

Do Online Recommendations Matter? – A Multimodal Investigation of Amazon’s Co-Purchase Network

Hsuanwei Michelle Chen
School of Information
San José State University
USA
hsuanwei.chen@sjsu.edu



ABSTRACT: Online recommendations are considered as a useful approach to encourage consumer purchases and drive product sales in e-commerce. This paper uses a multimodal approach to examine the effectiveness of online recommendations using the lenses of econometric modeling and text mining. Applying data on Amazon’s “customers who bought this item also bought” recommendation, commonly referred to as the co-purchase network, this study presents several findings. First, the effects of direct co-purchase recommendations on product sales are essentially strong, regardless of other potential factors that can also affect those sales. Second, the sales of one product can have positive effect on the sales of another, when the latter is suggested in co-purchase recommendations. Moreover, the similarity of the customer base between the two co-purchased products, which can be extracted using text mining from online reviews, will affect that product’s sales. This finding leads to the conclusion that online recommendations achieve by recommending related products to like-minded consumers, and that recommendation effect is significant. These results contribute to academics and practitioners by providing clear economic evidence of the value of using text mining approaches for online recommendations.

Categories and Subject Descriptors

H.3.5 [Online Information Services]: Web-based services; **I.2.7 [Natural Language Processing]:** Text analysis; **K.4.4 [Electronic Commerce]; K.6.0 [General]:** Economics

General Terms: Economics, Management

Keywords: Online Recommendations, E-Commerce, Virtual Communities, Network Economics, Word-of-Mouth

Received: 2 February 2015, Revised 12 March 2015, Accepted 13 March February 2015

1. Introduction

The innovation of Internet technologies has created a growing opportunity for companies and their customers to engage in online business, leading to greater so-called business-to-customer (B2C) e-commerce. Online companies are now exploiting innovative techniques, such as data mining, collaborative filtering, and social network analysis to offer effective or personalized tools that encourage consumer purchase behavior and increase product sales. Among these is a well-known supportive feature of B2C websites, the *recommendation*. Prior studies have shown that online recommendations influenced consumers’ perceptions of both transactional and social orientations of a certain website [1]. [2] and [3] further argued that personalized recommendations play a significant, but mixed, role in affecting the consumer purchase experience. In addition, recommendations have been shown to be effective in creating greater customer loyalty to e-commerce websites while generating higher sales [4].

One typical approach toward providing recommendations often adopted in e-commerce websites is the item-to-item recommendation, which recommends products to consumers based on the product itself. A well-known example is the “customers who bought this item also bought” recommendation provided by Amazon.com. This type of recommendation is referred to as a co-purchase network [5][6]. In a co-purchase recommendation, a list of items is suggested in a sorted way such that the product

placed on the leftmost position is considered the most likely to be co-purchased by consumers and induce a transaction. Many recent studies on information systems have recognized the role of co-purchase networks in affecting e-commerce outcomes. For example, [5] demonstrated that the structures of a co-purchase network do affect demand in e-commerce and lead to a long tail phenomenon.

While there is a growing interest in studying how item-to-item recommendations positively influence the shopping experience and product sales, two key points were argued in prior work. First, it remains questionable how to best accurately model the effect co-purchase recommendations have on actual product sales. More specifically, given that not only the products listed in the co-purchase recommendation, but also other factors occurring on e-commerce websites can impact product sales, how can one precisely understand how the sale of one product affects that of another recommended by the online company? This problem has become a major e-commerce issue. Another point raised in the previous work is whether consumers clearly purchase products based on product awareness [7][8] or simply on the recommendations. Take the co-purchase recommendation for example. Individual consumers may co-purchase the product shown on the recommendation list. However, it is very likely that consumers co-purchase that product simply because of their general awareness of the product, regardless of the recommendations suggested by the online companies.

