

# The Effect of Piracy Protection in Book Publishing

Imke Reimers\*

This version: October 6, 2014

## Abstract

Digitization has impacted media industries by lowering the cost of distributing cultural goods. With the option to download creative works legally, however, comes the possibility of doing so illegally. Efforts to prevent such piracy have mostly been public in nature, with mixed receptions. The book publishing industry showcases a more private approach to piracy protection, in which publishers hire private companies to protect their titles. I exploit a natural experiment where different book titles are protected by a piracy protection service at different times. I compare sales of similar book titles with and without piracy protection, before and after titles have been added by the service, in a difference-in-differences setting. I find that the effect of piracy protection varies across formats and titles. E-books, the closest substitute for online piracy, benefit from piracy protection by selling over 11% more units, while there is no significant effect on other formats. The effect is more pronounced if more effort is exerted. The effect is also larger for titles that have been successful prior to protection, indicating that book piracy has a promotional effect for lesser known works. The results are robust to different specifications.

---

\*Assistant Professor of Economics, Northeastern University, 304 Lake Hall, 360 Huntington Avenue, Boston, MA 02115, i.reimers@neu.edu

# 1 Introduction

Digitization has significantly decreased the cost of creating and distributing cultural goods in media industries. Consequently, digitization has affected firm and artist profitability in many settings and cases. Production and copying of creative works have become cheaper and easier, increasing the variety and quantity of products available for consumption. The wider variety could increase the level of competition and lower prices further. In many cases, these effects shift surplus from producers to consumers.

Digitization has another effect: as the *legal* distribution of creative works has become easier, so has the *illegal* distribution of those works. Thus, the question of the optimal intellectual property (IP) strategy arises for each artist and distributor, and in each industry. How does illegal distribution of content affect the legal sales of a work? Modern technology has also made monitoring and regulation of illegal activity more feasible. Can works be effectively protected from piracy? If they can, what should this protection look like and who should enforce this protection? This paper addresses these questions in detail using a novel dataset from the book publishing industry.

A large and growing literature addresses the effect of file sharing in the music and movie industries.<sup>1</sup> The findings are divided, with some work indicating that there is no significant effect (e.g. Oberholzer-Gee & Strumpf, 2007), while more recent work has found that regular sales have been significantly displaced by pirated versions (e.g. Zentner, 2006; Liebowitz, 2008; and Waldfogel, 2010). In the music industry, file sharing seems to have significantly decreased the legal sales of songs and albums.<sup>2</sup> The negative effect can in part be attributed to the arrival of Napster in 1999, a file sharing website that revolutionized the industry. The first legal option to download music did not arrive until almost two years later, when Apple introduced its iTunes Store in May 2003.

More recently, research on file sharing has turned to determining whether anti-piracy efforts can be effective. Most efforts to protect creative works have been public in nature, and many of these efforts have not been very well-received. Consider, for example, the internet blackout to campaign against the Stop Online Piracy Act (SOPA) and the Protect IP Act (PIPA) in January 2012.<sup>3</sup> Similarly, the French Parliament passed the Creation and Internet Law, an anti-piracy law

---

<sup>1</sup>See Smith and Telang (2012) for a detailed description of studies on those industries, and Peitz & Waelbroeck (2006) for a review of the theoretical literature.

<sup>2</sup>See Rob & Waldfogel (2004)

<sup>3</sup>See <https://www.eff.org/issues/coica-internet-censorship-and-copyright-bill>.

more commonly known as HADOPI, in 2009. This bill introduced a three strike policy in which consumers of illegal music were cut off the internet after two warning notices. While Danaher et al. (2014) find that awareness of this law increased legal music sales on iTunes, the bill faced controversy as France's Constitutional Council declared access to the internet a basic human right, and it was finally revoked in July 2013.<sup>4</sup>

Other efforts to prevent piracy tend to be public or at least concentrated in nature as well. In the movie industry, the abrupt shutdown of the cyberlocker Megaupload.com in January 2012 had a significant positive effect on box office revenues and digital movie sales of popular works, while the shutdown did not seem to have a positive effect on box office revenues of less well-known works (Peukert, Claussen & Kretschmer, 2013; Danaher & Smith, 2013). In music, the Recording Industry Association of America (RIAA) posed legal threats against file sharers, decreasing the level of file sharing although a substantial amount of illegal files remains available (Bhattacharjee et al. 2006). Another effort by the major music labels to control secondary distribution of content through digital rights management (DRM) does seem to have a positive effect on sales (Zhang, 2013).

Private efforts to prevent piracy seem to face less controversy than public efforts, but it is not clear whether we can rely on private action against copyright infringement, and if the gains from such protection outweigh the effort to protect a work. The book publishing industry provides a setting that allows me to answer these questions. In the book publishing industry, the effort to prevent piracy is led by private firms that are hired by the publishers. The largest of these firms is Digimarc, which searches the internet for infringing content specific to protected titles. Protection from piracy thus is title and publisher specific, rather than industry wide, and it is initiated by publishers and the firms they work with, rather than a public entity.

I estimate the effect of this anti-piracy effort in the book industry. I use a novel dataset consisting of weekly physical and monthly electronic book sales of titles that are offered electronically by one particular publisher (Rosetta Books), from 2010 to 2013. The dataset includes the level and success of piracy protection over the same time period, through the leading book piracy protection company, Digimarc. I follow sales of each title before and after their title-specific piracy protection begins in a difference-in-differences setting. I check whether (and how much) sales rise

---

<sup>4</sup>See <http://www.theguardian.com/technology/2013/jul/09/france-hadopi-law-anti-piracy>.

or fall when protection for a title begins, relative to other titles whose protection begins at another time. Moreover, I provide evidence that the timing of protection is consistent.

Piracy protection for a title involves two major steps: asking search engines to de-list the link to the pirated content, and requesting that the pirated content is taken down directly through the site host. Smith et al. (2014) show that the presence of pirate links in search results strongly influence the piracy behavior of consumers of movies. I separate the effectiveness of de-listing a link from search engines and of taking down infringing content in the book publishing industry.

By examining titles that have enjoyed different levels of reader attention in the past, I analyze the interaction of two demand-side effects of illegal downloading. First, a free version of the same product (title) competes with the legal nonzero-price options, potentially displacing legal sales.<sup>5</sup> Second, though, free options have a sampling and advertising effect, and this potentially increases legal sales.<sup>6</sup> Demand for fiction novels is reputational and depends in part on word-of-mouth advertising. This is also true for nonfiction books, albeit to a lesser degree.

Not surprisingly, I find that the effect of piracy protection on legal book sales depends on the awareness level of the title, the type of work (fiction or nonfiction), and the format of the edition. While physical formats are not affected by piracy protection, closer substitutes to online piracy, such as legally distributed e-books, see a mean differential increase in sales of 21.8%. This effect seems to come from the de-listing of links from search engines rather than a decrease in piracy sites offering the title, although the difference in the two effects is not significant. The effect of protection is stronger the more effort is exerted, and it takes its strongest effect a few months after the start of protection.

