

# **The Impact of Ebook Distribution on Print Sales: Analysis of a Natural Experiment**

Yu (Jeffrey) Hu<sup>‡</sup> and Michael D. Smith<sup>†</sup>

yuhu@gatech.edu, mds@cmu.edu

This Version: April 2013

Acknowledgements: The authors thank an anonymous publisher for generously providing data to support this research. Smith acknowledges generous support from Carnegie Mellon's iLab.

---

<sup>‡</sup> Scheller College of Business, Georgia Tech University, Atlanta, GA, 30308

<sup>†</sup> School of Information Systems and Management, Heinz College, Carnegie Mellon University, Pittsburgh, PA, 15213.

# **The Impact of Ebook Distribution on Print Sales: Analysis of a Natural Experiment**

## **ABSTRACT**

Digital distribution channels introduce several new strategic questions for the creative industries. Two notable questions are (1) how will participation in digital channels impact physical sales, and (2) where should digital releases occur relative to existing “physical” release dates. These questions are particularly salient for book publishing where firms have expressed concern that making ebooks available alongside hardcover titles may cannibalize hardcover sales.

We analyze the impact of ebook availability on physical sales using a natural experiment that occurred between April and June 2010. Across all titles we find that delaying ebook availability results in a small (and statistically insignificant) increase in print sales and a large decrease in total ebook sales. However, we also find this effect is moderated by consumers’ brand awareness and digital channel preference: delaying Kindle availability for books with strong brand awareness and weak digital channel preference results in a statistically significant increase in print sales for those titles.

These results contribute to a growing academic literature analyzing the impact of digitization on marketing strategies in the media industries. Our results also highlight the usefulness of natural experiments derived from rapidly changing media distribution environments for identifying the impact of digital distribution channels on sales.

**Keywords:** *Digital distribution, channel, publishing industry, natural experiment.*

## **1. Introduction**

Digital distribution channels introduce a variety of new challenges and opportunities for the creative industries. Ebooks represent one prominent (and understudied) setting where these challenges are on display. The last few years have witnessed significant growth in ebook platform adoption, and an associated growth in ebook sales. The Association of American Publishers reports that, ebook sales revenue was \$345 million from January to October 2010, a 171 percent growth from the same time period in 2009, and that ebook sales made up 9 percent of total trade book sales during this period in 2010 (Association of American Publishers 2010). Similarly, Amazon.com, the largest online seller of print and digital books, reported that Kindle sales surpassed Amazon's total hardcover sales in July 2010, and surpassed total print sales as of April 1, 2011 (Amazon.com 2011).

As the popularity of ebooks and other digital products grows, it is increasingly important for academics and practitioners to understand the dynamics of consumer behavior when facing dual physical and digital distribution channels. Answers to these questions will shape the future of the book publishing industry as it adjusts to new digital channels, and will help publishers and retailers make key marketing decisions in these changing environments.

Currently, the book industry is engaged in an active debate about how the use of ebook channels will impact print sales. On one side of the debate are ebook platform providers, such as Amazon, who claim that ebooks do not cannibalize print sales, but rather represent mostly incremental sales. For example, Jeff Bezos, CEO of Amazon, observed in an earnings call with analysts (SeekingAlpha 2009),

*“When people buy a Kindle, they actually continue to buy the same number of physical books going forward as they did before they owned a Kindle. And then incrementally, they buy about 1.6 to 1.7 electronic books, Kindle books, for every physical book that they buy.”*

On the other side of the debate many book publishers worry that because print books and ebooks offer essentially the same content in different formats, it is natural to think that consumption of ebooks would cannibalize sales of print books, particularly those that are priced significantly higher than their ebook counterparts. As Carolyn Reidy, CEO of Simon & Schuster, represented this view in a New York Times article (Rich 2009),

*“What a (book) consumer is buying is the content, not necessarily the format.”*

**Table 1: Delaying Ebooks**

<i><b>Event</b></i>	<i><b>Quote</b></i>
September 2009: HarperCollins delays the ebook release of Sarah Palin’s memoirs by 5 months after the hardcover release date.	<i>"The publishing plan is focused on maximizing velocity of the hardcover before Christmas."</i>  Brian Murray, CEO HarperCollins
November 2009: Viacom/Schribner delays the ebook release of Stephen King’s new novel by 6 weeks after the hardcover release date.	<i>"We think that this publishing sequence gives us the opportunity to maximize hardcover sales"</i>  Adam Rothberg, Spokesperson Schribner
Early 2010: Hachette Book Group delays the ebook release of nearly all of its titles by 3-4 months after the hardcover release date.	<i>"I can't sit back and watch years of building authors sold off at bargain-basement prices."</i>  David Young, CEO Hachette
Early 2010: Simon & Schuster delays the ebook release for 35 major titles by 4 months after the hardcover release date.	<i>"The right place for the e-book is after the hardcover but before the paperback,"</i>  Carolyn Reidy, CEO Simon & Schuster

In 2009 and 2010 these sorts of cannibalization concerns led many publishers to delay ebook releases in the hopes of not cannibalizing hardcover sales. Table 1 reviews several announcements that publishers made regarding their decisions to delay ebook releases.

However, discussions we have had with publishers suggest that these decisions have been made with little empirical analysis. In this paper we attempt to provide empirical evidence regarding how delaying the release of ebooks impacts hardcover sales, ebook sales, and total sales. To do this we use a novel natural experiment where, over the course of a 2-month period, a particular publisher stopped releasing new Kindle ebooks to Amazon, but continued to release new hardcover books to Amazon. This event, which we describe in more detail below, creates an environment where ebook release dates are exogenously delayed by between 1 to 8 weeks for a relatively homogeneous set of titles. This event provides a unique opportunity for studying the cross-channel effect between ebooks and print books.

Our analyses suggest that, in total, delaying ebook release dates relative to their hardcover counterparts results in an insignificant increase in hardcover sales, but results in a large *decrease* in ebook sales, total sales, and likely total profit.<sup>1</sup> However, we also find evidence that the cross-channel effect between ebooks and print books is moderated by the strength of consumers' brand awareness. For titles that have strong brand awareness among consumers, delaying ebook release dates leads to a significant substitution toward print books. In contrast, for titles that have weak brand awareness, we find a significant decline in print books when ebook release dates are delayed.

---

<sup>1</sup> We note that these lower total sales will also lead to lower profits in our setting given that publishers' margins on ebooks are very similar to those on higher priced print books. See for example Rich (2010) who notes that publisher profit (before overhead) on a \$26.00 print book is approximately \$4.05 while publisher profit on the same book sold in ebook format is between \$3.51-\$4.26 for a \$9.99 ebook and \$4.56-\$5.54 for the same ebook sold at \$12.99.

Furthermore, we find evidence that the cross-channel substitution between ebooks and print books is moderated by the strength of consumers' digital channel preference. The unique advantages of ebooks over print books include the ability to carry many books on a single lightweight device and to search inside the book, and such advantages become more salient for books with high weight and many pages and less salient for books with low weight and few pages. Accordingly, we find that delaying ebook release dates leads to a significant substitution toward print books for titles with low weight and few pages; for titles with high weight and many pages, delaying ebook release dates leads to a significant decline in print sales.