This paper thus proposes a multimodal approach using econometric estimation and text mining to 1) measure the sales effect one product has on another, specifically in Amazon's co-purchase network for the book market, and 2) examine whether this effect is due to online recommendations or general product awareness. To approach the first task, one notices that the direct link from one book to another as suggested in the co-purchase list is generally known to induce a sales impact on the co-purchased book. However, it is also argued that people usually ignore the fact that other books that are not apparently observed from this co-purchase network might also impact the book's sales. These "other books" correlate with both the book currently of interest and the book from co-purchase recommendations and serve as extraneous variables, leading to the issue of endogeneity [9]. How to precisely model this link effect is a major issue when evaluating the effectiveness of a co-purchase network. This study suggests that, when using book sales data from the "customers who bought this item also bought" section that is provided by Amazon.com, an instrument variable estimation model should be introduced, one aiming to examine the estimation approach so as to model the link effects of co-purchase recommendations more precisely. In particular, when the consumer is shopping for a certain book, referred to here as the *target book*, that book comes with a list of sorted recommended books that are suggested in the "customers who bought this

item also bought" section, referred to here as the *recommended books*. By introducing an instrument variable book that correlates with the target book but not with the recommended books, an instrument variable book that correlates with the target book but not with the recommended books, an instrument variable estimation model is presented, thus examining whether and how much this instrument will influence the direct link effect that the target book has on the recommended books.

The second task, which examines the roles played by online recommendations and product awareness when consumers purchase a product, is more challenging. This task is equivalent to the question about the degree the link sales effect of the target book has on the recommended books from the online recommendations. Therefore, online recommendations were further investigated by looking at the customer base of two related books. The proposed hypothesis is as follows. If two like-minded consumers are to be suggested with similar books, then the online recommendation works more effectively because it can identify similar consumers who share common interests and purchase preferences, and accordingly provide book recommendations. To understand this customer base further, the text mining approach was applied, specifically text clustering analysis, to automatically group the consumers based on their online reviews. This grouping can reveal consumer opinions, interests, or preferences pertaining to a specific book.

These findings provide two significant observations. First, when using an instrument book for model estimation, it is possible to examine the factors that are influencing product link effects on sales. The results of this study show that, although the endogeneity of target books is usually ignored, the effectiveness and value of the effects of co-purchase recommendations remain strong. This study further provides insights for practitioners. For e-commerce website providers or online companies, the findings recognize that using online recommendations will help affect product sales, thus less-weighting the possibility of these effects coming only from pure product awareness.

The rest of the paper is organized as follows. A literature review is presented in Section 2, including discussions of prior work pertaining to online recommendations, co-purchase networks, and e-commerce outcomes from such recommendations. Section 3 describes the data set adopted for this study. The text clustering approach is given in Section 4, followed by the theoretical models and estimation results in Section 5. Concluding remarks and future research opportunities are given in Section 6.

2. Literature Review

A good online recommendation is widely recognized as an effective tool for enhancing the consumer shopping experience as well as encouraging product sales for an online business. The idea of recommendation is rooted in the belief that by showing consumers compelling products,

an e-commerce website can maximize the possibility of a transaction. Generally, online recommendations include three main categories [10]. The first is a personalized recommendation, which recommends products based on the consumer's purchase histories. The second type of recommendation, called social recommendation, recommends a product to users according to the past shopping behavior of like-minded consumers. The third category, called item recommendation, recommends products to consumers based on the actual product.

One of the leading exemplars of online recommendations can be found on Amazon.com. Over the years, Amazon.com has invested money and brainpower to develop an intelligent and effective recommendation system that "taps into consumers' browsing route, purchase history, and shopping experience of other consumers," according to [11]. The item recommendation on Amazon.com is carried out through a "customers who bought this item also bought" list, referred to as "co-purchase" recommendations in this study. This recommendation is often based on basket pairings, which are considered more accurate than lifetime consumer pairings, and prior studies have shown that co-purchased items may increase the product sales [6]. Some interesting observations are offered on co-purchase recommendations by [12]. First, the items listed as "customers who bought this item also bought" are sorted in such a way that the higher rankings on the list are more likely to be involved in a transaction. Second, one of the original intentions of co-purchase recommendations was to provide products with similar titles to a customer to build an order up high enough to reach free shipping. This strategy, however, is not observed in some of the sorted "customers who bought this item also bought" list. This phenomenon suggests that Amazon.com might introduce advanced manipulations when generating the recommendation list. The idea of co-purchase recommendations is also consistent with the research in marketing over the past decades, which found that decomposing consumer utility for underlying product characteristics could significantly improve the accuracy of marketing models [13]. The genetics considered in co-purchase recommendations, therefore, is essentially a manifestation of product characteristics [14]. Figure 1 shows a typical example of such co-purchase recommendations on Amazon.com.

