How Consumers' Content Preference Affects Cannibalization: An Empirical Analysis of an E-book Market

Research-in-Progress

Kyunghee Lee
KAIST College of Business
85 Hoegiro Dongdaemoon-gu Seoul, Korea
kyunghee.lee@kaist.ac.kr

Kunsoo Han
McGill University
1001 Sherbrook St. W.
Montreal, QC H3E 1X8
Canada
kunsoo.han@mcgill.ca

Eunkyoung Lee
KAIST College of Business
85 Hoegiro Dongdaemoon-gu Seoul, Korea
greenpray@business.kaist.ac.kr

Byungtae Lee
KAIST College of Business
85 Hoegiro Dongdaemoon-gu Seoul, Korea
btlee@business.kaist.ac.kr

Abstract

Despite the increasing popularity of e-books and the growing maturity of e-book markets, there have been few studies on the e-book channel and its influence on the existing paper book channel. In order to fill this gap in the literature, we investigate the extent to which e-book release boosts or cannibalizes demand for the paper book. Using unique data on the actual sales of paper books and e-books, we conduct an empirical analysis on this question. Our results without addressing selection bias suggest that the e-book release boosts the demand for the paper book. However, this effect disappears once we control for selection bias by using matching. We also find that the impact of e-book release is moderated by consumers' content preference. Specifically, the e-book release increases paper book sales as well as total sales for those books with light contents that consumers prefer to consume through a digital channel. In contrast, the books that appeared on the bestseller lists experience significant demand cannibalization from e-book release.

Keywords: E-business, channel competition, consumer preference, e-book market, cannibalization

Introduction

With the availability of diverse electronic contents and dedicated e-book readers, e-books are gaining popularity and have taken over paper books in terms of sales in mature markets, thereby shifting the competitive dynamics of the existing book market. For example, in the US market, Amazon is now selling 180 e-books for every 100 hardcovers,[1] and this ratio is expected to further increase in the future. With the increasing significance of e-books as an alternative channel for consuming contents, it is important to understand the impact of the e-book channel on the existing paper book channel.

Publishers' motivation to release e-book is two-fold. First, by releasing an electronic version of a given book, publishers can advertise or increase the awareness of the book so that more consumers learn about the book and purchase either a print version or an e-book (Hu and Smith 2013). When consumers learn about the book through the release of e-book and purchase a print version, e-book release will boost print book sales. In contrast, releasing an e-book will cannibalize the paper book sales if consumers who would otherwise buy (more expensive) paper version end up purchasing (less expensive) electronic version. Second, publishers release e-books in order to tap the segment of consumers who have strong preference for e-books over paper books. Given that these consumers may already know about the book but have not bought a paper version because of their preference for digital content, publishers can increase their revenue by selling an electronic version to them.

To date, however, there have been few studies on e-books, especially on their impact on the existing print book channel with an exception of Hu and Smith (2013). This paucity of research on this important topic may be due to the lack of necessary data. To estimate the cross-channel impact, one needs actual book-level demand data. Moreover, the sample period needs to be long enough so that it is possible to compare the demand for paper books before and after e-book release. The number of sample book titles also needs to be large to ensure the representative of the sample. Using actual sales data from 1,828 randomly selected book titles over four years (2010-2013), we empirically examine the impact of e-book release in a quasi-experimental setting. Specifically, the richness of our data allows us to use matching and difference-in-differences (DID) analysis, thereby alleviating the concern for selection bias.

We find that on average e-book release does neither cannibalize nor boost the demand for paper books. We also find that the aggregated demand for the paper and electronic versions of a book experiences a significant increase following the e-book release. These results run counter to publishers' prevalent concern that their revenue would be damaged by e-book availability. More importantly, we show that the impacts of e-book release are moderated by consumers' digital content preference. Specifically, our results suggest that in book categories that users prefer to read as an e-book (e.g., novels), e-book release has a positive impact on the paper book sales as well as the total sales (paper book and e-book). In contrast, for bestseller books, e-book release significantly hurts both the paper book sales and total sales.

This study contributes to the literature by demonstrating that whether a new channel cannibalizes or complements the existing channel critically depends on consumers' preference toward the channels. Moreover, our study is among the first to examine the cross-channel dynamics in the e-book market using large-scale sales data, and contributes to the nascent literature on the e-book market. On a practical front, our findings can help publishers assess the impact of e-book release on their bottom line and make strategic choices regarding e-book release in such a way that maximizes the benefit (while minimizing the cannibalization) from the release of e-books.

