

# Can Private Copyright Protection be Effective? Evidence from Book Publishing

Imke Reimers\*

March 8, 2016

## Abstract

Digitization has impacted publishing, news, and entertainment industries in recent years by lowering the cost of access. With the option to download creative works legally, however, comes the possibility of doing so illegally. Public (legislative) and formal (legal) efforts to prevent copyright infringements have been controversial or inefficient. The book industry showcases an alternative approach in which private companies use relatively inexpensive network surveillance to protect individual titles. I estimate the effectiveness of such protection on legal sales of books that become protected using a difference-in-differences approach. I find a protection-related increase of e-book sales - the closest substitute for online piracy - of more than 14 percent, with effectiveness depending on popularity, genre, and search frequency. Most of the increase is due to prevention of “casual” infringements, rather than professional piracy. Sales of non e-book formats are not affected.

---

\*Department of Economics, Northeastern University, Boston, MA 02115, i.reimers@neu.edu. I am grateful to James Dana, John Kwoka, Gerard McCullough, and Joel Waldfogel for useful feedback and advice. I would like to thank seminar participants at Northeastern University, the University of Groningen, the ZEW Conference on Information and Communication Technologies, the University of Minnesota, the EC Workshop on the Interaction between Legal and Illegal Book Sales, the University of Zurich, and the NBER productivity lunch for helpful suggestions which improved the paper. Finally, I thank the publisher and the anti piracy company for generously providing data to support this research.

# 1 Introduction

Digitization has significantly decreased the costs of creating and distributing cultural goods in media industries. The collapse of traditional costs has increased the quantity and variety of products available to consumers. The wider variety is expected to increase the level of competition and lower prices further. Consequently, it is affecting firm and artist profitability. By themselves, these changes shift surplus from producers to consumers.

Digitization has had another effect though. As the *legal* distribution of creative works has become cheaper and easier, so has the *illegal* distribution of these works. Recent academic work shows that illegal distribution displaces legal sales in media industries, even though modern technology has made monitoring and regulation of illegal activity more feasible. Thus, the question of the “best” intellectual property (IP) strategy becomes extremely important for each artist and distributor, and in each industry. To what extent can creative works be effectively protected from piracy, what should this protection look like, and who should enforce this protection?

Many well-known efforts to protect copyright have been public, or legislative, in nature, with mixed success and reception.<sup>1</sup> Other efforts have been legal in nature, often attempting to protect the entire industry at once - an effort that can be very costly. A less formal type of protection, private copyright protection that targets individual works, has received less attention but might have clear advantages. Most importantly, the ability to target a subset of works could make the effort to prevent infringement more efficient.

The book publishing industry provides a setting that allows me to analyze such an effort to protect a work’s copyright. In the book publishing industry, targeted copyright protection efforts are led by private companies that are hired by publishers to search the Internet for infringing content specific to protected titles. Protection from piracy is title and publisher specific, rather than industry wide. It is initiated by publishers and authors, rather than by public officials, and the firms attempt to remove infringing content without the involvement of courts, which further lowers the costs of protection.

I estimate the effect of such a private anti-piracy effort on legal book sales. 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 January 2010 to December 2013. The dataset includes the intensity level and success rate of piracy protection over the same time

---

<sup>1</sup>These efforts are described briefly below.

period, through a large piracy protection company. I follow sales of each title in the publisher's catalog before and after their title-specific piracy protection begins in a difference-in-differences setting. I check whether (and by how much) sales rise or fall when protection for a title begins, relative to other titles which become protected at another time.

In this analysis, I take advantage of the fact that setting up protection of a work is a relatively labor intensive process that was implemented in shifts over several months by the anti-piracy company, and that the publisher seemed unaware that the order in which titles were submitted might affect their profits. I further use Google Trends data to provide evidence that the timing of protection is exogenous and consistent. Here, I take advantage of the fact that Internet search volume likely reflects consumer interest in a work, whereas it would not be affected by the work's copyright protection status.

A large and growing literature addresses the effect of file sharing in media industries.<sup>2</sup> While some early work indicating that there is no significant effect (Oberholzer-Gee and Strumpf, 2007), most recent work has found that regular sales have been significantly displaced by pirated versions (Zentner, 2006; Rob and Waldfogel, 2006; Liebowitz, 2008; Waldfogel, 2010). In the music industry, the negative effect of file sharing on legal sales 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 four years later, when Apple sold its first songs through the iTunes Store in May 2003.