The sale effects induced by the product links that are generated from the co-purchase recommendations have been widely studied in the previous literature. [5] investigated how such co-purchase network structures change demand distribution by network traffic between e-commerce product pages. The authors found that when the traffic distribution generated by the network was less skewed than the intrinsic traffic distribution, such network structures evened out demand across products, leading to a demand distribution with a longer tail. The background of co-purchase recommendations also emphasizes the rationale that product sales are impacted through

the product links that are built by the "customers who bought this item also bought" list. However, as shown in prior efforts, little is understood about other factors that might also affect these co-purchased recommendation links, thus producing biased model estimation. This study improves the current literature by introducing an instrument book model estimation to evaluate the effects of co-purchase recommendations on sales.

How to measure the effects of online recommendations is not an easy task. One possible and natural assumption resulting from online recommendations is that with the help of an effective recommendation system, online companies tend to recommend related, especially motivational, products to like-minded consumers. These consumers are viewed as sharing common interests and similar purchase preferences. One way to understand consumers using the Amazon.com book market is through their online reviews [15]. Consumers post or read online reviews and gain more information on the product from different perspectives. There is broad literature and various studies that have discussed whether or not the online reviews actually affect sales [16][17][18][19]. In this study particularly, online reviews are analyzed using a text mining approach in order to understand the consumer composition, or customer base. Whether and to what degree the similarity of the customer base between the target and the recommended books impacts sales of the latter can reveal the effectiveness of online recommendations. Text mining techniques were adopted in this study to automatically cluster online reviews into groups based on their textual features [20].

3. Data

This study focuses on Amazon.com book market because by far books are a product category with the largest number of individual titles, and the product set is also relatively stable over a period of time. Daily book information on Amazon.com was thus collected from randomly selected five categories. These five categories included: 1) Arts & Photography → General; 2) Business & Money → General; 3) Children's Books → Educational; 4) Literature & Fiction → General; and 5) Mystery, Thriller & Suspense → General. Selections from these five random categories cover books from various disciplines and are thus relatively unbiased. For each of the categories, the top 500 best sellers were collected, resulting in a total of 2,500 books taken into account.

The data was collected on a daily basis and lasted for a one-week period. These 2,500 books are referred to here as *target books*, standing for those books in which consumers are currently interested. For each of the 2,500 books, the respective "customers who bought this item also bought" information was also collected, consisting of the first three books (listed from left to right) ranked in the order of the possibility of their being co-purchased from Amazon's standpoint and also suggested by Amazon.com. These three ranked books are called

The Da Vinci Code Paperback – March 31, 2009
by **Dan Brown** (Author)
★★★★☆ 5,527 customer reviews

See all 67 formats and editions

| | | |
|------------------|----------------------|----------------------------|
| Kindle \$7.22 | Hardcover \$17.38 | Paperback \$7.60 |
|------------------|----------------------|----------------------------|

Read with our free app

| | |
|-----------------------------|---------------------------|
| 3434 Used from \$0.01 | 461 Used from \$0.01 |
| 674 New from \$0.98 | 112 New from \$3.23 |
| 172 Collectible from \$1.97 | 9 Collectible from \$9.77 |

PREMIUM MASS MARKET EDITION

#1 Worldwide Bestseller—More Than 80 Million Copies Sold

As millions of readers around the globe have already discovered, **The Da Vinci Code** is a reading experience unlike any other. Simultaneously lightning-paced, intelligent, and intricately layered with remarkable research and detail, Dan Brown's novel is a thrilling masterpiece—from its opening pages to its stunning conclusion.

Customers Who Bought This Item Also Bought

Angels & Demons: A Novel
(Robert Langdon)
> Dan Brown
★★★★☆ 3,195
Paperback
\$9.95 ✓Prime

The Lost Symbol
> Dan Brown
★★★★☆ 4,684
Paperback
\$9.91 ✓Prime

Deception Point
> Dan Brown
★★★★☆ 1,422
Paperback
\$12.84 ✓Prime

Figure 1. Co-Purchase Recommendations on Amazon.com(© 2015 Amazon.com)

recommended books in this study, which were correlated with the target books and then considered to have *product links* to them. A third type of book, *instrument books*,

recent 100 consumer online reviews and their corresponding reviewers are also extracted. Notice that if some books had a total number of reviews less than 100, then the entire set of reviews was extracted.

Book Title: The title name of each book.

Sale Price: The price on Amazon.com on the day the data was collected.

Sales Rank: The sales rank or the number associated with each book on Amazon.com on the day the data was collected. This number measures the relative demand of the book compared to other products. Given that it is the sales rank, the lower the number, the higher the sales of that particular book for the day.