The rest of the paper is organized as follows: Section 2 surveys relevant studies to lay the theoretical foundation of the current study. In Section 3, data and analytical strategies are discussed. Section 4 provides the preliminary results of the empirical analyses. Section 5 discusses implications of the findings and the next steps.

---

[1] http://www.wired.com/2010/07/amazon-more-e-books-than-hardcovers/

Theoretical Background

Cross-channel competition

Prior studies have examined whether a new channel cannibalizes or boosts the demand in the existing channel. Moorthy and Png (1992) investigates the timing of product introductions and its impact on cannibalization. They show that the lower quality product would cannibalize demand for the high quality product, and this magnitude could be reduced by decreasing the price of the high-quality product or delaying the release of the low-end model. In the context of the newspaper industry, Deleersnyder et al. (2002) show that the advent of online newspapers cannibalizes the demand for physical newspaper. They find that the degree of cannibalization, however, is relatively small, suggesting that the fear of the online channel in the industry has been somewhat overstated. Similarly, Danaher et al. (2010) find that the availability of the same content on iTunes does not harm DVD sales.

Enabled by the Internet, digital piracy has been considered a form of a new distribution channel. Although, in general, piracy negatively impacts the demand for physical channels (Danaher et al. 2013), some studies found that the magnitude of displacement of physical sales from digital piracy is insignificant. For example, file sharing does not reduce the legal sales of music (Oberholzer-Gee and Strumpf 2007) and the availability of pirated content has no effect on DVD sales (Smith and Telang 2009).

The aforementioned studies suggest that the impact of a new channel on the existing channel hinges on the distinct characteristics of the two channels, such as customer segments and product selection. Related to this point, Brynjolfsson et al. (2009) find that Internet retailers encounter severe competition from brick-and-mortar retailers when selling popular products, whereas the Internet retailers experience statistically zero competition when selling niche products. This is because customers in the offline stores experience higher search costs than online counterparts, which results in different product selection across the channels. In a similar vein, Vernik et al. (2011) demonstrate that the digital format of music (i.e., MP3 or AAC files) is more appreciated than the traditional format (i.e., CDs) by consumers, thus this preference can asymmetrically promote the sales for the digital contents. Based on this, we argue that the impact of e-book channel on paper book channel is essentially an empirical question and that this impact will vary across books depending on their distinct characteristics.

E-book market

Despite of the rapid growth of the e-book industry, there have been only a few studies on e-book. Hao and Fan (2014) investigate the optimal pricing strategies for publishers and retailers in the e-book channel, and show that publishers are worse off under the agency model. We found two empirical studies that pose similar research questions as our study (Bounie et al. 2013; Hu and Smith 2013). Bounie et al. (2013) show that the sales of the books written by successful authors are more likely to be cannibalized by e-book sales. However, given that they use monthly top 100 lists to infer the demand and the history of rank for a particular book cannot be fully tracked in this data, it is difficult to make a causal argument. In addition, due to data limitations, the authors capture the degree of cannibalization by the probability that a book that was in the bestseller print book list also included in the bestseller e-book list. Using actual sales data, our study examines the degree of cannibalization more directly.

The most closely related to our study is Hu and Smith (2013) who use a natural experiment to find that delaying e-book availability results in a small, yet insignificant increase in paper book sales, but a large decrease in e-book sales. They use a publisher's decision to remove its e-book contents from Kindle and restore the contents after two months an exogenous event. Although the research questions are similar, our study is different from Hu and Smith (2013; HS hereafter) in several important ways. First, while HS examine the impact of the timing (delay) of e-book release, we investigate the impact of e-book release (not the timing) using large-scale data that cover the periods before and after the e-book release. Second, while HS use data on books that have both print and electronic versions, our data include books with print version only as well as those with both paper and electronic versions. By constructing a control group that consists of the books of which e-book version has never been released, we address the selection bias— publishers may choose to release e-books only for those titles that can benefit from the release of an electronic version. Third, we provide additional insights by identifying book- and publisher-specific factors that influence the decision to release e-books; it is noteworthy that these factors explain more than

50% of the variance of the decision. Fourth, because our data cover a large portion of the Korean e-book market whereas HS uses data from a single publisher, we can assess the impact of e-book release at a much broader level. Finally, while HS use data from a mature US e-book market, we analyze data from the Korean e-book market that is still at a relatively early stage of development; therefore, our study provides implications relevant to publishers and retailers operating in emerging e-book markets.