The increase in e-book sales is larger and more significant among popular works, while e-book sales of less well-known works do not change significantly as a result of piracy protection, and demand for nonfiction titles seems to depend less on word-of-mouth advertising than demand for fiction titles.

Lastly, it is possible that limiting piracy for a title has a direct effect on supply side considerations. Oestreicher-Singer and Sundararajan (2013) show that digital rights management has an economically significant effect on pricing decisions of digital content. Preventing piracy effectively

---

<sup>5</sup>See Smith and Telang (2012).

<sup>6</sup>See Peitz and Waelbroeck (2006b).

limits the level of competition in the market. Theory predicts a resulting increase in prices. I test for this among e-books, and I do not find a significant change in prices as a result of piracy protection. It seems that book piracy protection has a larger effect on consumer behavior than on publisher behavior when the book has already been written and published. This result adds to other tests suggesting that endogeneity of piracy protection is not a major issue in my analysis.

The remainder of the paper proceeds as follows. Section 2 provides background on the publishing industry and on the role of piracy. I describe the empirical strategy of the experiment and provide intuition for the identification in section 3. The data are described in detail in section 4. I proceed with showing results in section 5, and I present robustness checks in section 6. I conclude with some implications in section 7.

## 2 Background and the Piracy Protection Experiment

Technological change has transformed media industries such as music, newspapers, movies, television, and books. The recorded music industry faced challenges from digitization when the Napster file-sharing service arrived in 1999. Since then, digitization has posed challenges to other content industries as well. Newspaper revenues, for example, have fallen by half since the late 1990s.<sup>7</sup> Some (but not all) of these challenges can be attributed to increased competition through free online versions.

The effect of file sharing and the effort to stop it is not yet well-studied in book publishing, as digitization in this industry has lagged behind other industries. While e-books can be read on computers and have therefore been available for over a decade, e-books are most useful to consumers when read on small hand-held devices, such as e-book readers or tablet computers. Sony had released some electronic book readers as early as 2004, but the first widely adopted e-reading device was Amazon's Kindle, which was introduced in November 2007, with no widespread illegal option to download present at the time. The e-reader and e-book markets have grown quickly since then. The share of US adults owning an e-book reader grew from 2 percent in April 2009 to 24 percent in September 2013, and the share of adults owning either an e-reader or a tablet (some

---

<sup>7</sup>See <http://stateofthemediamedia.org/2013/newspapers-stabilizing-but-still-threatened/newspapers-by-the-numbers/purchases-by-volume/>.

viable method for consuming e-books) reached 50 percent in January 2014.<sup>8</sup> Electronic books have similarly become increasingly popular in the past years. While e-books held a negligible market share among fiction books before 2007, about half of the weekly top 150 bestseller books have been sold primarily as e-books by late 2012 (see Waldfogel & Reimers, 2014).

With the option to download content online comes the possibility of doing so illegally. In other content industries, online piracy has been relatively concentrated. Although many book titles are available on well-known piracy sites (for example, Piratebay), the book industry does not face one large file sharing service, as was the case with Napster (music). The small size of book files as compared to music and movies makes it convenient to store them as intact files, so that many infringing files are located on websites rather than on peer-to-peer sharing services, making tracking and taking down illegal content more feasible.

Publishing houses have increasingly hired private firms in an attempt to limit the amount of online book piracy. These firms search for websites, cyberlockers, and file sharing services that offer specific book titles. This approach implies that books are protected individually.

This paper follows titles that are protected by Digimarc Guardian, one of the industry's leading anti-piracy services.<sup>9</sup> Its piracy protection strategy includes an automated process that finds suspected pirated content, followed by two tiers of human verification of those sites. The sites with confirmed pirated content are subject to two types of treatment: Digimarc sends requests to the search engines Google and Yahoo! to de-list the links, and it sends take down notices to the provider of the illegal content, essentially until the content is taken down.

These two actions affect different types of pirates. The de-listing from search engines may prevent "recreational" consumers of illegal content - those who rely on search engines to find illegal content - from reading pirated versions, while the taking down of infringing sites has a better chance of diverting more experienced pirates. Digimarc continues to search for infringing sites for the title as long as the title is under contract.

Digimarc has added most of the large publishing houses as clients over the past five years. The publishing companies include HarperCollins (February 2010), Macmillan (October 2010, al-

---

<sup>8</sup>See <http://www.marketingland.com/pew-50-percent-in-us-now-own-tablet-or-e-reader-70765>.

<sup>9</sup>The service was introduced by a company called Attributor, which was acquired by Digimarc in December 2012. It is now known as Digimarc Guardian.

though Macmillan now uses another service), Simon & Schuster (October 2011), and Random House (June 2013, now Penguin-Random House).<sup>10</sup> In addition, the company has added several smaller publishers.<sup>11</sup> Rosetta Books, which has signed with the company in June 2011, is one of those publishing houses.

This publisher has secured exclusive rights to publishing electronic versions of over 600 titles.<sup>12</sup> Rosetta's list of titles consists in large part of backlist titles, ranging from well-known classics (including, among others, works by Kurt Vonnegut, Arthur C. Clarke's *Space Odyssey* Books, and Stephen Covey's *The 7 Habits of Highly Effective People*), to works that are less well known today. The publisher also carries some original titles that are available only in electronic format. Digimarc started searching for, de-listing, and taking down, different sets of titles in Rosetta's catalog at different dates after June 2011.<sup>13</sup>

### 3 Empirical Strategy and Identification

Any analysis examining the effect of piracy protection on book sales is faced with a fundamental inference problem. For any given title that is protected from piracy I do not observe the counterfactual: what would the sales of this title have been if it had not been protected from piracy? However, I can observe and compare similar titles with and without piracy protection.

In an ideal experiment that determines the effect of piracy protection, the econometrician assigns titles randomly across groups and protects one set of titles. The marginal effect of online piracy protection on book sales is the change in sales of the treated titles as compared to the change in sales of the group of titles that was not treated. The empirical strategy takes advantage of the change in protection in a difference-in-differences analysis. The log of the (observed) sales  $q_{it}$  of

---

<sup>10</sup>[http://www.mediabistro.com/galleycat/harpercollins-and-harvard-business-school-publishing-will-use-attributors-piracy-protection-program\\_b11178](http://www.mediabistro.com/galleycat/harpercollins-and-harvard-business-school-publishing-will-use-attributors-piracy-protection-program_b11178),  
<http://www.marketwired.com/press-release/Contributor-Macmillan-Kensington-Publishing-Corp-Lead-Global-Initiative-Educate-1330505.htm>,  
<http://www.publishersweekly.com/pw/by-topic/digital/copyright/article/48974-s-s-signs-with-contributor.html>, and  
<http://randomnotes.randomhouse.com/anti-piracy-reporting-tool-added-to-author-portal/>.