We believe these results make three main contributions. First, these results inform an important managerial question currently facing the book industry. Second, these results add to a growing academic literature analyzing the impact of digital distribution channels on existing sales channels. We extend this literature by using a novel natural experiment and sales data collected from the understudied electronic publishing industry that is growing in importance. Third, these results provide evidence that consumers can form surprisingly strong channel or platform preferences between physical and digital products. Our framework on how cross-channel effects are moderated by consumers' brand awareness and digital channel preference can assist academics and managers in designing optimal marketing strategies for firms responding to the availability of new digital distribution channels.

## **2. Literature Review**

Our research is related to a large marketing literature analyzing cannibalization and release timing in media channels (e.g., Lehman and Weinberg 2000, Luan and Sudhir 2006, and Prasad et al. 2004). Within this literature, our research is most closely related to a growing literature

analyzing the degree to which legitimate digital distribution channels impact sales in physical channels. In this literature, Moorthy and Png (1992) analyze how a seller can sequentially introduce a high-end product and a low-end product. In an empirical setting, Deleersnyder *et al.* (2002) find that the introduction of online newspapers resulted in a relatively small cannibalization of physical newspaper sales; Biyalogorsky and Naik (2003) find that the introduction of online storefronts for music did not significantly cannibalize physical record sales; Waldfogel (2007) shows that Youtube viewing has only a small negative impact on television viewing; and Danaher *et al.* (2011) show that the presence of the iTunes distribution channel has no statistical impact on DVD sales (but results in a large impact on digital piracy).

There is also a long and related literature on the impact of digital piracy on sales in existing channels (see Oberholzer-Gee and Strumpf (2010) and Danaher, Smith, and Telang (Forthcoming) for reviews of this literature), in general finding that piracy harms physical sales (see for example Zentner (2005), Hui and Png (2003), Peitz and Waelbroeck (2004), Danaher and Waldfogel (2011), Rob and Waldfogel (2006), and Rob and Waldfogel (2007)). However, a few notable exceptions to this are Oberholzer and Strumpf (2007) who find little displacement of record sales from digital music piracy and Smith and Telang (2009) who find little displacement of DVD sales by digital movie piracy for catalog (i.e., older) content.

Our research differs from these papers in that we use a novel natural experiment to study substitution between digital and physical channels. This quasi-experimental setting allows us establish causality and avoid many confounding factors common in related cross-channel settings. Further, our research extends the current literature from music and movie industries to the understudied (and increasingly important) electronic publishing industry.

Our theory that cross-channel effects can be moderated by consumers' brand awareness and digital channel preference is related to several recent papers showing similar types of effects. Simester et al. (2009) show that the effects of current advertising on future sales are moderated by brand awareness among consumers. Brand awareness is also related to marketing buzz that can affect product sales, as shown by Elberse and Eliashberg (2003) and Luan and Sudir (2006). Furthermore, Vernik, Purohit, and Desai (2011) show that consumers may have strong preference for digital format in the music industry, and this preference can drive product sales. Frambach, Roest, and Krishnan (2007) demonstrate that consumers can develop higher preferences for the online purchasing channel when they have more favorable Internet experience. Our paper applies such existing concepts of consumers' brand awareness and channel preference to the setting of cross-channel effects between digital and physical channels.

Finally, our work relates to a literature studying online sales of books. In this literature, Ghose, Smith, and Telang (2006) show that Internet channels for used books result in a relatively small cannibalization of used book sales, Chevalier and Goolsbee (2003) show that the cross price elasticity between books sold at Amazon and Barnes and Noble is relatively low, and Kannan, Pope, and Jain (2009) analyze optimal pricing strategies for dual channel — print and ebook — publishers.

### **3. Theory**

Altering a focal product's availability through digital distribution channels can affect product sales in digital and physical channels in two ways: 1) it may influence the focal product's overall brand awareness; 2) it may also affect how its brand awareness is successfully converted into actual sales in the available channels. The second effect can be further divided into two sub-



effects: consumers who switch between channels when the product is not available in their chosen channel or consumers who leave the market for a particular product when it is not available in their chosen channel.

We use two simple equations to illustrate the aforementioned effects in our setting where ebook releases are delayed relative to hardcover availability. We consider a focal book (book  $i$ ) that can potentially be sold in two channels: a digital channel and a print channel. We let book  $i$ 's sales in the print channel be  $PS_i = A(X_i, D_i) * PC(X_i, D_i)$ , where  $A(X_i, D_i)$  is book  $i$ 's overall brand awareness among consumers, and  $PC(X_i, D_i)$  is the conversion rate from brand awareness to print sales. Similarly, book  $i$ 's sales in the digital channel are given by  $DS_i = A(X_i, D_i) * DC(X_i, D_i)$ , where  $A(X_i, D_i)$  is again brand awareness, and  $DC(X_i, D_i)$  is the conversion rate from brand awareness to digital sales. Note that we implicitly assume that brand awareness and the conversion rate from brand awareness to print (or digital) sales are affected by a series of marketing actions, denoted by  $X_i$ , and the delay of the book's digital release, denoted by  $D_i$ .

Taking the natural log of book  $i$ 's print and digital sales, and differentiating them with respect to the delay in digital release  $D_i$ , we obtain the following two equations:

$$\frac{\partial \ln (PS_i)}{\partial D_i} = \frac{\partial \ln (A(X_i, D_i))}{\partial D_i} + \frac{\partial \ln (PC(X_i, D_i))}{\partial D_i}, \quad (1)$$

$$\frac{\partial \ln (DS_i)}{\partial D_i} = \frac{\partial \ln (A(X_i, D_i))}{\partial D_i} + \frac{\partial \ln (DC(X_i, D_i))}{\partial D_i}. \quad (2)$$

Equations (1) and (2) illustrate that the effect of book  $i$ 's digital delay on its print (or digital) sales is the sum of two effects: the effect on its overall brand awareness and the effect on the conversion rate from brand awareness to print (or digital) sales.<sup>2</sup>

### The Effect on Same-channel (Digital-channel) Conversion

Delaying the focal title's digital release should have a negative effect on conversion from brand awareness to same-channel (digital-channel) sales. Previous research on product stockouts has shown that consumers who face stockouts of a certain brand may switch to other sizes and varieties of the same brand, or switch to a competing brand (e.g., Emmelhainz et al. 1991). Similarly, ebook consumers may switch from the focal title's ebook to its print book when its ebook is delayed, simply because ebooks and print books offer the same content and could be strong substitutes for each other; ebook consumers may also switch from the focal title's ebook to other competing titles' ebook versions if they exhibit strong preference for the ebook channel. Both types of switching point to declines in the conversion from brand awareness to digital-channel sales.

It is not clear how the effect on same-channel conversion is moderated by the title's brand awareness. Strong brand awareness increases the possibility of switching to the same title's print version; while at the same time, it limits switching to other titles' ebooks by making those ebooks a poor substitute to the focal title's ebook. Likewise, it is not clear how the effect on the same-channel conversion is moderated by the preference for digital channel. Strong channel preference limits the possibility of switching to the same title's print version; while at the same

---

<sup>2</sup> We note that all the effects mentioned above are expressed in relative terms, because for instance,  $\frac{\partial \ln(PS_i)}{\partial D_i} = \frac{\partial PS_i / PS_i}{\partial D_i}$ .

time, it increases switching to other titles' ebooks by making those ebooks a good substitute to the focal title's ebook. Therefore, we conclude that the effect on same-channel conversion, denoted by  $\frac{\partial \ln(DC_i)}{\partial D_i}$ , is likely to be strongly negative for all titles.