Methods

Data description

We obtained large-scale data from Kyobo, the largest book retailer in Korea. Starting as a bricks-and-mortar company, Kyobo has expended their business boundary to the online market and digital contents. The e-book sales account for approximately 3% of paper book sales of the company, which shows that the Korean e-book market is in its infancy. To stay ahead of competition in the e-book market, Kyobo has introduced a series of own e-book readers and has also tried to increase the amount of e-book contents by persuading publishers who have been reluctant to produce digital versions of their books. In 2013, Kyobo had approximately 130,000 titles in e-book version and the annual e-book sales in 2012 increased by roughly 40 percent compared with the previous year.

We obtained weekly sales records of the 1,828 randomly selected book titles of which the e-book version was released between January 2011 and December 2013. The total number of e-books available in Kyobo bookstores is nearly 130,000, but excluding the titles without a paper version left approximately 40,000 book titles. Out of these titles, we randomly selected 1,828 titles that constitute the sample for our analysis. In our data, the e-book release dates are distributed over three years from 2011 to 2013. Note that we collected the paper book sales data from January 2010, one year prior to the earliest e-book release date in order to ensure that every book has an enough pre-release period to compare the demand between before and after e-book release.

Table 1 describes the variables contained in the data set. Demand is captured in two variables: PAPER_SALES represents the paper version sales for book i in week t; and TOTAL_SALES is the aggregated sales (both paper book and e-book) for book i in week t. PRICE represents price for the print version of book i in week t. AGE indicates the number of days from the release of the paper version of book i until week t. NTITLES represents the total number of book titles published by the publisher of book i. BEST is a dichotomous variable indicating whether or not book i was on the monthly top 200 bestseller list during the sample period at least once. In order to estimate the effect of the e-book release on sales, we create a dichotomous variable indicating the post-e-book-introduction period. Specifically, RELEASE equals to one after the e-book release date, and zero otherwise.

Table 1. Summary statistics					
Variable	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
PAPER_SALES	246,494	4.099	24.472	0	997
TOTAL_SALES	246,494	4.306	24.944	0	1,000
RELEASE	246,494	0.475	0.499	0	1
E-PREFER	246,494	0.322	0.467	0	1
BEST	246,494	0.062	0.241	0	1
PRICE	246,494	10,428.850	8,028.731	0	270,480
AGE	246,494	988.974	917.788	0	6,407
NTITLES	246,494	842.707	1,090.875	0	7,276

According to the interviews with the managers of Kyobo, consumers' preference for e-book varies across book categories, which may lead to differential cross-channel impacts across categories. In particular, our interviews revealed that a specific category containing romance, martial arts, and fantasy novels is much more strongly preferred as an e-book by consumers, compared to the other categories. Thus, we create a

dummy, E-PREFER, that represents this category.

Empirical model

To estimate the impact of e-book release, we first consider the following empirical model:

$$y_{it} = \alpha + \mathrm{X}\beta + \gamma I_{it} + v_i + \delta_t + \varepsilon_{it} \tag{1}$$

where y_{it} denotes two types of demands (PAPER_SALES and TOTAL_SALES); I_{it} is our main parameter, RELEASE. X represents the explanatory variables including PRICE, AGE, E-PREFER, BEST, and NTITLES. In addition to the explanatory variables, we include v_i, category fixed effects, and δ_t, year (2010, 2011, 2012, and 2013) and month (January - December) fixed effects.

We are primarily interested in the coefficient of RELEASE. A positive (negative) coefficient would imply that releasing an e-book version boosts (cannibalizes) the paper book sales. As mentioned earlier, we conjecture that this effect is moderated by consumer's content preference. Specifically, for those categories of books that consumers prefer to read as an e-book, the impact of e-book release may be different from other categories. E-PREFER category consists of romance, martial arts, and fantasy novels whose content can be lightly consumed. We include the interaction term, RELEASE×E-PREFER, to capture the moderating effect of this category.

To further address unobserved book-level heterogeneity, random- and fixed-effect models are considered. Random effect model is more efficient in the estimation of parameters and is also able to deal with time-invariant variables while it imposes an assumption that the book-specific effects are orthogonal to the other independent variables. On the other hand, fixed-effect model does not require this assumption, but has an issue in dealing with time-invariant variables. Although the result of Hausman test favors fixed-effect model, we report the results from random-effects model as well, because we have time-invariant variables, E-PREFER and BEST, that are important in our analyses.