<sup>11</sup>See <http://www.digimarc.com/guardian/customers> for a list of Digimarc's customers.

<sup>12</sup>While electronic editions can only be published by Rosetta, physical editions can be published by other publishers. See <http://go-to-hellman.blogspot.com/2011/11/creative-commons-media-neutrality-and.html> for more information.

<sup>13</sup>The timing is almost entirely determined by the publisher. Some customers opt for full catalog coverage (with new titles added as they approach release), while others choose to protect only a smaller set of titles which may evolve over time (often handled as monthly updates). Rosetta seems to fall into the latter category of publishers.

title  $i$  in week  $t$  are a function of the works' observable characteristics and its protection status  $protection_{it}$  in that week.

The effect of this piracy protection is identified because I observe sales of titles when they are protected and when they are not protected. Since the move into piracy protection happens at different times for different titles, it is unlikely that any changes in sales when a title moves into piracy protection are due to an exogenous time-dependent shock that affects overall demand for books. Moreover, a change in the level of piracy for one title is unlikely to affect legal demand for another title (while it might affect the level of piracy for other titles). I estimate the effect of piracy protection as

$$\log(q_{it}^k) = \alpha \cdot protection_{it} + \delta_i^k + \mu_t^k + \epsilon_{it}, \quad (1)$$

where  $q_{it}^k$  denotes the unit sales of title  $i$  in week  $t$  in format  $k$ , where  $k \in \{\text{hardcover, trade paperback, mass market paperback, audio, and e-books}\}$ , and  $protection_{it}$  is a dummy variable that is 1 if the title is under piracy protection in week  $t$ . The effect of book piracy protection on regular book sales is given by  $\alpha$ , where book piracy protection causes a change in regular sales of  $e^\alpha - 1$  percent.<sup>14</sup> The log specification assumes that piracy changes sales by a common percentage, rather than by a common absolute amount. One would expect that the absolute sales of a title by Stephen King will be affected more strongly by piracy than a more obscure title that is not consumed much at all.

I also make use of the fact that I observe unit sales of over 130 titles (depending on format), over a period of 170 weeks, by controlling for title  $i$  and format  $k$  ( $\delta_i^k$ ) and adding week-year fixed effects ( $\mu_t^k$ ), which are allowed to vary across formats (e-books have become more popular over time). The title fixed effects pick up the overall popularity of the title, as well as genres and the author's level of popularity. Time fixed effects control for changes over time that affect all titles similarly, such as changes in the economic environment, or the release of a new e-reader or tablet.<sup>15</sup>

The identification strategy assumes that the timing of piracy protection is uncorrelated with

---

<sup>14</sup>As I observe some periods with zero physical sales for some titles, I use  $\log(q_{it}^k + 0.00001)$  as the dependent variable for physical formats. In that case, the marginal effect is interpreted with a bias. This bias understates the true effect but goes to zero quickly (the bias is  $\frac{\Delta q}{q(q+0.0001)}$ ).

<sup>15</sup>Since the titles in my dataset have originally been published several years ago, I do not include a time trend that follows the work's time since publication in the main regression. I do, however, include a quadratic function of age and an indicator variable that equals 1 if a new edition has recently been published in section 6.1 to show that the results are robust.

factors that determine the outcome of interest - in this case, log-unit sales. However, the decision to protect a title from piracy may well be endogenous: a publisher may be more interested in protecting those titles that have gained popularity lately than in protecting those that do not seem to attract much reader interest at all. For example, the publisher would have liked to increase piracy protection for Richard Matheson's 1954 novel, *I Am Legend*, when it was adapted as a blockbuster movie in 2007. The change in regular book sales would then be the result of a combination of two effects: an increase in demand due to the movie's promotional effect on the novel, and the change in demand from a change in piracy protection. Disentangling these effects is difficult.

I address the potential endogeneity of piracy protection in three ways. First, running the regression from equation (1) separately for protected titles and for titles that are never protected shows that even after controlling for protection, sales for unprotected titles decrease significantly more over time than sales for protected titles. I therefore ignore those titles that never enter piracy protection as those titles may be inherently different from those that move into piracy protection.

Second, most titles in my dataset are "old" - most of them have been published before 2000 - so that their demand does not change much over time, and it is unlikely that a change in demand is caused by title specific market changes. Following the demand for titles over time shows no irregular increase in demand shortly before they move into piracy protection.

Lastly, it is possible that a change in piracy protection is accompanied by supply side reactions. For instance, eliminating a zero-price competitor can affect a publisher's pricing strategy. I include e-book prices in several estimations of the effect on e-book sales to account for such a supply side adjustment.

As piracy protection has two elements, I use both the overall protection and a measure for the effectiveness of take down notices as my independent variables to identify the leading mechanism behind the success of piracy protection. I also estimate the role of intensity of protection by analyzing the effect of different frequencies of protection. I further separate the effect of piracy protection on well-known works and less well-known works, and on fiction and nonfiction titles to determine what types of works are more likely to be protected successfully. While my analysis is restricted to one publisher, the observed titles vary widely in popularity, quality, and genre, allowing for some inference to other publishers and works.

## 4 Data

I follow the demand for and piracy protection of a set of book titles over four years, from 2010 to 2013. The underlying dataset consists of 653 titles whose electronic versions are exclusively published through RosettaBooks.<sup>16</sup> Most of the works have been originally published several years ago, going back as far as the first half of the twentieth century. There also are a few titles that have been published in the past ten years. The analysis includes the subset of the titles whose piracy protection status has changed between 2011 and 2013.

On the demand side, I observe weekly sales data for physical book formats: hardcover, trade paperback, mass market paperback, and audio versions, through the Nielsen BookScan database. These sales date back to 2002, and I focus on the time period from 2010 to 2014. They include all editions of the title that are listed by Nielsen.

I also observe monthly e-book unit sales on the title level directly through Rosetta (the exclusive publisher of the titles' e-books), from July 2011 to December 2013. These editions are sold through Amazon, Apple's iBooks store, and Barnes and Noble. The publisher carries a few popular titles that sell far more than 100,000 copies annually, and some that sell less than 1,000. Table 1 shows mean title sales in each format in 2013. Trade paperbacks and e-books seem to be the most popular formats.

On the piracy dimension, I obtain detailed information through Digimarc, the company that protects Rosetta's works from illegal file sharing. I observe when the title was taken under contract to be protected, the date that each illegal link to each title was found, when take down notices were sent to websites offering specific titles, and when these piracy sites were taken down. The titles in my dataset have 31 different protection start dates between June 2010 and December 2013. Titles are protected with different priority levels, with 1 being the lowest, and 3 being the highest. The priority level of a title can vary over time, although it does not do so often.