#### The Effect on Cross-channel (Print-channel) Conversion

Delaying digital releases can potentially lead to an increase in the conversion rate from awareness to cross-channel (print-channel) sales. More interestingly, the size of this effect is moderated by the strength of the focal title's brand awareness and the consumer's preference for the digital channel. When the focal title's ebook is delayed, ebook consumers may switch from the ebook to the print version, or switch from the focal title's ebook to other competing titles' ebook versions.

Switching to the focal title's print book is more likely to happen if the focal title has strong brand awareness, while switching to other competing titles' digital versions is more likely to happen if brand awareness is weak. Therefore, we believe that the effect on cross-channel conversion, denoted by  $\frac{\partial \ln(PC_i)}{\partial D_i}$ , is likely to be strongly positive for titles with strong brand awareness and weakly positive for books with weak brand awareness.

In addition, there are reasons to believe that ebook and print markets could be distinct due to ebook consumers' preference for digital distribution channels. Marketing theory shows that markets can be segmented based on consumer's channel preferences (Kotler 2002). Thus, it is conceivable that ebook consumers may not even consider purchasing print books. Ebooks may also have unique features that further enhance consumers' preference for ebooks. When the focal title's ebook is delayed, ebook consumers may switch to the digital copy of competing titles, and

such a scenario is more likely for titles where consumers have strong digital channel preference. These preferences might be more salient where specific ebook titles have a stronger advantage over print titles. For example, consumers may be less likely to switch from ebook to print titles for unusually heavy or unusually long print books. Therefore, we expect consumers' digital channel preference to be weaker for titles with relatively lower weight and fewer pages than for other books, and as a result, consumers will be more likely to switch to print books when short or lightweight books are unavailable in digital channels than they would be for heavier or longer titles. In other words, the effect on cross-channel conversion, denoted by  $\frac{\partial \ln(PC_i)}{\partial D_i}$ , is likely to be strongly positive for books with low weight and few pages, and weakly positive for titles with high weight and many pages.

Finally, we note that consumer patience can potentially play a role in driving both same-channel conversion and cross-channel conversion, particularly if consumers know how long the delay of digital releases will be (see for example Luan and Sudir (2006)). However, as we discuss in more detail below in Section 4, in our experiment consumers had no information about when Kindle titles would be returned to the Amazon store, reducing the importance of this factor in our setting.

#### The Effect on Overall Brand Awareness

Many marketing actions can affect books' overall brand awareness among consumers. For instance, publishers often promote their books through traditional media channels, book reviews, and book signings. With the rise of Internet channels, online word-of-mouth has become an increasingly important source of brand awareness for books and an important driver of product sales (e.g., Luan and Sudir 2010). Publishers often actively court, and even pay, top reviewers to

create online word-of-mouth for their new books (Coster 2006). Importantly, online word-of-mouth can occur from both print book and ebook consumers. As such, delaying a title's digital release may suppress the online word-of-mouth that would have been generated by ebook consumers had the book not been delayed, resulting in harm to the book's overall brand awareness.

Furthermore, if a title has generated much publicity via book critics, media mentions, or book-signing events, delaying its digital release may have a smaller (relative) impact on brand awareness than for books with less public "buzz." In other words, the effect on overall brand awareness, denoted by  $\frac{\partial \ln(A_i)}{\partial D_i}$ , is likely to be weakly negative for titles with strong brand awareness and strongly negative for titles with weak brand awareness.

Figure 1 summarizes these observations regarding how reducing a title's availability on digital channels may change overall brand awareness, cross- and same-channel conversion rates, and how such effects vary across books with different levels of brand awareness and digital channel preference. We also aggregate the respective individual effects to obtain the expected net effect on print sales, on digital sales, and on overall sales. In sum, we have the following observations.

Observation 1: For books with strong brand awareness, delaying digital releases will lead to an increase in print sales, a decrease in digital sales, and a weak decline in overall sales.

Observation 2: For books with weak brand awareness, delaying digital releases will lead to a decrease in print sales, a decrease in digital sales, and a strong decline in overall sales.

Observation 3: For books with weak digital channel preference, delaying digital releases will lead to an increase in print sales, a decrease in digital sales, and a decline in overall sales.

Observation 4: For books with weak awareness, delaying digital releases will lead to a decrease in print sales, a decrease in digital sales, and a decline in overall sales.

**Figure 1: How Delaying Digital Releases May Affect Print and Digital Sales**

<i>Book</i>	<i>Effect on Overall Awareness</i>	<i>Effect on Cross-channel Conversion</i>	<i>Net Effect on Print Sales</i>	<i>Effect on Overall Awareness</i>	<i>Effect on Same-channel Conversion</i>	<i>Net Effect on Digital Sales</i>	<i>Net Effect on Overall Sales</i>
Book with strong brand awareness	Weakly -	Strongly +	+	Weakly -	Strongly -	-	Weakly -
Book with weak brand awareness	Strongly -	Weakly +	-	Strongly -	Strongly -	-	Strongly -
Book with weak digital preference	-	Strongly +	+	-	Strongly -	-	-
Book with strong digital preference	-	Weakly +	-	-	Strongly -	-	-

#### **4. Data and Description of Natural Experiment**

An ideal starting point for understanding how delayed release dates for ebooks affect sales would be a pure experiment where a publisher randomly assigns books to treatment groups with different amounts of delay. Lacking this data, our research employs a natural experiment that approximates this ideal setup.

Specifically, we use data from a publisher that stopped distributing new Kindle titles to Amazon from April 1, 2010 and June 1, 2010. Prior to April 1, the policy of this publisher was to release Kindle titles on the same day that a book is initially released in print (typically in a hardcover version). Starting on April 1, as a result of a dispute with Amazon, this publisher stopped releasing new Kindle titles to Amazon, while still releasing new print titles. On June 1, 2010 the

publisher returned to the Kindle store, releasing all Kindle titles that had been released in print from April 1 through May 31 and returning to their previous policy of releasing Kindle ebooks on the same day as the print (hardcover) version was released.<sup>3</sup>

**Table 1: Kindle Release “Experiment”**

<i>Week(s)</i>	<i>Print Release</i>	<i>Kindle Release</i>	<i>Kindle Delay</i>
Before April 1, 2010	Print and Kindle titles released same day		0 weeks
Week of April 4	April 4	June 1	8 weeks
Week of April 11	April 11	June 1	7 weeks
Week of April 18	April 18	June 1	6 weeks
Week of April 25	April 25	June 1	5 weeks
Week of May 2	May 2	June 1	4 weeks
Week of May 9	May 9	June 1	3 weeks
Week of May 16	May 16	June 1	2 weeks
Week of May 23	May 23	June 1	1 week
After June 1, 2010	Print and Kindle titles released same day		0 weeks

This results in the release schedule of Kindle and print titles shown in Table 1. Notably for our purposes, if the titles released in April and May are statistically similar to those released in March and June, this creates a “natural experiment” where a sample of Kindle titles are delayed by between 1 to 8 weeks relative to their print counterparts.

To analyze the results of this “experiment,” we obtained data from the publisher in question, covering print and Kindle unit sales, prices, and release dates for all books released from March 1, 2010 through June 30, 2010. One confounding factor in our experiment is that Apple’s iBookstore opened on April 3 and included content from this publisher. To take this event into account, we also obtained this publisher’s sales on the iBookstore for the titles in our data.