Co-Purchase Rank: This attribute is only associated with the three books listed as co-purchase recommendations. The rank represents the position of the product on that list. For example, if the book is listed at the leftmost position, then its co-purchase rank is assigned a 1. Hence, there were 1-3 co-purchase ranks.

For each of the target and recommended books, the most

4. Text Analysis

To understand the roles played by online recommendations and product awareness when consumers purchase a product, online recommendations need to be investigated by looking at the customer base of two related books. A text clustering analysis is thus adopted to automatically group the consumers based on their online reviews. The difficulty of dealing with web extraction data is caused by the essence of these data, which pertains to several unique characteristics. First, the scale of online textual data is growing rapidly. Second, considering each word or phrase as a “dimension” makes the data highly dimensional. Third, there are usually word and semantic ambiguities. For instance, words can contain multiple meanings (baseball bat vs. flying bat) or be synonyms (buy vs. purchase). In addition, web data are likely to be noisy, consisting of small talks, abbreviations, spelling mistakes, etc. They

are usually not well structured either. Therefore, before conducting a clustering analysis of online reviews, the following preprocessing steps were first performed, making an effort to attenuate any harm that these characteristics might bring to the study.

Stemming -Stemming is used to reduce data dimensionality by identifying a word using its stem, base, or root while at the same time reserving its semantic meaning. For example, a stemming algorithm reduces the words “flying” and “flew” to the root word, “fly”. In this study, the widely-used stemming algorithm, the Porter Stemmer [21] implemented in CLUTO [22], was adopted.

Stop Words Removal - Removing stop words is considered a crucial step toward text preprocessing. By identifying the most common words that are unlikely to help with text mining, these words can be eliminated from the data, thus reducing noisy and irrelevant information. In this study, a manually coded stop word list was provided, which contains 390 of the most common words such as “you”, “the”, “of”, “a”, etc., and then used to clean the textual data.

For a clustering analysis, textual review data are represented by the vector-space model, which is frequently used in information retrieval, collaborative filtering, and text classification [20]. CLUTO, a software tool specifically designed and used for clustering high-dimensional data, was adopted for text clustering analysis in this study, as the data was considered complex, large-scaled, and highly dimensional [22].

In this study, the clustering of online reviews was used to represent the consumer composition for each book, forming the customer base. Therefore, for each individual book i with a maximum 100 reviews clustered, a corresponding customer base B_i was formed, consisting of the K clusters $b_i^1, b_i^2, \dots, b_i^K$, in which the number of members in each cluster was $|b_i^1|, |b_i^2|, \dots, |b_i^K|$, respectively. To measure the customer similarity of two books and , the following similarity index was computed:

$$sim(i, j) = \frac{1}{\sqrt{(|b_i^1 - b_j^1|) + (|b_i^2 - b_j^2|) + \dots + (|b_i^K - b_j^K|)}} \quad (1)$$

The similarity index takes the form of the reciprocal of the Euclidean distance between the customer clusters of books i and j , meaning that the smaller the Euclidean distance of the two books’ customer bases is, the more similar these two books are (i.e., have a greater similarity index). For example, if the Euclidean distance of the customer clusters of two books is 2, then the similarity index will be 0.5. On the other hand, if the Euclidean distance of the customer clusters of two books is 4, then the similarity index will be 0.25. The two books in the former case are considered more similar than those in the latter case, as the similarity index of the former is greater.

Euclidean distance is adopted for this study, as it is being widely used in clustering problems, including clustering text [23]. It also satisfies the four conditions of a true distance metric, namely, (1) the distance between any two points must be non-negative; (2) the distance between two objects must be zero if and only if the two objects are identical; (3) the distance may be symmetric; and (4) the measure satisfies triangle inequality. Euclidean distance is also the default measure used for the K-means algorithm.