Digimarc searches the Internet for websites, cyberlockers, and peer-to-peer services that make books available illegally for free. The company sends take down notices to those sites with increasing levels of urgency (and threats of legal action), until the site is taken down. Table 2 shows summary statistics on the number of sites found per title, the number of notices sent per

---

<sup>16</sup>See <http://www.rosettabooks.com/books/books-a-z/> for Rosetta's catalog.

site, and measures of the company's success, for all titles that were protected at some point during the observed time period, including those that were never without protection.

The anti-piracy service seems to have some success in reducing file-sharing, and this success seems to come reasonably quickly. Over 50% (10,417) of all infringing sites are taken down within one week of them being found. Yet, it is possible that Digimarc does not find all websites, and that taking down one website will not prevent another one from offering the same title. The fact that the company finds close to 88 sites per title on average, and it successfully takes down over 83% of those, indicates that the company at least makes a dent in the online book piracy landscape, at least for the titles that Rosetta Books covers. The relatively high success rate may be due to the nature of piracy sites. As book files are small, there are more small sites with infringing material, and the operators of such small sites may be more easily intimidated by such take down notices.

Despite Digimarc's apparent success, the number of detected piracy sites that have not been taken down increases on average over time since not all take down notices are successful. Figure 1 illustrates the cumulative number of sites that are detected by Digimarc, those that are successfully taken down, and the difference between the two as the number of sites that are currently active, for Richard Matheson's *I Am Legend*.<sup>17</sup> The number of active sites decreases at times, allowing me to use the net change in the number of sites as a measure of success of take down notices.

Google's transparency project and ChillingEffects help determine which links have been requested to be de-listed. Most of the links are in fact taken off the search engine. Between May 2013 and September 2014, only 358 out of 10,198 requests (3.5%) from Digimarc regarding Rosetta Books have not been honored. However, there is no information about which links remained up and, more importantly, *when* the link was de-listed. I therefore do not use the actual de-listing as an independent variable. Instead, I approximate the effect of de-listing as the difference between the effect of the title moving into piracy protection at all (de-listing plus take downs), and the effect of a decrease in active sites (successful take downs of pirated content that outweigh the emergence of new sites).

I collect additional information on the titles' popularities and qualities through the crowd-sourced online book database Goodreads. This website provides over 10 million book reviews across

---

<sup>17</sup>The fact that I do not observe the number of infringing sites before protection, and that the number of active detected sites increases over time, make it difficult to use protection as an instrument for the level of piracy. For that reason I instead estimate the effect of protection itself.

700,000 titles.<sup>18</sup> The list of Rosetta’s titles includes more fiction than nonfiction works as 85% of the e-books and 74% of the physical titles in the analysis are fiction works. The mean (median) number of reviews for the titles in my analysis is 4,855 (483) for ebooks, and 20,821 (1,500) for physical books, and the mean (median) star rating (on a scale from 1 to 5) is 3.75 (3.75) for e-books, and 3.85 (3.86) for physical books. I use this information when disentangling the effect of piracy on different types of titles.

## 5 Results

I follow unit sales observations for different formats of the titles from January 2010 to December 2013 (physical sales), and from July 2011 to December 2013 (e-book sales). I estimate several versions of equation (1) for different formats.

### 5.1 Baseline: Piracy Protection

Table 3 shows the baseline estimation results for all titles that moved into protection with Digimarc at some point during that time period. The independent variable of interest is an indicator variable that is one if the title is under contract with Digimarc. That is, a title is protected for all time periods after the first site has been found, regardless of the number of sites that were found and/or taken down in any given time period. The dependent variable is log-unit sales of format  $k$  of title  $i$  in time period  $t$ . Note that the number of titles is different across formats because some titles are not available in all formats.

The effect of piracy protection on legal book sales depends on the edition’s format. Physical editions of a title are not necessarily close substitutes to free electronic versions of the same title. Consequently, the effect of piracy protection is insignificant at the 10% level for the physical formats, and even negative (but highly insignificant) for hardcovers and trade paperbacks. Generally, the effect on physical formats is estimated with a lot of error, causing me to focus much of my analysis on e-books.

E-books can be regarded as the closest substitutes to pirated versions. Column 5 shows that piracy protection has a large and statistically significant positive effect (at 5%) on sales of

---

<sup>18</sup>See <http://www.goodreads.com/api>.

e-books. Moving a title into piracy protection increases regular e-book sales by 21.8% ( $= e^{0.197} - 1$ ). It seems that protection against piracy can be effective for e-books.

## 5.2 Timing

It seems that protection from piracy significantly increases legal e-book sales, and how promptly this piracy protection takes effect can give us more insight into how this protection works. I analyze this question for e-books, using the time (in months) before and after protection starts as the explanatory variables of interest as in equation (2):

$$\log(q_{it}) = \left( \sum_{m=-6}^6 \alpha_m \cdot 1(t - \text{protection\_start} = m) \right) + \delta_i^k + \mu_t^k + \epsilon_{it}, \quad (2)$$

where *protection\_start* is the first month of protection.

Figure 2 shows that piracy protection takes around 4 months to have an effect, indicating that it takes Digimarc quite some time to find and deactivate “enough” infringing content. This also suggests that the age of a title plays an important role when determining the effectiveness of protection. New titles need to be protected quickly as traditionally most units of a title are sold within the first few months of its initial publication.

Figure 2 also addresses concerns about endogeneity of piracy protection. It is possible that a title is protected because it has just become popular. This concern would be larger for more recent titles than the ones in this dataset since works that are several decades old are less likely to see a sudden spike in overall demand.

Moreover, if piracy protection was endogenous, demand for the title would increase in the time periods leading up to piracy protection. This does not seem to be the case for e-books (nor for other formats) as the percent change in unit sales leading up to protection is close to zero, and never significantly positive.

## 5.3 Supply Side Effects - Prices

While I find no evidence that sales increase right before a title becomes protected, it is possible that the publisher adjusts other attributes of their legal editions when the title moves into piracy

protection. For example, taking away the zero-price option decreases the level of competition for the remaining editions. As a result, one might expect a change in the price level of those editions.

If prices increase as a result of the lower level of competition, then the positive effect of piracy protection on regular unit sales would be countered by a negative effect of higher prices on unit sales, and the above results would underestimate the true effect of piracy. If, on the other hand, prices decrease when a title is protected, then the results reported above would be overestimated: my estimates would pick up a combination of piracy protection and lower prices.<sup>19</sup>

I address this using monthly average e-book prices. A difference-in-differences analysis with e-book prices as the dependent variable indicates that e-book prices decreased by 6.3% as a result of a move into piracy protection, although this effect is insignificant at 10%. A negative correlation of prices and protection would indicate that the results above overestimate the effect of piracy protection.