We then divide this sample into two groups as follows:

---

<sup>3</sup> Just as importantly for our discussion below, based on discussions with the publisher in question and based on searches of press articles in both Lexis-Nexis and Proquest, there was no prior public indication that Kindle titles were going to be removed from Amazon for this publisher, and the first time there was any discussion of the return date for this publisher was on May 26, just 5 days prior to their restoration onto Amazon.

1. “Control” Group: books that are released from March 1 to March 31 and from June 1 to June 30. These books are released simultaneously in both print and Kindle formats.
2. “Experiment” Group: books that are released from April 1 to May 31. These books are released first in print and then one to eight weeks later (on June 1) in Kindle format.

Overall, we have 83 “control” titles and 99 “experiment” titles.

#### 4.1. Comparing The Control and Experiment Groups

For this to be a valid setting for a natural experiment we first must establish that books in the control group have a similar profile to books in the experiment group on observable dimensions — which in our case include the number of pages, weight, height, length, width and list price of the books. Summary statistics on these dimensions are provided in Table 2 separately for books in the control group and books in the experiment group. The statistics in Table 2 show that the control and experiment titles are very similar along all these observable dimensions.

**Table 2: Summary Statistics**

		<i>Mean</i>	<i>S.D.</i>	<i>Min</i>	<i>Max</i>	<i>Median</i>	<i>Obs.</i>
“Control” Group	Number of Pages	349.6	128.1	160.0	1184.0	304.0	83
	Weight (pound)	1.2	0.4	0.5	3.6	1.1	83
	Height (inch)	1.3	0.3	0.8	2.5	1.3	83
	Length (inch)	8.8	0.6	6.9	10.0	9.0	83
	Width (inch)	6.0	0.4	4.8	7.8	6.1	83
	List Price (\$)	26.2	3.1	16.0	45.0	26.0	83
“Experiment” Group	Number of Pages	339.6	88.9	192.0	704.0	320.0	99
	Weight (pound)	1.1	0.3	0.3	2.3	1.2	99
	Height (inch)	1.3	0.2	0.8	2.3	1.3	99
	Length (inch)	8.9	0.4	7.1	10.6	9.1	99
	Width (inch)	6.1	0.4	5.1	8.0	6.2	99
	List Price (\$)	25.9	2.0	22.0	38.0	26.0	99

We can test whether any of the observable variables can be used to predict a particular book belonging to the experiment group. To accomplish this, we estimate the following Probit model:



$$P(Experiment_i = 1 | X_i) = \Phi(X_i\beta) , \quad (3)$$

where  $Experiment_i$  is an indicator of whether book  $i$  belongs to the experiment group, and  $X_i$  is a vector of independent variables including book  $i$ 's number of pages, weight, height, length, width, and list price. The results from estimating this model are reported in Table 3.

The results in Table 3 show that none of the coefficients is statistically significant at the 5% confidence level. This is again consistent with the hypothesis that books in the control group are materially similar to — and thus a reasonable control for — books in the experiment group.

**Table 3: Results from A Probit Model**

	<i>Experiment Group Dummy</i>
Constant	-3.989 (2.026)
Number of Pages	-0.0008 (0.003)
Weight	-0.007 (0.005)
Height	0.010 (0.009)
Length	0.007 (0.004)
Width	0.0002 (0.005)
List Price	-0.083 (0.062)
Log Likelihood	-119.6
Number of Obs.	182

Standard errors are in parentheses.

\*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ .

## 4.2. Initial Analyses of Book Sales

When studying the sales and sales patterns of these books, we focus on the sales in the first twenty weeks since each book's initial release in their respective channels. We do this because

this is the period where the majority of sales occur, allowing us to keep the sales numbers comparable across different books.

Summary statistics can provide an initial test for whether there are differences in sales patterns between the control and experiment titles. These summary statistics are provided in Table 4 for sales of books in the control and experiment groups in the first twenty weeks since each book's release in the respective channel (print and Kindle). We note that digital sales of the books in the experiment group would be unambiguously lower if we studied digital sales in the first twenty weeks since each book's print release: this is because digital sales in the first few weeks would have been zero for the digital books in the experiment group.

**Table 4: Summary Statistics (Weeks 1-20 since Each Version's Release)**

		<i>Mean</i>	<i>S.D.</i>	<i>Min</i>	<i>Max</i>	<i>Median</i>	<i>Obs.</i>
"Control" Titles	Total Sales	248.8	825.2	0	19,474	60	1,660
	Print Sales	123.7	483.6	0	10,408	32	1,660
	Digital Sales	125.2	417.6	0	9,066	20	1,660
"Experiment" Titles	Total Sales <sup>†</sup>	168.7	381.2	0	5,573	53	1,980
	Print Sales	132.5	364.3	0	5,573	34	1,980
	Digital Sales	40.1	86.9	0	1,247	13	1,980
	Digital Sales after Print Release <sup>†</sup>	36.2	86.9	0	1,247	8	1,980

<sup>†</sup> Sales are shown for weeks 1-20 after print release

Looking first at the "Control" titles, these summary statistics suggest that digital sales are a significant sales channel. Digital sales make up nearly half of total sales for this publisher, which is generally consistent with available data regarding sales patterns at Amazon (e.g., Amazon.com 2011). More importantly for our study, these summary statistics suggest that in the first 20 weeks after an ebook was made available for sale, ebook sales in the experiment group are significantly lower than in they are in the control group. These summary statistics also show that print sales,

while slightly higher in the experiment group than in the control group, are nowhere near large enough to compensate for the lost ebook sales.

To further test for this effect, we first run a linear regression of book sales onto a dummy variable that equals one for all the books in the experiment group. We then run a quantile regression to replace the linear regression so that the results are less likely to be driven by any potential outliers in our sample. The model we estimate for each version (print or Kindle) of book  $i$  in each week  $t$  is simply:

$$Sales_{it} = \beta_0 + \beta_1 Experiment_i + \varepsilon_{it} \quad (4)$$

where  $Sales_{it}$  represents the sales generated by a particular version of book  $i$  in week  $t$  and  $Experiment_i$  is a dummy variable that equals one if book  $i$  belongs to the experiment group. We study print sales first followed by digital sales, and report the results from the linear regression models in Table 5.

When examining print sales (Column 1 of Table 5), we find that there are, on average, 8.9 more print sales for the experiment group than for the control group, but that this difference is statistically insignificant. On the other hand, examining digital sales (Column 2 of Table 5), we find that there are 85.0 fewer digital sales for the experiment group than for the control group. Thus, initial analyses show that delaying Kindle releases has not led to a statistically significant increase in print sales while leading to a significant decrease in digital sales. Moreover, Column 3 of Table 5 shows that there are 80.1 fewer total sales for the experiment group than for the control group, indicating that (as expected given the statistically insignificant increase in print sales) the net effect of delaying ebook releases is an overall loss in total sales (and as we discuss in more detail below likely total revenue). Note that when a book's Kindle release is delayed, it

will have zero digital sales (by definition) until its Kindle version is released. To ensure our comparison is valid, we compare digital sales of “experiment” titles after each book’s digital release with digital sales of “control” titles, as shown in Column 2 of Table 5. If one were to study digital sales after print release, the drop in digital sales as a result of Kindle delays would be larger—89.0 fewer digital sales. We report these results in Column 4 of Table 5.