An important empirical issue to consider in our data is selection bias. In other word, publishers may choose to release e-books only for those titles that can benefit from the release of an electronic version. For instance, a small-sized publisher may want to avoid competition with the incumbents in the trade book category and target a niche market; this publisher would strategically release e-book versions of the niche products. Such selection may jeopardize the validity of our results, and we address this issue by using a matching method where each sample is matched with subjects in non-experimental control group that have similar observable attributes (Heckman et al. 1997, 1998). We collected additional data on book titles for which an electronic version has never been released within our sample period, and used it for constructing the control group. Specifically, we randomly chose 6,000 book titles out of the population of the Korean book market. Then, we modeled e-book release decision as follows:

$$\Pr(ebook = 1 \mid X) = \Phi(X'\beta) \tag{2}$$

where Pr denotes probability; Φ is the CDF of normal distribution; and X involves price, the number of pages, book size, age, publisher's sales information, bestseller, and book categories. Based on these observable attributes, for each book the model estimates the propensity of being "treated," that is, releasing e-book version. Then, each treated book is paired with a non-treated book that is identical to the treated book in terms of the propensity scores. In this way, the treated group can be directly compared with its counterfactual samples, which allow us to precisely measure the impact of e-book release. Then, we estimate the following DID model to estimate the impact of e-book release:

$$Demand_{it} = \alpha + \beta_0 Treated_i + \beta_1 Release_{it} + \gamma Treated_i \times Release_{it} + X_{it} + \mu_i + \tau_t + \varepsilon_{it} \tag{3}$$

Treated equals one if a book belongs to the treatment group (with both paper and electronic versions), and zero if it belongs to the matched control group (with a paper version only). *Release* equals one after the release of an e-book version for book *i*, and zero before the release date. Note that because books in the control group do not have an e-book release date, we use the release date of its matched sample in the treatment group. For example, if for a treated book *k*, an electronic version is released on December 12, 2012, then we apply the same release date for the one-to-one matched book *k´* in the control group even though the latter does not have an electronic version.

The coefficient on the interaction term, *Treated×Release*, is of our primary interest and captures the impact of e-book release on demand in the treatment group relative to the control group. If γ is positive (negative), it indicates that e-book release has a positive (negative) impact on the demand of books in the treatment group as books in the control group that have the same propensity score for having an electronic version.

Results

Impacts of e-book release on demand (without accounting for selection bias)

Table 2 represents the results of estimating the impact of e-book release based on equation (1). Columns 1-4 show the results of random-effects models and columns 5-8 present the results of fixed-effects model. We are mainly interested in the coefficient of RELEASE, which indicates the period after e-book versions are released. The coefficient of RELEASE is positive and significant (*t*=3.47, *p*<0.001), suggesting that releasing e-book versions increases the paper book sales. This result is opposed to the conventional belief that e-book can damage the demand for paper books (Hu and Smith 2013). In Model (2) where the dependent variable is the total sales from both paper and e-book versions, we get essentially the same result although the magnitude of the effect is much greater. Specifically, the coefficient of RELEASE in column (1) suggests around 1% increase ($e^{0.013} = 1.013$) in demand after the e-book release while in column (2) this magnitude is amplified to 20% ($e^{0.185} = 1.2032$).