Table 4 shows that this overestimation does not significantly change the effect of piracy protection on unit sales, although the effect loses some statistical significance. Column 2 shows that controlling for price changes, piracy protection still increases regular unit sales by 11.4%. This effect remains significant at the 10% level. However, since prices do have a significant effect on sales (with a price elasticity of -1.39), I control for e-book prices in the remainder of the paper.

Column 3 of table 4 also shows that piracy protection has a significant positive effect on revenues for RosettaBooks, increasing revenues from e-book sales by 14.0%. The publisher's decision to protect a title from piracy then should depend on the cost of protection compared to this benefit. In a broader sense, protection is efficient if the effort it takes to protect a title is smaller than the gain from protection.

## 5.4 De-listings and Takedowns

The above estimates show the combined effect of exactly two actions: de-listing links from search engines and taking down sites with infringing content. In order to guide anti-piracy efforts in a

---

<sup>19</sup>A decrease in prices as the level of competition decreases seems counterintuitive at first, but it is possible. When a zero-price file sharing option is available, the publisher of the higher-quality, nonzero price option may target high-valuation consumers exclusively. Without the low-quality competitor, the publisher can target all types of consumers, but may have to lower its price to reach the low-valuation consumers. See Petrin (2002) for a structural analysis of this effect in the car industry.

more efficient direction, it is instructive to know which of these actions causes the effectiveness of piracy protection. I add another explanatory variable to equation (1): the direction of the change in the number of active sites for title  $i$  compared to the previous time period. A decrease in the number of active sites indicates that take down notices were successful in that time period, while an increase in active sites indicates that more piracy sites emerged than were removed. I further control for the overall piracy “supply” by including the number of detected sites that have not yet been taken down for each title.

The relative effect of take down notices and de-listings can be inferred from the relative significance of the two coefficients. If a decrease in active sites (successful takedown) does not significantly increase legal sales for a title, it must be that most of the effect of protection is due to the de-listing of links. Alternatively, if the success of take down notices has a significant positive coefficient while making the effect of overall protection insignificant, then the effectiveness of piracy protection must be due mostly to take down notices rather than de-listings.

Table 5 suggests that de-listings have a stronger effect on e-book sales than do take downs. The coefficient on overall protection remains almost unchanged while a successful attempt to take down infringing sites (a decrease in the number of sites for a title) has a small and insignificant effect. The difference between the effects is insignificant though, and both actions may be critical in the success of e-book piracy protection. Again, neither type of protection significantly affects physical book sales.

## 5.5 Intensity of Piracy Protection

Protecting a creative work from piracy requires some effort. Digimarc manually checks every suspected infringements before sending out de-listing requests and take down notices. The success of piracy protection needs to be assessed in relation to the effort it takes to protect the work. The level of effort depends on the frequency with which a titles or site is crawled. Digimarc protects titles with three levels of priority as guided by the observed level of piracy. The more a title’s copyright is infringed, the more regularly its contents will be crawled for on the internet, and the more effort the title’s protection requires.

Titles that are assigned a low priority are crawled approximately once every month, while

medium priority corresponds to crawling about once per week, and high priority means the title is searched for every day. After controlling for the differences in the number of active piracy sites, one might assume that high priority titles are about 7 times as costly to protect as are medium priority titles. Are those differences in costs outweighed by differences in effectiveness?

Table 6 shows that daily crawling has a much larger positive effect than more sporadic crawling. When a title is protected with high priority, protection increases e-book sales significantly, with a point estimate of 47.7%. The effect for low and medium priority protection is less clear. The trend holds true for mass market paperbacks - a format that can be seen as a close substitute to online piracy - as well.

## 5.6 Identifying Displacement and Promotion Effects

The above results show the overall effect of piracy protection on unit sales, but book piracy can affect different titles in different ways. The effect of piracy depends on the title's popularity as well as the readers it attracts. These differences are explored in the following two sections. In this section, I analyze in which ways the effect of piracy protection depends on the title's popularity. In the next section, I explore the importance of a title's genre.

Piracy has two counteracting effects. The illegal versions can steer consumers away from legal options (displacement effect), but the consumer may also find it easier to hear about a book if free versions are more easily available, potentially increasing legal sales (promotional effect). The promotional effect could be large enough that a publisher might be harmed by piracy protection for some of its books. The relative extent of these effects depends on the past success of a title, as well as its (perceived) quality.

If a title has been well-known for many years, it does not rely on an additional promotional effect of pirated versions. For well-known and successful titles, book piracy mostly displaces sales of legal editions. On the other hand, book titles that are not very well-known yet rely more heavily on word-of-mouth advertising. For those titles, a free pirated version of the book can actually spur demand for the title by making it more widely known and accessible to more people.

I use the number of Goodreads reviews as a proxy for each title's success and level of consumer awareness. For instance, Stephen Covey's *The 7 Habits of Highly Effective People* has

162,647 reviews on Goodreads, while the median e-book title in my dataset has 483 titles.<sup>20</sup> I use this variation across titles to separately identify the two effects by interacting the protection term with an indicator that is one if the work is well-known in equation (1) by treating a title as well-known if it has more than 500 reviews, and as obscure if it has fewer reviews.<sup>21</sup> The results are robust to different cutoff points.

The size and even the direction of this promotional effect depends on the title’s quality. If the title is “good” (its readers like it), the promotional effect is expected to be positive and rather strong. If on the other hand the title is “bad” (its readers do not like it), a reader may discourage others from buying the book. This is especially true for more obscure titles. Hence I interact a measure of title quality with the protection variables for the obscure titles in a triple-differences analysis. To measure a title’s quality, I use its rating on Goodreads, which goes from 1 to 5 stars.<sup>22</sup> The lowest observed title rating is 3.0, and the mean ratings are 3.75 for e-books, and 3.85 for physical books. I treat a title as “good” if its Goodreads rating is above 3.8. Again, the results are robust to different cutoff points.

Table 7 summarizes the effect of piracy protection on well-known titles and on those that are more obscure. I control for the number of active infringing sites to ensure the difference across popularities is not due to different levels of piracy as more popular titles are more likely to be pirated more.

As expected, the effect on electronic versions of popular titles seems to be larger than that on more obscure titles, although the difference is not significant at the 10% level. After controlling for the number of infringing sites and differences in prices, sales of popular titles increase by 17.0%, while sales of the more obscure titles increase by 7.9 to 8.6%, where this effect is estimated with some error. This differences supports the notion that obscure titles have more to gain from piracy, while well known titles mostly face a displacement effect of piracy. The differences across other formats are less clear.

The work’s quality does not significantly change the effect of piracy protection, as “good” unknown titles benefit more from protection (or are hurt less by it) than “bad” unknown titles,

---

<sup>20</sup>Number of reviews as of August 2, 2014.