**Table 5: Initial Analyses of Sales Data (Weeks 1-20 since Each Version’s Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	123.7** (10.4)	125.2** (7.1)	248.8** (15.3)	125.2** (7.1)
<i>Experiment</i>	8.9 (14.1)	-85.0** (9.6)	-80.1** (20.8)	-89.0** (9.6)
R <sup>2</sup>	0.0001	0.021	0.004	0.023
Number of Obs.	3,640	3,640	3,640	3,640

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release

These analyses are sufficient for our data if the control group serves as a valid control for our experiment group. This statement is true as long as other factors that may influence book sales — such as book-level characteristics, over-time seasonality or sales trends — are orthogonal to the experiment treatment. Our discussions with the publisher who provided these data suggest that this is likely true: the publisher was unaware of and had not planned any systematic differences between the control and experiment groups. In fact, the publisher provided us with its publishing schedule for the first six months of 2010 in October 2009, and at that time, no one from the publisher expected a dispute to occur with Amazon in 2010. However, in spite of these reassurances, in the next section we use a more sophisticated model to control for the possibility of other factors unrelated to our experiment that may influence book sales.

## 5. Extended Model and Results

The first factor we control for in our extended model is the number of weeks since release. We do this because book sales typically follow a declining curve over time. The price of a book may also influence sales. Thus, we add control variables for the (logged) weekly average prices of print and digital versions. We also control for a set of book-level characteristics, including the number of pages, weight, height, length, width, and the number of pre-print-release Amazon reviews. Finally, we add a set of weekly fixed effects for each calendar week in our data. These weekly dummy variables will capture any variation in book sales over time that are driven by seasonality or the effect of the opening of iBookstore on April 3. Since weekly print and digital sales are count data, and we find evidence of over-dispersion, we use a negative binomial regression model (Hausman et al. 1984). Accordingly, we estimate the following model for each version (print or digital) of book  $i$  in each week  $t$ :

$$f(Y_{it}|X_{it}) = \frac{e^{-\mu_{it}} \mu_{it}^{Y_{it}}}{Y_{it}!}, Y_{it} = 0, 1, 2, 3, \dots \quad (5)$$

where  $Y_{it}$  represents the sales generated by a particular version of book  $i$  in week  $t$ ;  $E(Y_{it}|X_{it}) = \mu_{it} = \exp(X_{it}\beta + \varepsilon_{it})$  is the conditional mean;  $\varepsilon_{it}$  is the unobserved heterogeneity and is assumed to follow a log-gamma distribution, with  $\varepsilon_{it} \sim \Gamma(\theta, \theta)$  (Greene 2002). The explanatory variables  $X_{it}$  include *Experiment<sub>i</sub>*, a dummy variable that equals one if book  $i$  belongs to the experiment group; *NumWeeks<sub>it</sub>*, the number of weeks from the release of that version of book  $i$  until week  $t$ ;  $\ln(\text{PrintPrice}_{it})$ , the natural log of price for the print version of book  $i$  in week  $t$ ;  $\ln(\text{DigitalPrice}_{it})$ , the natural log of price for the digital version of book  $i$  in week  $t$ ; *BookVar<sub>i</sub>* is a series of book level characteristics such as book  $i$ 's number of pages, weight, height, length,

width, and number of pre-print-release Amazon reviews; and  $Week_t$ , a series of dummy variables for each calendar week.

An advantage of this model is that it allows us to directly control for any factors that may have affected book sales during this timeframe that may differ between the control and experiment groups. We note that if sales of the control books are statistically representative of what sales of the experiment books would have been if the Kindle version had been released along with the print version (as we believe is the case), then this test is unnecessary. However, if the market environment that the experiment group books faced is systematically different from the market environment the control group faced, then this more sophisticated model can be used to control for some of these outside factors. Our focus is on the coefficient of the *Experiment* variable.<sup>4</sup>

### 5.1. The Effects on Print Sales and Digital Sales

Column 1 of Table 6 reports our results from estimating the model in Equation (5) with print sales as the dependent variable. These results show that the coefficient on the *Experiment* variable is statistically insignificant, indicating that delaying Kindle releases does not lead to a significant change in print sales consistent with our summary statistics above. Our control variables have the expected signs and approximate magnitudes. According to the coefficient -0.137 on *NumWeeks* in Column 1 of Table 6, print sales decline at roughly 12.8% (i.e.,  $e^{-0.137}-1=-0.128$ ) per week.

Column 2 of Table 6 reports the results when this more sophisticated model is applied to analyze digital sales. These results show a strong decrease in sales for delayed Kindle titles: the (statistically significant) -0.591 coefficient on variable *Experiment* means that delaying Kindle

---

<sup>4</sup> Our results remain qualitatively unchanged whether price variables are included or not.

releases leads to a roughly 44.6% (i.e.,  $e^{-0.591}-1=-0.446$ ) decrease in digital sales. Column 3 of Table 6 reports the results when the dependent variable is total sales. The (statistically significant) -0.214 coefficient on variable *Experiment* means that total sales decrease by 19.3% (i.e.,  $e^{-0.214}-1=-0.193$ ) as a result of delaying Kindle releases. Finally, Column 4 of Table 6 shows the results when digital sales after each title's print release are used as the dependent variable.

The results in Table 6 are consistent with the results of our initial analyses in Table 5: delaying Kindle releases leads to a large decline in ebook sales, while print sales are not significantly affected by the delay of Kindle releases.

**Table 6: Analyses of Sales Data (Weeks 1-20 after Each Version's Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	6.750** (0.583)	6.496** (0.939)	7.689** (0.603)	7.577** (1.035)
<i>Experiment</i>	-0.049 (0.041)	-0.591** (0.058)	-0.214** (0.041)	-1.014** (0.076)
<i>NumWeeks</i>	-0.137** (0.005)	-0.124** (0.008)	-0.124** (0.005)	-0.040** (0.009)
Other Control Variables	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies
Log Likelihood	-19043.8	-16556.7	-21007.3	-15759.4
Number of Obs.	3,640	3,640	3,640	3,640

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release

## 5.2. The Effects Vary across Strong-Awareness Books and Weak-Awareness Books

As discussed in Section 3, the effect of delaying a book's digital release on the book's print and digital sales is likely moderated by whether the book has strong brand awareness among consumers. In this section we explore such moderating effects. To do this, we use a book's pre-print-release Amazon reviews as a proxy for its brand awareness. More precisely, we rank all the books by the number of pre-print-release reviews observed on Amazon. We then split our sample

at the 67<sup>th</sup> percentile, which is 3 pre-print-release Amazon reviews, defining the top one-third books as “strong-awareness books” (with *StrongAwareness*=1) and the bottom two-thirds books as “weak-awareness books” (with *WeakAwareness*=1). Note that the number of pre-print-release Amazon reviews is not driven by sales after the book’s release; hence, it is an exogenous variable that is not directly affected by whether the book’s digital version is delayed. We then interact *StrongAwareness* and *WeakAwareness* with *Experiment* and replace the *Experiment* variable in equation (5) by interaction terms *Experiment\*StrongAwareness* and *Experiment\*WeakAwareness*. The results of estimating this new model are reported in Table 7.<sup>5</sup>

These results show that delaying digital releases leads to a statistically significant decrease in digital sales of both strong- and weak-awareness titles: the coefficient on *Experiment\*StrongAwareness* is -0.653 (translating to a 48.0% drop in sales since  $e^{-0.653}-1=-0.480$ ) and the coefficient on *Experiment\*WeakAwareness* is -0.549 (translating to a 42.2% drop since  $e^{-0.549}-1=-0.422$ ), with both coefficients being highly significant.

