundararajan, A. "Recommendation Networks and the Long Tail of Electronic Commerce," *Working Paper*, October 2006.
- Rosenthal, M. "Barnes and Nobel Book Store vs. Amazon Stocking", (accessed May 6, 2008) <http://www.fonerbooks.com/2005/08/barnes-and-noble-book-store-vs-amazon.html>, 2005.
- Rosenthal, M. "Amazon Sales Rank for Books", (accessed May 6, 2008) <http://www.fonerbooks.com/surfing.htm>, 2008.
- Sen, S. and Lermann, D. "Why are you telling me this? An examination into negative consumer reviews on the web," *Journal of Interactive Marketing* (21:4), Autumn 2007, pp. 76-94.
- Stiglitz, J.E. "Imperfect Information in the Product Market", in *Handbook of Industrial Organization*, Schmalensee, R. and Willig, R.D. (Eds.), Elsevier, Amsterdam, 1989, pp. 769-847.

Stigler, G. and Becker, G. "De Gustibus Non Est Disputandum," *American Economic Review* (67:2), 1977, pp. 76-90.  
Welt-Online, (accessed May 3, 2008) [http://www.welt.de/kultur/article1888908/Die\\_Bestseller\\_Methoden\\_der\\_Charlotte\\_Roche.html](http://www.welt.de/kultur/article1888908/Die_Bestseller_Methoden_der_Charlotte_Roche.html), 2008.