<sup>21</sup>Using Goodreads reviews as a measure of popularity rather than actual sales numbers helps avoid one source of endogeneity as actual sales are the outcome of interest.

<sup>22</sup>Four of the e-book titles do not have a Goodreads rating.

although the difference is insignificant at 10% for all formats except mass market paperbacks. This may be due to the fact that in a world in which information becomes more and more readily available (for instance, through Goodreads), potential readers do not rely on close peers to form an expectation on the work's quality, regardless of how obscure the title is.

## 5.7 Fiction and Nonfiction Titles

The effect of piracy protection further depends on the type of demand that the work faces. Some genres benefit more from word-of-mouth advertising than others. In the extreme case, demand for textbooks likely remains constant over time (as long as the number of students remains unchanged), so that piracy mostly displaces book sales without creating much additional demand. While I do not observe textbooks, I do observe some nonfiction books, most of which are either self help books, or educational in history or society. If demand for these works depends little on word-of-mouth advertising as well, one would expect the effect of piracy protection to be larger for fiction than for nonfiction works.

Table 8 shows the effect of piracy protection on fiction and on nonfiction titles. The estimation now includes interaction terms of piracy protection and indicator variables for the two types of works. Editions of nonfiction titles do benefit more from piracy protection than do those of fiction titles for most formats for most formats, although the difference is statistically insignificant at the 10 percent level throughout.

## 6 Robustness Checks

The results above are robust to small changes in the definition of success of protection. They are also robust to specifications that include interactions of time fixed effects with dummy variables for types of titles. In what follows, I present in more detail results that include more variables, and that have different interpretations of protection and its success.

## 6.1 Recently Published Editions and Title Age

The main analysis does not control for the age of the title and its editions because the titles have been published several years before they moved into piracy protection. However, many of those titles have been republished more recently. If an edition of a title becomes newly available, there are two possible effects on piracy and sales. First, a new edition can direct attention to the title and spur demand - an increase in regular sales is expected even if we also expect more piracy.

The second effect plays a role as I observe demand for different formats. Suppose Rosetta Books publishes a new e-book edition. This edition increases competition for the other formats (including pirated versions). As a result, we would expect a negative effect on both sales and piracy.

I collect the publication dates of all e-book editions (through Rosetta Books), as well as of the most recently published print and audio editions for each title (as observed in the Nielsen Bookscan database). In addition, I observe each title's original publication date. To determine if regular sales (and hence the inferred effect of piracy protection) are driven by recently published editions or by the title's age, I add three more indicator variables that are one if a new edition has been published in the two months prior to the month in question as well as a quadratic function of title age (in months) to the analysis in equation (1).

Table 9 shows that recent publications and title age do not significantly change the effect of piracy protection.<sup>23</sup> The coefficients on protection are close to those in tables 3 and 4. The effect of protection on e-book sales even increased and is now significant at the 1% level.

## 6.2 Site Traffic as a Measure of Success

Section 5.4 uses the change in the number of available infringing sites in order to identify the effects of de-listings and take down notices. It is possible though that those sites that are successfully taken down do not see a lot of traffic anyway. In that case, take down notices are not as successful as they originally appear.

I proxy for traffic to the infringing content by using site ranking information from Alexa Internet, Inc., as retrieved on September 25, 2014. This company provides traffic data, global

---

<sup>23</sup>Note that I could not find the initial publication date for five titles.

rankings and other information on millions of websites (the lowest observed ranking in my dataset is 18,943,656), as gathered from a sample of “millions of Internet users using one of over 25,000 different browser extensions.”<sup>24</sup> Note that the sites reported on Alexa.com are the mother sites to the infringing links, meaning that I do not observe traffic to the infringing content itself.

I use the inverse of the mother site’s ranking, assuming that traffic on the internet follows a Pareto distribution, and that traffic to the infringing content is proportional to traffic on the main site. Figure 3 shows that traffic to infringing sites is not perfectly correlated with the number of active infringing sites. I therefore estimate the two effects treating take down notices as successful if the estimated traffic to infringing content for a title decreases in a given time period.

Table 10 shows that the relative effects of de-listing a site and of a successful attempt to take down infringing sites are similar to those found when using the number of active infringing sites (table 5). The effect of de-listings appears to be stronger, although the difference remains insignificant at 10%.

### **6.3 Protection Only When Sites are Removed**

The above results assume that a title is constantly protected through Digimarc. That is, Digimarc constantly searches the Internet for infringing sites, and it constantly removes their links. However, the company searches the Internet more sporadically (once a week or once a month) for some of the titles. It is also possible that the company does not find some of the infringing links. A title would only be “protected” when the company actively and successfully tries to limit piracy.

Table 11 shows the estimation results when protection is only turned on during the time periods when an infringing site is taken down. This is a stricter definition of piracy protection, so that larger effects may be expected.

While the effect remains insignificant for most physical formats, it becomes significant (at 5%) for mass market paperbacks. This format constitutes the cheap, low quality paperback versions that can be found at airport bookstores. These are closer substitutes to free pirated versions than, for example, hardcover editions, and the significant positive effect is not surprising. Interestingly,

---

<sup>24</sup><http://www.alexa.com/about>, retrieved September 25, 2014.

the effect on e-books is smaller than it is under the assumption of constant protection, presumably because protection takes some time to see an effect.

## 7 Conclusion and Implications

Digitization has significantly lowered the cost of production and distribution of cultural goods over the past decade. This has led to large shifts in market structure and competitive behavior, leading to an increase in the variety of products that can be consumed. Some of these new products infringe the copyright of existing work. Economists have shown that file sharing and online piracy have had large impacts on other media industries, but a lack of data has made an analysis of the book publishing industry difficult in the past.

While piracy seems to hurt regular sales, it is less clear whether protection from piracy can be effective. The effort to protect a work from piracy can take on different forms. Most efforts have been public or centralized in nature, with mixed receptions and success. The book publishing industry showcases a more private effort to prevent piracy. The effect of piracy protection in this industry can indicate whether we can rely on private action to prevent copyright infringements.

This paper is the first to empirically analyze the interaction of online piracy and the legal market for books. It finds that piracy protection significantly increases regular unit sales of e-books, while the effect on physical formats is not as clear. Most of the effectiveness can be attributed to the removal of links to infringing sites from search engines, and the effectiveness of such protection is directly related to the effort that is exerted. It seems that this private protection can effectively increase sales. It remains to be seen whether such protection is efficient in the sense that the gains from protection outweigh the effort to protect a work.

The effect varies by the title's level of popularity. For well known books and those by popular authors, online piracy mainly poses a threat to regular book sales, while authors who are just starting out could benefit from the additional platform. My results support this idea, at least for e-books. The promotional effect of free online versions has already been assumed by several emerging authors who have started offering their titles, or excerpts of their titles, for free on their own websites. This strategy allows authors to control the level of sampling and promotion, while they have little to no control over online piracy.