When studying print sales, we find that the coefficient on *Experiment\*StrongAwareness* is 0.238 (translating to a 26.9% increase since  $e^{0.238}-1=0.269$ ) and the coefficient on *Experiment\*WeakAwareness* is -0.266 (translating to a 23.4% drop since  $e^{-0.266}-1=-0.234$ ), where both coefficients are highly significant. In other words, delaying digital releases of strong-awareness books leads to a significant *increase* in print sales of such books while delaying digital releases of weak-awareness books leads to a significant decrease in print sales.

When studying total sales, we find that delaying digital releases of strong-awareness books leads to an insignificant decline in total sales of such books, and that delaying digital releases of weak-

---

<sup>5</sup> We note that our results are robust to a variety of other threshold levels for splitting the sample.



awareness books leads to a significant decline in total sales of such books (i.e., a 25.5% decline since  $e^{-0.295} - 1 = -0.255$ ).

All of these results shown in Table 7 are consistent with Observations 1 and 2 at the end of our theory discussions in Section 3. Our interpretation of the results in Table 7 is that we have found evidence that consumers' brand awareness can moderate the effect that delaying digital sales has on print and digital sales.

**Table 7: Analyses of Sales Data for Strong-Awareness and Weak-Awareness Titles (Weeks 1-20 after Each Version's Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	7.591** (0.590)	6.353** (0.943)	8.012** (0.611)	7.558** (1.032)
<i>Experiment*StrongAwareness</i>	0.238** (0.053)	-0.653** (0.073)	-0.099 (0.054)	-1.117** (0.093)
<i>Experiment*WeakAwareness</i>	-0.266** (0.047)	-0.549** (0.066)	-0.295** (0.048)	-0.946** (0.084)
<i>NumWeeks</i>	-0.137** (0.005)	-0.124** (0.008)	-0.124** (0.005)	-0.039* (0.010)
Other Control Variables	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies
Log Likelihood	-19006.7	-16555.8	-21001.5	-15757.6
Number of Obs.	3,640	3,640	3,640	3,640

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release

### 5.3. The Effects Vary across Books with Low Weight and Books with High Weight

As noted in Section 3, the effect of delaying a book's digital release on the book's print and digital sales are likely moderated by whether the book's consumers have strong or weak preference for digital channels. In this section we explore such moderating effects. As discussed in Section 3, the unique advantages of ebooks over print books, such as the capability to carry many ebooks on a single device and search inside the book, may be less salient for titles that

have few pages (content) and are that are lighter than other titles. Therefore, we explore how a book's weight can moderate the effects of delaying its digital release. More precisely, we rank all the books by their weight. We then split our sample at the 33<sup>th</sup> percentile, which is 1.00 pound, defining the bottom one-third books as "low-weight books" (with *LowWeight*=1) and the top two-thirds books as "high-weight books" (with *HighWeight*=1). Next, we interact *LowWeight* and *HighWeight* with *Experiment* and replace the *Experiment* variable in Equation (5) by interaction terms *Experiment\*LowWeight* and *Experiment\* HighWeight*. The results of estimating this new model are reported in Table 8.<sup>6</sup>

These results show that delaying digital releases leads to a statistically significant decrease in digital sales of both low-weight and high-weight titles: we find that the coefficient on *Experiment\*LowWeight* is -0.403 (translating to a 33.2% drop since  $e^{-0.403}-1=-0.332$ ) and the coefficient on *Experiment\*HighWeight* is -0.699 (translating to a 50.3% drop since  $e^{-0.699}-1=-0.503$ ), with both coefficients being highly significant.

When studying print sales, we find that the coefficient on *Experiment\*LowWeight* is 0.334 (translating to a 39.7% increase since  $e^{0.334}-1=0.269$ ) and the coefficient on *Experiment\*HighWeight* is -0.278 (translating to a 24.3% drop since  $e^{-0.278}-1=-0.243$ ), with both coefficients being highly significant. In other words, delaying digital releases of low-weight books leads to a significant increase in print sales of such books; in contrast, delaying digital releases of high-weight books leads to a significant decrease in print sales of such books.

---

<sup>6</sup> We note that our results are robust to a variety of other threshold levels for splitting the sample.

When studying total sales, delaying digital releases of low-weight books leads to an insignificant decline in total sales of such books, and delaying digital releases of high-weight books leads to a significant decline in total sales of such books (i.e., a 30.1% decline since  $e^{-0.358}-1=-0.301$ ).

These results are consistent with Observations 3 and 4 in Section 3. Our interpretation of the results in Table 8 is that we have found evidence that a book's weight can moderate the effect of delaying digital sales on print and digital sales. As a robustness check, we have also tested whether our results would hold if we use the number of pages as a moderating factor. These robustness check results are reported in Table A1 of the Appendix and are qualitatively the same as the results in Table 8 for weight.

**Table 8: Analyses of Sales Data for Titles with Low and High Weight (Weeks 1-20 after Each Version's Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	6.630** (0.579)	6.697** (0.943)	7.680** (0.602)	7.818** (1.042)
<i>Experiment*LowWeight</i>	0.334** (0.058)	-0.403** (0.079)	0.041 (0.059)	-0.825** (0.100)
<i>Experiment*HighWeight</i>	-0.278** (0.046)	-0.699** (0.065)	-0.358** (0.047)	-1.119** (0.083)
<i>NumWeeks</i>	-0.137** (0.005)	-0.125** (0.008)	-0.124** (0.005)	-0.040** (0.010)
Other Control Variables	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies
Log Likelihood	-18,993.8	-16,550.3	-20,986.4	-15,754.8
Number of Obs.	3,640	3,640	3,640	3,640

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release

#### 5.4. Sensitivity: Adding iBooks Sales Does Not Change Our Results

One could argue that at least part of the drop in Kindle sales for the experiment group could be attributed to the opening of iBookstore on Apr 3 if the decrease in sales came from consumers

who substituted from the (unavailable) Kindle channel to the iBookstore. We first note that iBookstore purchases can only be viewed on Apple iOS devices (iPhone, iPod Touch, iPad), reducing the potential market segment that could make the tradeoff between the Kindle and iBookstore. We also note that we have already controlled for this effect in part by adding a set of weekly dummy variables.

However, a more comprehensive way to control for this effect is to add all iBook sales to Kindle sales and study whether our original results still hold. To do this, we first add weekly iBook sales to Kindle sales to create a new digital sales variable. We then repeat our analyses shown in Table 7 and Table 8, and report these new results as Table A2 and Table A3 in the Appendix. Of course this represents a very conservative test given that it includes both iBookstore purchases from customers who may have switched from the Kindle to the iBook channel, and purchases from new iBookstore customers who were not previously Kindle customers (and therefore whose sales are unrelated to the experiment).

Notably, the results in Table A2 are very similar to those in Table 7, while the results in Table A3 are very similar to those in Table 8. Note that the number of observations varies slightly in some columns due to the unavailability of iBook sales data in later weeks. Since the coefficients change only minimally when switching from Table 7 and Table 8 to Table A2 and Table A3, we conclude that our results are robust to including iBook sales.