Given the calculation of the customer base similarity, Table 1 and Table 2 present the summary statistics for the analyzed data. In Table 1, the means and standard deviations for the sales ranks of all collected data are presented. Specifically, as discussed earlier, the collected data can be further divided into Target Books (the books in which consumers are currently interested in viewing), Recommended Books (books recommended by Amazon.com and ranked from left (#1) to right (#3) on the recommendation list), and Instrument Books (books derived directly from Target Books, but with a time lag 1). Table 1 shows that the means and standard deviations for the sales ranks of collected data are the same scale, regardless of what type of book to which each belongs. Table 2 presents the means and standard deviations for the similarity between all the target books and their three recommended books (ranked #1, rank #2, and rank #3, respectively). The statistics shows that while the means for the similarities are on a smaller scale, they still tend to be consistent.

| Mean (Standard Deviation) | | | | | |
|------------------------------|--------------|-------------------|---------|---------|------------------|
| | Target Books | Recommended Books | | | Instrument Books |
| | | Rank #1 | Rank #2 | Rank #3 | |
| Sales | 9.96 | 8.46 | 8.68 | 8.75 | 9.94 |
| Rank | (2.59) | (2.14) | (2.17) | (2.21) | (2.61) |

Table 1. Statistics Summary of Amazon.com Co-Purchase Data: Sales Rank

| Similarity $sim(i, j)$ | Target Books |
|---------------------------|---------------|
| Recommended Book: Rank #1 | 0.014 (0.002) |
| Recommended Book: Rank #2 | 0.013 (0.001) |
| Recommended Book: Rank #3 | 0.013 (0.002) |

Table 2. Statistics Summary of Similarity (mean, standard deviation) of Co-Purchased Books

5. Theoretical Model and Estimation Results

The need to have an instrument variable for model estimation derived from the following ideas. When considering a linear model,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2)$$

$$E(\varepsilon) = 0, \text{Cov}(x_j, \varepsilon) = 0, j = 1, 2, \dots, k-1, \quad (3)$$

where x_k might be correlated with ε . It means that the explanatory variables x_1, x_2, \dots, x_{k-1} are exogenous, while x_k is considered endogenous in (2).

An observable variable z_1 that is not in equation (2) is then needed to satisfy the following two conditions. First, z_1 should be uncorrelated with ε , that is,

$$\text{Cov}(z_1, \varepsilon) = 0 \quad (4)$$

This requirement leads to the condition wherein $x_1, x_2, \dots, x_{k-1}, z_1$ are exogenous. The second condition requires the linear projection of x_k onto all the exogenous variables, which means

$$x_k = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \dots + \delta_{k-1} x_{k-1} + \theta_1 z_1 + \gamma_k \quad (5)$$

By definition, the linear project error is equal to zero, that is $E(\gamma_k) = 0$, and (γ_k) is uncorrelated with all the explanatory variables. Notice that the key assumption for this estimation is that $\theta_1 \neq 0$ [9].

In this setting, analogously, to study the sales effect the target book has on each of the three recommended books ranked from the co-purchased network, the sales rank of the target book is the explanatory variable, denoted as S_t . The sales rank of the recommended book of rank k ($k = 1, 2, \dots, 3$) is regarded as the dependent variable and denoted as S_r^k . Therefore, the preliminary estimation model for recommended book of rank k can be stated as

$$\log(S_r^k) = \beta_0 + \beta_1 \log(S_t) + \varepsilon \quad (6)$$

However, it is assumed that the error term may be correlated with because of other factors on the e-commerce website. One natural factor taken into account is that there exists an omitted book that also affects the sales of the target book, S_t , but that book is not obviously observable. This book is considered an instrument book, and its sales rank is denoted as S_i . Furthermore, for such a book to be a valid instrument for $\log(S_t)$, it is assumed that the logarithm for the sales of this book is uncorrelated with and that $\theta_1 \neq 0$ in the following equation:

$$\log(S_t) = \delta_0 + \theta_1 \log(S_i) + \gamma \quad (7)$$

Based on these arguments, in this study, this omitted book

is collected from the pool of the books that correlate with the target books but not with the recommended books. One simple candidacy set contains the target book sales with a time lag equal to 1. That is, at time p ,

$$\log(S_t^p) = \delta_0 + \theta_1 \log(S_t^{p-1}) + \gamma \quad (8)$$

Applying the model in equation (6) to the co-purchased book data from Amazon.com, the ordinary least squares (OLS) method was first performed for estimating the sales of the recommended books ranked 1, 2, and 3, which are assumed to be affected by the sales of the target books, but without considering the issue of instrument books. The OLS method estimates the unknown parameters in a linear regression model while minimizing the differences between the observed and predicted responses [9]. The instrument books are then introduced to the model, leading to a two-stage least squares (2SLS) estimation. 2SLS is the extension of the OLS method that is used when the dependent variable's error terms are correlated with the independent variables [9], which applies to the case here. The comparative results are shown in Table 3.