Table 2. The results of RE and FE model with time fixed effects								
	Random-Effects				Fixed-Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	lnPAPER_SALES	lnTOTAL_SALES	lnPAPER_SALES	lnTOTAL_SALES	lnPAPER_SALES	lnTOTAL_SALES	lnPAPER_SALES	lnTOTAL_SALES
RELEASE	0.013*** (3.47)	0.185*** (46.65)	-0.088*** (-20.80)	0.028*** (6.32)	0.013*** (3.56)	0.186*** (46.80)	-0.088*** (-20.77)	0.029*** (6.40)
E-PREFER	-0.926*** (-16.78)	-0.858*** (-15.38)	-1.067*** (-19.32)	-1.062*** (-19.03)				
RELEASE× E-PREFER			0.297*** (59.54)	0.431*** (81.13)			0.297*** (59.59)	0.431*** (81.17)
RELEASE × BEST			-0.986*** (-58.44)	-0.959*** (-53.29)			-0.989*** (-58.56)	-0.963*** (-53.42)
BEST	2.153*** (25.05)	2.159*** (24.87)	2.781*** (32.12)	2.771*** (31.66)				
lnAGE	-0.256*** (-129.00)	-0.251*** (-118.11)	-0.255*** (-130.35)	-0.252*** (-121.08)	-0.256*** (-127.24)	-0.250*** (-116.06)	-0.254*** (-128.48)	-0.251*** (-118.90)
lnPRICE	-0.273*** (-32.96)	-0.266*** (-29.97)	-0.275*** (-33.73)	-0.269*** (-30.93)	-0.288*** (-34.26)	-0.283*** (-31.31)	-0.289*** (-35.02)	-0.284*** (-32.24)
Constant	4.672*** (23.31)	4.562*** (22.30)	4.739*** (23.72)	4.673*** (22.95)	4.639*** (57.08)	4.563*** (52.31)	4.653*** (58.27)	4.593*** (53.91)
Category fixed effects	Included	Included	Included	Included				
Book fixed effects					Included	Included	Included	Included
Time fixed effects	Included	Included	Included	Included	Included	Included	Included	Included
N	220,391	220,369	220,391	220,369	220,391	220,369	220,391	220,369
R-sq	0.345	0.315	0.351	0.326	0.724	0.692	0.734	0.706
Adjusted R-sq					0.722	0.689	0.732	0.704
Note: t statistics in parentheses; * p<0.05, ** p<0.01, *** p<0.001; Overall R-sq for the random effect models								

In column (3), we add two interaction terms to capture the moderating effects of E-PREFER and BEST.

The coefficient of the interaction term, E-PREFER×RELEASE, is positive and significant (t=59.54, p<0.001), suggesting that books in E-PREFER category account for a substantial portion of the positive demand impact from e-book release. Quite contrary to E-PREFER, best-selling books (BEST) show a large and negative impact of e-book release on demand (t=58.44, p<0.001). This indicates that the release of e-book versions of best-selling books cannibalizes the paper book sales. Since we control for the moderating effect of E-PREFER and BEST, the negative coefficient of RELEASE indicates that the average impact of e-book release for the book categories excluding E-PREFER and BEST is negative (cannibalization); e-book release is associated with around 9% decrease ($e^{-0.088} = 0.9157$) in the paper book sales.

When considering the total sales (column (4)), the coefficient of the interaction between RELEASE and E-PREFER becomes greater than that in column (3). Moreover, the coefficient of RELEASE becomes positive and significant (t=6.32, p<0.001), suggesting that the total sales for books that do not belong to E-PREFER and BEST increases by roughly 3% after releasing their e-book version.

Other control variables show the expected signs and these results are consistent across all the models. The negative coefficient of E-PREFER indicates that books in this category are less popular than others. Even in terms of the total sales combining paper book and e-book versions, E-PREFER category underperforms other categories. In addition, best-selling books, obviously, have a larger demand than other books, suggested by the positive sign of BEST. As expected, PRICE is negatively associated with the demand, and the number of days since the paper book release date, lnAGE, is negatively associated with the demand, suggesting that the demand for a book decreases over time.

While the above results are interesting, they may be subject to selection bias. Next, we present the results of estimating the impact of e-book release based on matching and difference-in-differences analysis in order to address potential selection bias.

Matching and difference-in-differences analysis

For matching, we first estimate a selection (probit) model described earlier (see equation (2)). The results of estimating the probit model are represented in Table 3. Coefficients represent the impact of a given variable on the likelihood of a book to be released in an electronic version. For example, the negative coefficient of *lnPrice* means that the higher the price of a book, the lower the likelihood of the e-book version of this book is released. The larger the number of pages a book has, the lower the likelihood of the book having an electronic version. Moreover, older books have a higher probability of having an e-book version than new ones. Best-selling books are also more likely to be released in e-book than less popular books.