My analysis is of course subject to limitations. I analyze a set of titles that were originally published several years ago and that are not necessarily the current top sellers. The dynamics that affect demand are different for recently published titles than for titles that have been available for a long time. They also differ across the titles' levels of popularity. The length of time it takes for protection to see an effect is especially large when looking at recently published works, indicating that piracy protection may be most effective for titles whose sales have already leveled off. Future research should therefore examine the effect of piracy on more recent titles.

Still, the analysis in this paper includes titles of varying levels of popularity and quality, and a variety of genres. It shows that well-known titles benefit more from piracy protection than titles that are more obscure, and that the effect depends on the work's genre. These patterns may also apply to new works by established authors as compared to new works by unknown authors. Overall, the results extend to other publishers, although the magnitude of the effect may vary depending on the work's and its editions' attributes.

As media industries continue to grow and transform in this digital age, copyright and piracy issues will continue to play a large role in governing the distribution of content. While private protection of a title seems to increase digital book sales, the optimal level of copyright protection depends on the ease of distribution, on the artists' popularity and objectives, and on the market structures of the legal and illegal markets.

## References

- [1] Bhattacharjee, Supid, Ram D. Gopal, Kaveepan Lertwachara, James R. Marsden. "Impact of Legal Threats on Online Music Sharing Activity: An Analysis of Music Industry Legal Actions." *Journal of Law and Economics* 45, no. 1 (2006).
- [2] Danaher, Brett, and Michael D. Smith. "Gone in 60 Seconds: The Impact of the Megaupload Shutdown on Movie Sales." Available at SSRN 2229349 (2013).
- [3] Danaher, Brett, Michael D. Smith, and Rahul Telang. "Piracy and Copyright Enforcement Mechanisms." No. w19150. National Bureau of Economic Research, 2013.

- [4] Danaher, Brett, Michael D. Smith, Rahul Telang, and Siwen Chen. “The effect of graduated response anti-piracy laws on music sales: evidence from an event study in France.” forthcoming, *Journal of Industrial Economics* (2014).
- [5] Liebowitz, Stan J. “Research Note-Testing File Sharing’s Impact on Music Album Sales in Cities.” *Management Science* 54, no. 4 (2008): 852-859.
- [6] Oberholzer-Gee, Felix, and Koleman Strumpf. “The effect of file sharing on record sales: An empirical analysis.” *Journal of Political Economy* 115, no. 1 (2007): 1-42.
- [7] Oestreicher-Singer, Gal, and Arun Sundararajan. “Are digital rights valuable? Theory and evidence from ebook pricing.” mimeo (2013).
- [8] Peitz, Martin, and Patrick Waelbroeck. “Why the music industry may gain from free downloading - The role of sampling.” *International Journal of Industrial Organization* 24, no. 5 (2006): 907-913.
- [9] Peitz, Martin, and Patrick Waelbroeck. “Piracy of digital products: A critical review of the theoretical literature.” *Information Economics and Policy* 18, no. 4 (2006): 449-476.
- [10] Petrin, Amil K. “Quantifying the Benefits of New Products: The Case of the Minivan.” *Journal of Political Economy* 110, no. 4 (2002): 705-729.
- [11] Peukert, Christian, Jörg Claussen, and Tobias Kretschmer. “Piracy and movie revenues: evidence from megaupload.” (2013).
- [12] Rob, Rafael, and Joel Waldfogel. “Piracy on the high C’s: Music downloading, sales displacement, and social welfare in a sample of college students.” Working paper no. 10874. National Bureau of Economic Research, 2004.
- [13] Sivan, Liron, Michael D. Smith, and Rahul Telang. “Do Search Engines Influence Media Piracy? Evidence from a Randomized Field Study.” Available at SSRN 2495591 (2014).
- [14] Smith, Michael D., and Rahul Telang. “Assessing the academic literature regarding the impact of media piracy on sales.” Available at SSRN 2132153 (2012).

- [15] Waldfogel, Joel. “Music file sharing and sales displacement in the iTunes era.” *Information Economics and Policy* 22, no. 4 (2010): 306-314.
- [16] Waldfogel, Joel, and Imke Reimers. “Storming the Gatekeepers: Digital Disintermediation in the Publishing Industry.” (2014).
- [17] Zentner, Alejandro. “Measuring the Effect of File Sharing on Music Purchases.” *Journal of Law and Economics* 49, no. 1 (2006): 63-90.
- [18] Zhang, Laurina. “Intellectual Property Strategy and the Long Tail: Evidence from the Recorded Music Industry.” (2013).

Table 1: Annual title sales, 2013

Sales	N*	Mean	Std. Dev.	Min	Max
Hardcover	60	383.862	1,527.05	0	10,773
Trade Paperback	88	5,743.53	21,974.79	0	167,570
Mass Market	56	831.576	1,970.38	0	11,654
Audio	43	560.767	3,183.42	0	20,923
E-Book	149	3,456.28	5,946.94	2	40,500

\*Titles that changed their piracy protection status.

Table 2: Online Piracy Prevention - Take down notices

Variable	N	Mean	Std. Dev.	Min	Max
Sites per title	251	87.899	183.025	1	1984
Notices per site	20,041	1.596	1.259	0	29
Success	20,041	0.835	0.371	0	1
Success after first notice	20,041	0.745	0.436	0	1
Time until success*	16,739	17.834	56.433	0	857

\*Sites that are successfully taken down

Table 3: Baseline Results: Piracy Protection

	(1)	(2)	(3)	(4)	(5)
VARIABLES	Hardcover	Trade PB	Mass Market	Audio	E-books
Protection	-0.0649 (0.338)	-0.167 (0.192)	0.103 (0.136)	0.342 (0.415)	0.197** (0.0760)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,652
R-squared	0.024	0.043	0.039	0.041	0.075
Number of Titles	60	88	59	43	150

Robust standard errors are clustered by title

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Table 4: Price Effect: E-books

VARIABLES	(1)	(2)	(3)
	Log price	Log sales	Log revenue
Protection	-0.0648 (0.0394)	0.108* (0.0550)	0.131** (0.0585)
Log price		-1.386*** (0.130)	
Time FE	monthly	monthly	monthly
Title FE	yes	yes	yes
Observations	3,652	3,652	3,652
R-squared	0.051	0.467	0.103
Number of Titles	150	150	150

Robust standard errors clustered by title

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 5: De-listings and take downs

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Hardcover	Trade PB	Mass Market	Audio	E-books
Under contract	-0.0688 (0.163)	-0.172 (0.180)	0.113 (0.137)	0.290 (0.420)	0.103* (0.0552)
Successful takedown	0.0775 (0.0833)	0.00880 (0.0836)	0.115 (0.107)	0.0632 (0.164)	0.0239 (0.468)
Active sites	0.00422 (0.0115)	-0.00205 (0.0105)	0.0166* (0.00910)	-0.0118 (0.0173)	0.00331 (0.00587)
Log price					-1.385*** (0.131)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,652
R-squared	0.027	0.043	0.040	0.041	0.47
Number of Titles	60	88	59	43	150