These results suggest the following observations. By using instrument books for the model estimation, the sales effects (coefficients and significance) of the target book on the recommended books ranked 1, 2, or 3 remain unchanged. This is an interesting finding, and it shows that although there is often a concern of whether other factors on an e-commerce website are impacting product sales concurrently and changing the effects of co-purchase recommendations, there is indeed a reason for people to believe that direct product links from the target book to the recommended books can strongly reflect the impacts on product sales. This finding also complies with the value of co-purchase recommendations that is usually recognized.

Based on the above results, the original OLS estimation model is further examined. Recall that not only the sales effect of the target book on the recommended books was considered. In this study, moreover, it is also investigated whether the similarity of the customer base of the target and the recommended books impact the sales of the latter. In other words, in this model, another explanatory variable should be introduced, i.e., the similarity of target book t and recommended book r of rank, k denoted as $\text{Sim}(t, r_k)$, which can be computed in equation (1). That is,

$$\log(S_r^k) = \beta_0 + \beta_1 \log(S_t) + \beta_2 \log \text{Sim}(t, r_k) + \varepsilon \quad (9)$$

Applying model in equation (9) to the co-purchased book data, the estimation results are given in Table 4.

In Table 4, two effects on the sales of the recommended books are estimated: 1) the sales of the target book and 2) the similarity of the customer base between the target and recommended books. Two interesting patterns appear. First, both the sales of the target book and the similarity of the customer base have a significant positive effect on

the sales of the recommended books. This finding is consistent with the general understanding and motivation for the use of co-purchase recommendations: the *product links* are presented for consumers to relate different products together, thus aiming to increase the visibility of these products and also encourage shopping transactions. In addition, the “customers who bought this item also bought” list is designed to provide consumers with effective and intelligent recommendations. Although the mechanisms of how Amazon.com creates such recommendation lists are not publicly known, it is natural to assume that a successful recommendation should be able to provide similar, like-minded consumers with correlated products. This assumption is further verified in the proposed model, namely, the similarity of customer base between the target and recommended books does have a significant effect on product sales of the recommended books.

Another pattern shown in Table 4 is that the lower the rank (in terms of the listed position) of the recommended book in the “customers who bought this item also bought” list, the less influential the sales of the target book as

well as the similarity of the customer base between books will be on product sales. This is an interesting finding and in fact is consistent with the nature of consumer purchase behavior and the design of co-purchase recommendations. The recommended book that ranked first at the leftmost position in co-purchase recommendations is usually considered more likely to be explored and purchased, given that the book is intentionally designed by Amazon.com to become the strongest recommendation, such that it will match consumers’ purchase interest, and be the most visible and reachable product at first sight. As a result, consumers’ purchasing decisions are more affected by co-purchase recommendations if those co-purchased products are ranked higher. In other words, co-purchased products that are listed lower in the recommendation list may not catch consumers’ attentions or match their purchase interests as much as higher-ranked products will. Table 4 provides economic evidence to support these arguments.

The model is further modified to examine to what degree the sales effect is due to online recommendations by looking at the similarity of the customer base for two

| Recommended Book | Variable | Coefficient | OLS | 2SLS |
|------------------|-------------|-------------|----------|----------|
| | | | Estimate | Estimate |
| Rank #1 | Constant | β_0 | 8.089*** | 8.094*** |
| | $\log(S_t)$ | β_1 | 0.038* | 0.037* |
| Rank #2 | Constant | β_0 | 8.266*** | 8.235*** |
| | $\log(S_t)$ | β_1 | 0.042** | 0.045** |
| Rank #3 | Constant | β_0 | 8.332*** | 8.308*** |
| | $\log(S_t)$ | β_1 | 0.043** | 0.045** |

Significance level: *** $p < 0.01$; ** < 0.05 ; * < 0.1

Table 3. Estimation Results for Sales of Recommended Books Ranked 1, 2, and 3: OLS and 2SLS

| Recommended Book | Variable | Coefficient | Estimate |
|------------------|--------------------|-------------|-----------|
| Rank #1 | Constant | β_0 | 23.674*** |
| | $\log(S_t)$ | β_1 | 0.109*** |
| | $\log Sim(t, r_1)$ | β_2 | 3.837*** |
| Rank #2 | Constant | β_0 | 21.495*** |
| | $\log(S_t)$ | β_1 | 0.097*** |
| | $\log Sim(t, r_2)$ | β_2 | 3.243*** |
| Rank #3 | Constant | β_0 | 19.929*** |
| | $\log(S_t)$ | β_1 | 0.090** |
| | $\log Sim(t, r_3)$ | β_2 | 2.841*** |