### **5.5. Sensitivity: Extending to Week 40 Does Not Change Our Results**

Thus far we have restricted our analyses to weeks 1 through 20 after each version's release for each title. As an additional robustness check on our results, in this section extend our analyses in Table 7 and Table 8 to use data from weeks 1 through 40 for each version's release. We report

these new results as Table A4 and Table A5 in the Appendix. Note that the number of observations changes slightly in some columns due to the unavailability of sales data in later weeks. The results in Table A4 are very similar to those in Table 7, while the results in Table A5 are very similar to those in Table 8. Thus, we conclude that our results are robust to extending our analyses to weeks 1 through 40 after each version’s release for each title.

## **6. Discussion**

Our research analyzes how the availability of a product in a digital channel impacts sales in physical and digital channels — in our case in the digital and physical channels for book sales. This question is important to both managerial and academic audiences. From a managerial perspective, content owners across a variety of industries are making decisions about whether to include digital products in their existing set of distribution channels and where to place these products into their existing product release cycles. These decisions are particularly salient for book publishers, many of whom have experimented with releasing ebook versions in between the (high margin) hardcover release date and the (lower margin) paperback release date. From an academic perspective, our work adds to a growing literature analyzing the impact of new digital distribution channels on physical sales.

We analyze this question using a novel “natural experiment” where a publisher stopped releasing new ebook content to Amazon for a period of about two months. Because this publisher still released print copies of their titles to Amazon during this period, the net effect of this event is that books released during this timeframe were delayed on the electronic channel relative to the print channel by between one to eight weeks.

Our results suggest that delaying the publication of the ebook relative to its print version causes a large and persistent decrease in overall digital sales. Moreover, while there may be some substitution between the (unavailable) Kindle channel and the print channel for books with strong brand awareness and weak digital channel preference, in total this substitution does not compensate for the lost digital sales. Thus, the net effect of delayed Kindle releases is an overall loss in sales and (as noted above), based on the best available data, a net loss in revenue and profit to the publisher.

These results also point to the possibility that, in general, consumers often are relatively more tied to their mode of consumption (physical or digital) than they are to a particular product. Said another way, when facing new digital channels, publishers and other media firms have frequently conceptualized the consumer's decision process as being driven by product choice first and then channel choice. This conceptualization is seen in the frequent strategy to delay digital availability as a way of retaining physical sales. Our results suggest that this framing may be wrong, at least for products for which consumers do not have strong brand awareness. Rather, our results suggest that for many products consumers may choose their channel (digital versus physical) first, and then restrict their choice set to the products available in this channel.

Finally, we also note that our results do not suggest that digital channels will not cannibalize aggregate physical sales in the long term; they certainly will. Rather, we believe our results suggest that as a managerial question, given that digital channels exist and that some consumers have preferences for these channels versus physical channels, that refusing to provide books in digital channels is unlikely to result in increased short-term physical sales for a particular title.

However, we also note that our results are not without limitations. First, and most notably, our results are based on the assumption that the books released immediately before and after our “experiment” period are good controls for the books released during the experiment. We believe this is true based on the empirical tests above and based on conversations with the publisher in question, however, we cannot conclusively rule out the possibility that unobserved differences between the control and experiment groups may be driving some of our observed results. Second, our results are situated within a particular market (books) and at a particular stage of development of that market. Finally, our results suggest a difference in response for books with strong and weak brand awareness, and for books with weak and strong digital channel preference, a finding that we believe deserves further investigation using data from other industries and settings.

## References

- Amazon.com. 2011. Amazon.com Now Selling More Kindle Titles than Print Books. Press Release, May 19. Available from <http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=1565581>, last accessed August 18, 2011.
- Association of American Publishers. 2010. AAP Reports October Book Sales. December 8. Available from <http://www.scribd.com/doc/44902780/AAP-Reports-October-Book-Sales> (last accessed April 30, 2013).
- Biyalogorsky, E., P. Naik. 2003. Clicks and Mortar: The Effect of On-line Activities on Off-line Sales. *Marketing Letters*, 14(1) 21-32.
- Chevalier, J., A. Goolsbee. 2003. Measuring Prices and Price Competition Online: Amazon and Barnes and Noble. *Quantitative Marketing and Economics*, 1(2) 203-222.
- Coster, H. 2006. "The Secret Life of An Online Book Reviewer," *Forbes*, December 1.
- Danaher, B., S. Dhanasobhon, M.D. Smith, R. Telang. 2010. Converting Pirates without Cannibalizing Purchasers: The Impact of Digital Distribution on Physical Sales and Internet Piracy. *Marketing Science*, 29(6) 1138-1151.
- Danaher, B., J. Waldfogel. 2011. Reel Piracy: The Effect of Online Movie Piracy on Film Box Office Sales. Working paper. University of Pennsylvania, Philadelphia, Pennsylvania.
- Danaher, B., M.D. Smith, R. Telang. 2014. Piracy and Copyright Enforcement Mechanisms, Forthcoming in *Innovation Policy and the Economy*, Volume 14, J. Lerner and S. Stern, eds., University of Chicago Press, Chicago, Illinois.



- Deleersnyder, B., I. Geyskens, K. Gielens, M.G. Dekimpe. 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) 337-348.
- Elberse, A., J. Eliashberg. 2003. Demand and Supply Dynamics for Sequentially Released Products in International Markets: The Case of Motion Pictures. *Marketing Science* 22 (3) 329-354.
- Emmelhainz, M. A., J. R. Stock, L. W. Emmelhainz. 1991. Consumer responses to retail stockouts. *Journal of Retailing* 67(2) 138–147.
- Frambach, R., H. Roest, T. Krishnan. 2007. The Impact of Consumer Internet Experience on Channel Preference and Usage Intentions Across the Different Stages of the Buying Process. *Journal of Interactive Marketing* 21 (2) 26-41.
- Ghose A., M D Smith, R Telang R. 2006. Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact. *Information Systems Research*, 17(1), 3-19.
- Greene, W.H. 2002. *Econometric Analysis*, Prentice Hall, New York, NY.
- Hausman, J., B.H. Hall, Z. Griliches. 1984. Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica* 52, 909-938.
- Hui, K., I. Png. 2003. Piracy and the Legitimate Demand for Recorded Music. *Contributions to Economic Analysis and Policy*, 2(1).
- Kannan, P. K., B. Kline Pope, S. Jain. 2009. Pricing Digital Content Product Lines: A Model and Application for the National Academies Press. *Marketing Science*, 28(4) 620-636.
- Kotler, P. 2002. *Marketing Management*. Prentice-Hall, Upper Saddle River, NJ.

- Lehmann, Donald R., Charles B Weinberg. 2000. Sales through Sequential Distribution Channels: An Application to Movies and Videos. *Journal of Marketing* 64(July) 18-33.
- Luan, Jackie Y., K. Sudhir, 2006. Optimal Inter-Release Timing for Sequentially Released Products. Working paper. Yale School of Management, Yale University, New Haven, Connecticut.
- Luan, Y. J., and Sudir, K. 2010. Forecasting Marketing-Mix Responsiveness for New Products. *Journal of Marketing Research*, 47(3), 444-457.
- Moorthy, K., and I. Png. 1992. Market Segmentation, Cannibalization, and the Timing of Product Introductions. *Management Science* (38:3), 345-359.
- Oberholzer, F., K. Strumpf. 2007. The Effect of File Sharing on Record Sales. An Empirical Analysis. *Journal of Political Economy*, 115(1) 1-42.
- Oberholzer-Gee, F., K. Strumpf. 2010. File-Sharing and Copyright. In *Innovation Policy and the Economy*, Volume 10, J. Lerner and S. Stern eds., University of Chicago Press, Chicago, IL, pp. 19-55.
- Peitz, M, P. Waelbroeck. 2004. The Effect of Internet Piracy on Music Sales: Cross-Section Evidence. *Review of Economic Research on Copyright Issues*, 1(2) 71-79.
- Prasad, Ashutosh, Bart Bronnenberg and Vijay Mahajan .2004. Product Entry Timing in Dual Distribution Channels: The Case of the Movie Industry. *Review of Marketing Science*, 2(1).
- Rich, M. 2009. Steal This Book (for \$9.99). *New York Times*, May 17, p.WK3.
- Rich, M. 2010. Math of Publishing Meets the E-Book. *New York Times*, February 28. Available from <http://www.nytimes.com/2010/03/01/business/media/01ebooks.html>, last accessed April 30, 2013.