Robust standard errors are clustered by title

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 6: Piracy Protection - Priority Levels

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Low priority	-0.243 (0.254)	-1.006** (0.404)	-0.840 (0.963)	-0.00248 (0.505)	0.0120 (0.0838)
Medium priority	0.316 (0.633)	-0.299 (0.341)	-0.893 (0.777)	0.956* (0.508)	-0.0417 (0.0696)
High priority	-0.000427 (0.260)	-0.0787 (0.153)	0.397** (0.155)	-0.176 (0.397)	0.390*** (0.125)
Active sites	0.0276 (0.0227)	-0.00204 (0.0114)	0.0100 (0.00831)	-0.00737 (0.0163)	-0.00487 (0.00686)
Log price					-1.387*** (0.130)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,652
R-squared	0.028	0.048	0.052	0.046	0.47
Number of Titles	60	88	59	43	150

Robust standard errors are clustered by title

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 7: Displacement and Promotional Effects

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Hardcover	Trade PB	Mass Market	Audio	E-books
Protection*popular	0.111 (0.422)	-0.175 (0.213)	0.391* (0.220)	0.500 (0.458)	0.157* (0.0865)
Protection*obscure*“good”	0.0140 (0.283)	-0.0752 (0.300)	0.797 (0.143)	-0.0539 (0.466)	0.0828 (0.0796)
Protection*obscure*“bad”	-0.620 (0.403)	-0.215 (0.951)	-2.046** (1.005)	-1.053 (1.168)	0.0758 (0.0776)
Active sites	0.0258 (0.0240)	-0.00200 (0.0105)	0.00990 (0.00877)	-0.0147 (0.0171)	0.00245 (0.00593)
Log price					-1.410*** (0.136)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,593
R-squared	0.028	0.043	0.059	0.045	0.48
Number of Titles	60	88	59	43	146

Robust standard errors are clustered by title

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 8: Fiction and Nonfiction Titles

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Hardcover	Trade PB	Mass Market	Audio	E-books
Protection*fiction	-0.352 (0.282)	-0.262 (0.248)	0.0459 (0.137)	0.373 (0.459)	0.105* (0.060)
Protection*nonfiction	0.583 (0.750)	0.0368 (0.224)	0.558* (0.296)	0.256 (0.942)	0.125 (0.0942)
Log price					-1.386*** (0.130)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,652
R-squared	0.027	0.044	0.040	0.041	0.47
Number of Titles	60	88	59	43	150

Robust standard errors are clustered by title

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 9: Recent Publications and Title Age

VARIABLES	(1) Hardcover	(2) Paperback	(3) Mass Market	(4) Audio	(5) E-books
Protection	0.0926 (0.347)	-0.145 (0.192)	0.156 (0.142)	0.303 (0.399)	0.165*** (0.0552)
Recent e-book	0.113 (0.290)	0.220 (0.141)	-0.0603 (0.130)	0.221 (0.362)	0.289*** (0.0742)
Recent print	1.853* (1.036)	-0.652 (0.471)	0.464 (0.355)	-0.799 (0.575)	0.171 (0.130)
Recent audio	0.0330 (0.280)	0.451 (0.369)	0.0284 (0.250)	2.611*** (0.512)	0.149 (0.141)
Title Age	-0.0114 (0.0286)	-0.0407** (0.0164)	-0.0580** (0.0263)	-0.0141 (0.0377)	-0.00896 (0.00628)
(Title Age) <sup>2</sup>	0.0000484*** (0.0000173)	0.0000148 (0.0000118)	0.0000538** (0.0000215)	0.0000134 (0.0000216)	0.0000223 (6.49 <sup>-6</sup> )
Log price					-1.448*** (0.132)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,588
R-squared	0.033	0.046	0.047	0.058	0.48
Number of Titles	60	88	59	43	145

Robust standard errors are clustered by title

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 10: De-listings and take downs: using traffic information

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Under contract	-0.0560 (0.342)	-0.157 (0.182)	0.154 (0.141)	0.300 (0.414)	0.101* (0.0546)
Successful takedown	0.0757 (0.128)	0.0101 (0.0759)	0.00409 (0.0931)	0.00912 (0.164)	0.0459 (0.0336)
Traffic on active sites	0.210 (0.433)	0.110 (0.175)	0.494** (0.237)	-0.319 (0.416)	0.143 (0.0987)
Log price					-1.385*** (0.130)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,652
R-squared	0.025	0.044	0.043	0.042	0.47
Number of Titles	60	88	59	43	150

Robust standard errors are clustered by title

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Table 11: "Sporadic" Piracy Protection

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Protection	0.108 (0.187)	0.0125 (0.0795)	0.203** (0.0885)	-0.0300 (0.196)	0.144** (0.0560)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,652
R-squared	0.024	0.043	0.040	0.040	0.47
Number of Titles	60	88	59	43	150

Robust standard errors are clustered by title

\*\*\* p&lt;0.01, \*\* p&lt;0.05, \* p&lt;0.1

Figure 1: Active piracy sites - I Am Legend



Figure 2: Protection and e-Book Sales

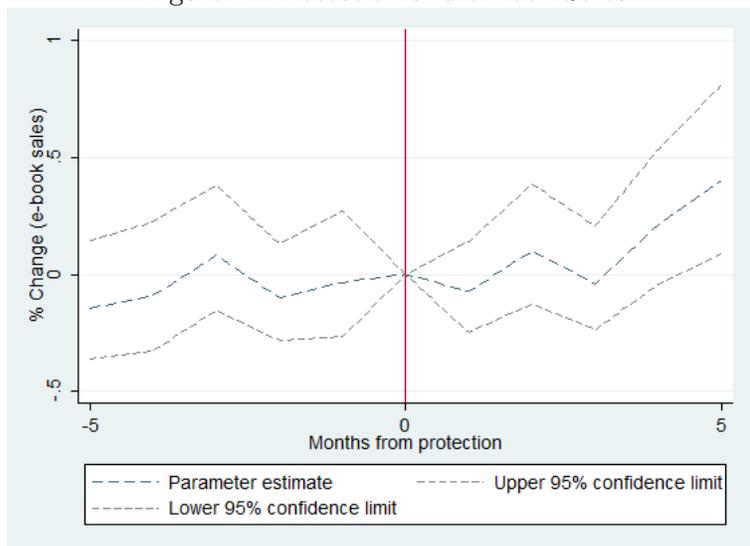


Figure 3: Active piracy sites and estimated traffic - I Am Legend