Table 4. How Target Sales and Customer Base Similarity Affect Recommended Product Sales

books. In order to compare β_1 to β_2 in model equation (9) for their effects on the sales of recommended book of

rank k , equation (9) is re-parameterized as follows:

$$\log(S_r^k) = B_0 + B_1 z_1 + B_2 z_2 + u \quad (10)$$

| Recommended Book | OLS | | |
|------------------|----------|-------------|-------------|
| | Variable | Coefficient | Coefficient |
| Rank #1 | Constant | B_0 | 23.273 |
| | z_1 | B_1 | 3.845 |
| | z_2 | B_2 | 3.631* |
| Rank #2 | Constant | B_0 | 21.995 |
| | z_1 | B_1 | 3.369 |
| | z_2 | B_2 | 3.171* |
| Rank #3 | Constant | B_0 | 19.929 |
| | z_1 | B_1 | 2.931 |
| | z_2 | B_2 | 2.751* |

* Significantly larger than zero at a 95% confidence interval.

Table 5. Degree of the Effect of Online Recommendations on Sales of Recommended Books Ranked 1, 2, and 3.

where $z_1 = \log Sim(t, r_k) + \log S_t / 2$ and $z_2 = \log Sim(t, r_k) - \log S_t / 2$. Therefore, to examine whether online recommendations are a significant determinant of product sale effects, B_2 in equation (10) is tested whether it is significantly larger than zero. Applying model (10) to the collected, co-purchased book data, including the target books and the recommended books ranked 1-3, Table 5 shows the estimation results using OLS regression. Table 5 illustrates to what degree the effects on the sales of recommended books are due to a similarity of customer base. This observation reveals whether online recommendations significantly determine sales effect. In other words, the weight that customer base similarity contributes to sales effect was examined. The results show that the similarity of the customer bases for the target and recommended books, which reveal the effectiveness of co-purchase recommendations, does play an important role in affecting product sales of the recommended books. Moreover, this importance remains existent even with lower ranked recommended books. While for lower ranked books, as suggested in co-purchase recommendations, the impacts of sales of the target book and the similarity are claimed to be less, the similarity of the customer base of books still holds a crucial effect, leading to the finding that online recommendations do matter when suggesting co-purchased books.

6. Conclusions and Future Directions

This paper combines econometric modeling and text mining techniques to provide a multimodal analysis and evidence from “customers who bought this item also bought” recommendations on Amazon.com by examining the effects of 1) the sales of one book on another to which a product link is suggested by Amazon.com as a connect,

and 2) the impacts of customer base similarity for two books on the sales of one of them. Upon econometric model estimations, three interesting observations emerge. First, by introducing an instrument variable estimation, the effect that the direct product links have on sales is essentially strong. More specifically, while other factors (e.g., the instrument books taken into account in the study) might also affect product sales in a direct or indirect way and which has been an issue when evaluating co-purchase recommendations, the influence of direct co-purchase recommendations is still considered dominant. Second, the study of the similarity of the customer base for two related books helps to understand whether or not these online recommendations take effect by suggesting related connections in terms of purchase, i.e., books to like-minded consumers. This result suggests that both product sales and customer base similarity significantly affect sales of another product that is linked, or related, as suggested by Amazon.com. The third observation helps one understand to what degree consumers are affected by online recommendations to perform the designated purchase behavior. These findings point out that online recommendations do play an important role in influencing consumers in the course of online business. More importantly, this result indicates that it is online recommendations, rather than consumers’ product awareness, that drive the purchase behavior when recommendations are available, such as in the case of any “customers who bought this item also bought” list.

This study contributes to academics by introducing an econometric perspective to the understanding of the effects of online recommendations on product sales. In particular, the combination of econometric analysis and data mining techniques helps unravel the problem using an

econometric model, but with hidden, previously unobservable information, which was extracted, or *mined*, using text mining. This study also recognizes the effectiveness of online recommendations, which thereby can help practitioners such as online companies to verify, examine, and understand their designs for recommendation and that value.