Table 3. The e-book release decision model		
	e-book release	
lnPrice	-0.937***	(-10.66)
lnNum_page	0.176***	(2.81)
lnAge	0.103***	(3.51)
lnPub_pQuantity	-0.365***	(-21.18)
lnPub_eQuantity	0.407***	(29.49)
Bestseller	0.640***	(4.12)
Constant	7.379***	(8.96)
Book size dummy	Included	
Book category dummy	Included	
N	5,728	
Pseudo R-sq	0.563	
Note: *t* statistics in parentheses; * *p*<0.10, ** *p*<0.05, *** *p*<0.01		

In addition to the book attributes, the model incorporates two publisher-specific factors. *Pub_pQuantity* is the total number of paper books a given publisher has published; its negative sign suggests that large publishers with a large number of paper book publications are less likely to publish e-books. *Pub_eQuantity* captures the total number of e-books published by a given publisher, and its positive sign indicates that publishers having published more e-books are more likely to publish e-books. The Pseudo-R2 of the decision model is 0.563, which suggests that the book and publisher attributes included in the model explain more than 50% of the variance of publishers' decision to release an e-book in our data.

Using the estimation results, we calculated propensity score for each book in the treatment group as well as control group. Based on the scores, each book in the treatment group was matched with a book in the control group on a one-to-one basis. Out of the 1,828 treated book titles used in our initial analysis, we failed to find matched sample for 270 book titles. After removing these titles, our final sample consists of 1,558 matched pairs. After matching, there is no significant difference between treatment and control groups in terms of the book and publisher attributes used in the selection model. This further shows that our matching effectively removed any difference in observable attributes between the two groups.

Table 4 presents the results of estimating the DID model (equation (3)) using the matched samples. As explained above, we are primarily interested in the coefficient of *Release×Treated*, which captures the average treatment effect of e-book release on sales. As can be seen from columns (1) and (3), this coefficient is not statistically significant. This suggests that after controlling for the selection bias, the positive impact of e-book release on paper book sales disappears (see Table 2). This confirms our conjecture that publishers may decide to release e-book version for those books that can benefit from the release. As shown in columns (2) and (4), the treatment effect of e-book release on total book sales remains to be positive and significant.

Table 4. The results of the DID model				
	(1)	(2)	(3)	(4)
	lnPaper_sales	lnTotal_sales	lnPaper_sales	lnTotal_sales
Release	0.00177 (0.27)	0.0126 (1.94)	0.00279 (0.43)	0.0137* (2.11)
Treated	-0.335*** (-8.19)	-0.339*** (-8.28)	-0.328*** (-8.04)	-0.332*** (-8.13)
Release×Treated	0.00186 (0.25)	0.0616*** (8.27)	-0.00357 (-0.46)	0.0432*** (5.49)
Release×Treated×Novels			0.103*** (7.94)	0.168*** (12.88)
Release×Treated×Best			-0.878*** (-22.29)	-0.857*** (-21.65)
Bestsellers	1.684*** (13.94)	1.682*** (13.91)	1.876*** (15.52)	1.868*** (15.44)
lnAge	-0.334*** (-119.93)	-0.336*** (-120.12)	-0.332*** (-119.26)	-0.334*** (-119.50)
lnPrice	-0.149*** (-15.92)	-0.149*** (-15.79)	-0.151*** (-16.10)	-0.150*** (-15.96)
Book category dummy	Included	Included	Included	Included
Year/month dummy	Included	Included	Included	Included
Constant	4.455*** (18.95)	4.457*** (18.94)	4.459*** (19.01)	4.465*** (19.01)
N	183,420	183,406	183,420	183,406
R-sq	0.202	0.196	0.200	0.195
Note: *t* statistics in parentheses; * *p*<0.05, ** *p*<0.01, *** *p*<0.001; Overall R-sq for the random effect models				

In order to examine to the moderating effect of book characteristics, we interacted *Release×Treated* with two binary variables: *Novels* and *Best*. Because there is no paper book category directly corresponding to the E-PREFER category, we use *Novels* category, a broader category than E-PREFER, which includes romance, martial arts, and fantasy novels. *Best* represents bestsellers as explained earlier. Similar to the results in Table 2, we find that e-book release significantly boosts the paper book sales in Novels category, and that the sales of bestsellers are cannibalized by the release of their e-book versions (see columns (3) and (4) in Table 4).