- Rob, R., J. Waldfogel. 2006. Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students. *Journal of Law and Economics*, 49(1) 29-62.
- Rob, R., J. Waldfogel. 2007. Piracy on the Silver Screen. *Journal of Industrial Economics*, 55(3) 379-393.
- SeekingAlpha. 2009. Amazon.com, Inc. Q4 2008 Earnings Call Transcript, Q&A. Available from <http://seekingalpha.com/article/117508-amazon-com-inc-q4-2008-earnings-call-transcript> (last accessed April 30, 2013).
- Simester, D., Y. J. Hu, E. Brynjolfsson, E. Anderson. 2009. Dynamics of Retail Advertising: Evidence from a Field Experiment. *Economic Inquiry* 47 (3), 482-499.
- Smith, M., R. Telang. 2009. Competing with Free: The Impact of Movie Broadcasts on DVD Sales and Internet Piracy. *Management Information Systems Quarterly*, 33(2) 312-338.
- Vernik, D., Purohit, D., and Desai, P. 2011. Music Downloads and the Flip Side of Digital Rights Management. *Marketing Science* (30:6), 1011-1027.
- Waldfogel, J. 2007. Lost on the web: Does Web Distribution Stimulate or Depress Television Viewing? *Information Economics and Policy*.
- Zentner, A. 2005. File Sharing and International Sales of Copyrighted Music: An Empirical Analysis with a Panel of Countries. *Topics in Economic Analysis & Policy* 5(1), Article 21.

## Appendix

**Table A1: Analyses of Sales Data for Titles with Many and Few Pages (Weeks 1-20 after Each Version's Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	6.528** (0.580)	6.478** (0.941)	7.490** (0.603)	7.485** (1.040)
<i>Experiment*FewPages</i>	0.307** (0.056)	-0.558** (0.077)	0.044 (0.056)	-0.928** (0.094)
<i>Experiment*ManyPages</i>	-0.246** (0.044)	-0.605** (0.061)	-0.346** (0.045)	-1.058** (0.081)
<i>NumWeeks</i>	-0.134** (0.005)	-0.124** (0.008)	-0.122** (0.005)	-0.038** (0.010)
	Book	Book	Book	Book
Other Control Variables	characteristics, Weekly dummies	characteristics, Weekly dummies	characteristics, Weekly dummies	characteristics, Weekly dummies
Log Likelihood	-18,995.5	-16,556.5	-20,982.5	-15,758.2
Number of Obs.	3,640	3,640	3,640	3,640

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release

**Table A2: Analyses of Sales Data for Strong-Awareness and Weak-Awareness Titles, with iBook Sales (Weeks 1-20 after Each Version's Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	7.591** (0.590)	6.058** (0.942)	8.061** (0.611)	7.309** (1.038)
<i>Experiment*StrongAwareness</i>	0.238** (0.053)	-0.514** (0.073)	-0.063 (0.055)	-0.987** (0.094)
<i>Experiment*WeakAwareness</i>	-0.266** (0.047)	-0.550** (0.065)	-0.297** (0.048)	-1.035** (0.085)
<i>NumWeeks</i>	-0.137** (0.005)	-0.135** (0.007)	-0.125** (0.005)	-0.042* (0.010)
	Book	Book	Book	Book
Other Control Variables	characteristics, Weekly dummies	characteristics, Weekly dummies	characteristics, Weekly dummies	characteristics, Weekly dummies
Log Likelihood	-19006.7	-16611.5	-20668.8	-16124.1
Number of Obs.	3,640	3,407	3,494	3,494

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release

**Table A3: Analyses of Sales Data for Titles with Low and High Weight, with iBook Sales (Weeks 1-20 after Each Version's Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	6.630** (0.579)	6.227** (0.933)	7.576** (0.626)	7.505** (1.042)
<i>Experiment*LowWeight</i>	0.334** (0.058)	-0.252** (0.080)	0.088 (0.060)	-0.740** (0.101)
<i>Experiment*HighWeight</i>	-0.278** (0.046)	-0.699** (0.065)	-0.363** (0.047)	-1.168** (0.084)
<i>NumWeeks</i>	-0.137** (0.005)	-0.137** (0.008)	-0.124** (0.005)	-0.043** (0.010)
Other Control Variables	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies
Log Likelihood	-18,993.8	-16,596.7	-20,651.4	-16,114.2
Number of Obs.	3,640	3,407	3,494	3,494

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release

**Table A4: Analyses of Sales Data for Strong-Awareness and Weak-Awareness Titles (Weeks 1-40 after Each Version's Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	7.513** (0.547)	6.850** (0.833)	8.076** (0.563)	5.891** (0.872)
<i>Experiment*StrongAwareness</i>	0.271** (0.044)	-0.603** (0.063)	-0.083 (0.044)	-0.606** (0.065)
<i>Experiment*WeakAwareness</i>	-0.351** (0.039)	-0.569** (0.057)	-0.320** (0.038)	-0.637** (0.059)
<i>NumWeeks</i>	-0.103** (0.004)	-0.091** (0.005)	-0.089** (0.004)	-0.065* (0.006)
Other Control Variables	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies
Log Likelihood	-25,166.1	-21,200.3	-28,540.6	-22,102.2
Number of Obs.	5,320	4,894	5,320	5,320

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release

**Table A5: Analyses of Sales Data for Titles with Low and High Weight (Weeks 1-40 after Each Version's Release)**

	<i>Print Sales</i>	<i>Digital Sales</i>	<i>Total Sales</i>	<i>Digital Sales After Print Release<sup>†</sup></i>
Constant	6.341** (0.538)	7.139** (0.830)	7.693** (0.553)	6.115** (0.876)
<i>Experiment*LowWeight</i>	0.358** (0.049)	-0.435** (0.069)	0.038 (0.049)	-0.439** (0.073)
<i>Experiment*HighWeight</i>	-0.352** (0.039)	-0.669** (0.057)	-0.372** (0.038)	-0.726** (0.058)
<i>NumWeeks</i>	-0.103** (0.004)	-0.092** (0.005)	-0.089** (0.004)	-0.066** (0.006)
Other Control Variables	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies	Book characteristics, Weekly dummies
Log Likelihood	-25,157.0	-21,195.1	-28,522.8	-22,095.4
Number of Obs.	5,320	4,894	5,320	5,320

Standard errors are in parentheses. \*\*Significantly different from zero,  $p < 0.01$ . \*  $p < 0.05$ . <sup>†</sup> Sales are shown for weeks 1-20 after print release