These research findings also open up several opportunities for future directions and study. First, future research can validate and extend the findings of this study for other forms of online recommendations, such as “What Other Items Do Customers Buy After Viewing This Item?,” a type of social recommendation found on Amazon.com. Secondly, the idea of examining the value of co-purchase recommendations using an instrument variable could be further extended. More complicated mechanisms involving the endogeneity of such variables could also be investigated. Indeed, the selection of an instrument variable is an important issue and definitely deserves further consideration and examination.

Acknowledgement

The author would like to thank Dr. Bin Gu and Dr. Prabhudev Konana for providing valuable insights and suggestions in the preparation of this paper.

References

- [1] Kumar, N., Benbasat, I. (2006). The Influence of Recommendations and Consumer Reviews on Evaluations of Websites. *Information Systems Research*, 17 (4) 425-429.
- [2] Senecal, S., Nantel, J. (2004). The Influence of Online Product Recommendations on Consumers' Online Choices. *Journal of Retailing*, 80, 159-169.
- [3] Shen, A. (2014). Recommendations as Personalized Marketing: Insights from Customer Experiences. *Journal of Services Marketing*, 28 (5) 414-427.
- [4] Ansari, A., Essegai, S. Kohli, R. (2000). Internet Recommendation Systems. *Journal of Marketing Research*, 37, 363-375.
- [5] Oestreicher-Singer, G. Sundararajan, A. (2012). Recommendation Networks and the Long Tail of Electronic Commerce, *MIS Quarterly*, 36 (1) 65-83.
- [6] Basuchowdhuri, P., Shekhawat, M. K. Saha, S. L. (2014). Analysis of Product Purchase Patterns in a Co-Purchase Network. *2014 Fourth International Conference of Emerging Applications of Information Technology*, Dec. 19-21, 355-360.
- [7] Macdonald, E. K., Sharp, B. M. (2000). Brand Awareness Effects on Consumer Decision Making for a Common, Repeat Purchase Product: A Replication. *Journal of Business Research*, 48, 5-15.
- [8] Chi, H. K., Yeh, H. R. Yang, Y. T. (2009). The Impact of Brand Awareness on Consumer Purchase Intention: The Mediating Effect of Perceived Quality and Brand Loyalty. *The Journal of International Management Studies*, 4 (1) 135-144.
- [9] Wooldridge, J. M. (2012). *Introductory Econometrics: A Modern Approach*. Cengage Learning.
- [10] Brusilovski, P., Kobsa, A., Nejdl, W. (2007). *The Adaptive Web: Methods and Strategies of Web Personalization*. Springer.
- [11] Iskold, A. (2007). The Art, Science and Business of Recommendation Engines. http://readwrite.com/2007/01/16/recommendation_engines
- [12] Jannach, D., Zanker, M., Felfernig, A., Friedrich, G. (2010). *Recommender Systems: An Introduction*. Cambridge University Press.
- [13] Green, P. E., Devita, M. T. (1974). A Complementary Model of Consumer Utility for Item Collections. *Journal of Consumer Research*, 1, 56-67.
- [14] Vafopoulos, M. N., Theodoridis, T., Kontokostas, D. (2011). Inter-Viewing the Amazon Web Salespersons: Trends, Complementarities and Competition. <http://ssrn.com/abstract=1850360>
- [15] Chen, H. (2012). The Impact of Comments and Recommendation System on Online Shopper Buying Behavior. *Journal of Networks*, 7 (2) 345-350.
- [16] Hu, N., Liu, L., Zhang, J. (2008). Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects. *Information Technology Management*, 9 (3) 201-214.
- [17] Godes, D., Mayzlin, D. (2004). Using Online Conversations to Study Word-of-Mouth Communication. *Marketing Science*, 23 (4) 545-560.
- [18] Floyd, K., Freling, R., Alhoqail, S., Cho, H. Y. Freling, T. (2014). How Online Product Reviews Affect Retail Sales: A Meta-analysis. *Journal of Retailing*, 90 (2) 217-232.
- [19] Zhu, F., Zhang, X. (2010). Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, 74, 133-148.
- [20] Feldman, R., Sanger, J. (2006). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.
- [21] Jones, K. S., Willet, P. (1997). *Readings in Information Retrieval*. Morgan Kaufmann.
- [22] CLUTO – Software for Clustering High-Dimensional Datasets.
- [23] Huang, A. (2008). Similarity Measures for Text Document Clustering. *The 2008 Proceedings of the New Zealand Computer Science Research Student Conference*, 49-56, Christchurch, New Zealand.