Discussion and Future Works

This research investigates the impacts of e-book release on demand for the paper book using unique large-scale book-level sales data. Our findings based on the treatment group only (without matching) suggest that releasing e-book version has a positive impact on the paper book sales as well as the total book sales from paper and e-book versions. However, after controlling for selection bias through propensity score matching, our DID analysis results show that the positive impact of e-book release on paper book sales is no longer significant. This implies that publishers are indeed more likely to release e-book version for books that can benefit more from the release. In addition, we find that the impact of e-book release on paper book sales is moderated by consumers' digital content preference. In particular, our results indicate that e-book release boosts the paper book sales more in the book categories with light contents that consumers prefer to consume through a digital channel (e.g., romance, fantasy novels). Finally, our finding that the e-book sales cannibalize the paper book sales of bestsellers more than non-bestsellers justifies publishers' reluctance to release an e-book version when a book is selling well.

Our study has limitations. First, we obtained data from the Korean e-book market, which is at a relatively early stage. Therefore, one should take caution in generalizing the results. Future studies should examine more mature markets (e.g., US) to see if the impact of e-book release varies depending on the maturity of the market. Second, because our data are from a single book retailer, we cannot examine the role of competition between book retailers. Although we expect our results would not be affected much by the consideration of competition because the focal book retailer is the largest in Korea, collecting data from multiple book retailers and analyzing the role of competition in the context of multi-channel publishing will be an interesting avenue for future research. Finally, due to the lack of information about consumers, we cannot

This research is expected to make important contributions. First, our study is among the first to examine the impact of e-book release using large-scale sales data, and provides meaningful implications about the e-book market that remains understudied. In particular, our findings shed light on the channel competition between e-books and paper books. Second, the different results we report based on the treatment group only (a significant positive impact of e-book release on paper book sales) and based on matching and DID analysis (insignificant impact of e-book release) highlight the importance of addressing selection bias in a setting where treatment (e-book release in our case) is endogenous. Without properly controlling for the selection bias, one can reach erroneous conclusions. On a practical front, our findings can help book publishers and retailers, especially those operating in less mature e-book markets, assess the impact of releasing an electronic version on their revenue. Our results can also inform them about the factors that moderate the impact of e-book release. Taken together, our analysis results can help practitioners make better decisions regarding e-book release

References

Bounie, D., Eang, B., Sirbu, M., and Waelbroeck, P. 2013. "Superstars and Outsiders in Online Markets: An Empirical Analysis of Electronic Books," *Electronic Commerce Research and Applications* (12:1), pp. 52–59.

Brynjolfsson, E., Hu, Y. (Jeffrey), and Rahman, M. S. 2009. "Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition," *Management Science* (55:11), pp. 1755–1765.

Danaher, B., Dhanasobhon, S., Smith, M. D., and Telang, R. 2010. "Converting Pirates Without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy," *Marketing Science* (29:6), pp. 1138–1151.

Danaher, B., Smith, M. D., and Telang, R. 2013. "Piracy and Copyright Enforcement Mechanisms," *NBER Working Paper 19150* .

Deleersnyder, B., Geyskens, I., Gielens, K., and Dekimpe, M. G. 2002. "How Cannibalistic Is the Internet Channel? A Study of the Newspaper Industry in the United Kingdom and The Netherlands," *International Journal of Research in Marketing* (19:4), pp. 337–348.

Hao, L., and Fan, M. 2014. "An Analysis of Pricing Models in the Electronic Book Market," *SSRN Electronic Journal* .

Heckman, J. J., Ichimura, H., and Todd, P. 1998. "Matching As An Econometric Evaluation Estimator," *Review of Economic Studies* (65:2), pp. 261–294.

Heckman, J. J., Ichimura, H., and Todd, P. E. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme," *The Review of Economic Studies* (64:4), pp. 605–654.

Hu, Y. J., and Smith, M. D. 2013. "The Impact of Ebook Distribution on Print Sales: Analysis of a Natural Experiment," *SSRN Electronic Journal* .

Moorthy, K. S., and Png, I. P. L. 1992. "Market Segmentation, Cannibalization, and the Timing of Product Introductions," *Management Science* (38:3), pp. 345–359.

Oberholzer-Gee, F., and Strumpf, K. 2007. "The Effect of File Sharing on Record Sales: An Empirical Analysis," *Journal of Political Economy* (115:1), pp. 1–42.

Smith, M. D., and Telang, R. 2009. "Competing with Free: The Impact of Movie Broadcasts on DVD Sales and Internet Piracy 1," *MIS Quarterly* (33:2), pp. 321–338.

Vernik, D. A., Purohit, D., and Desai, P. S. 2011. "Music Downloads and the Flip Side of Digital Rights Management," *Marketing Science* (30:6), pp. 1011–1027.