More recently, research on file sharing has turned to determining whether anti-piracy efforts can be effective, taking as given that piracy likely harms sales in an industry. Most efforts to protect creative works have been legislative, and many have been controversial. Consider, for example, the Internet blackout to campaign against the Stop Online Piracy Act (SOPA) and the Protect Intellectual Property Act (PIPA) in January 2012.<sup>3</sup> Similarly, the French Parliament passed the Creation and Internet Law, an anti-piracy law 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

---

<sup>2</sup>See Smith and Telang (2012) for a detailed description of studies on those industries, and Peitz and Waelbroeck (2006a) for a review of the theoretical literature.

<sup>3</sup>The acts were aimed to facilitate the combat of online copyright infringement by (among others) barring advertising networks from conducting business with infringing sites, web search engines from linking those sites, and requiring Internet service providers to block access to them. See <https://www.eff.org/issues/coica-internet-censorship-and-copyright-bill>.

access to the Internet a basic human right, and the bill was finally revoked in July 2013.<sup>4</sup>

Most legal efforts to prevent piracy tend to be broad in scope 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, but the shutdown did not have a positive effect on box office revenues of less well-known works (Peukert et al., 2013; Danaher and Smith, 2014). In music, the Recording Industry Association of America (RIAA) made 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, 2014).<sup>5</sup>

Legal efforts to prevent piracy face less *controversy* than legislative efforts, but it is not clear if the gains from such protection outweigh the effort that is exerted to protect a work, especially when protection is broad. I address these questions by examining the effectiveness of a less formal form of piracy protection - one in which private firms offer surveillance of piracy activity for specific titles - by comparing the resulting increases in revenue with the costs of protection. My work thus lends support to Luo and Mortimer (2015), who find that private enforcement in the form of ex-post infringement settlement requests of stock photography can indeed be successful. I show that these attempts can lead to increases in revenue that outweigh the cost of protection, especially when legal action can be avoided.

Specifically, I find that the effect of piracy protection on legal book sales depends on the popularity of the title, the type of work, 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 protection-related increase in sales of at least 14%. My results suggest that the effect is driven more by the de-listing of links from search engines than by a decrease in piracy sites offering the title. This indicates that casual pirates are more easily deterred than more serious pirates. Moreover, this type of piracy protection increases e-book revenue by up to ten times the cost of protection. I find that private and targeted piracy protection can be effective, and if the “right” set of works is protected with an appropriate level of intensity, this protection can even be cost efficient.

The remainder of the paper proceeds as follows. Section 2 provides background on the publishing

---

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

<sup>5</sup>DRM is a technology which restricts a consumer’s ability to copy (and thus redistribute) the content they have purchased.

industry and on the role of piracy. I describe the data in detail in section 3, and I describe the empirical strategy of the experiment and provide evidence that the effect is identified in section 4. I proceed with showing results about the the effectiveness and efficiency of protection in section 5, and I examine the mechanism with which protection works in section 6. I conclude with policy implications in section 7.

## 2 Industry Background and Piracy Protection

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>6</sup> Some of these challenges can be attributed to increased competition through free online versions.

The effects of file sharing and the effort to stop it are not yet well-studied in book publishing, as digitization in this industry has lagged behind other industries.<sup>7</sup> While e-books can be read on computers and have therefore been available for over a decade, they are most useful to consumers when read on small hand-held devices, such as e-book readers or tablet computers. Sony had released 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 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 USA Today Top 150 bestseller books have been sold primarily as e-books by late 2012 (see Waldfogel and Reimers, 2015).

Online piracy has been relatively concentrated in other content industries. Although many book titles are available on well-known piracy sites (for example, peer-to-peer networks such as Piratebay), the book industry does not face one large file sharing service, as was the case with

---

<sup>6</sup>See <http://stateofthemediamedia.org/2013/newspapers-stabilizing-but-still-threatened/newspapers-by-the-numbers/purchases-by-volume/>.

<sup>7</sup>Hardy et al. (2014) conduct a field experiment in which they limit piracy for a set of more recent book titles in Poland. While they find that piracy can in fact be limited, they do not find an effect of such action on legal sales. My paper looks at a different set of titles and a slightly more comprehensive protection strategy. While their strategy consists only of take down notices, the strategy here also includes requests to de-list links from search engines.

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

Napster in the music industry.<sup>9</sup> The small size of book files as compared to music and movies makes it convenient to store them as intact files, so that much of the infringing content is located on websites rather than on peer-to-peer networks. This in turn makes tracking and taking down illegal content more feasible.

The small file size for books may also lead to fewer “professional” pirates. Most book content (65 percent) is either in html or pdf format - formats that are easy to upload and download but can take away from the reading experience. Only twenty percent of the infringing content in my dataset is clearly recognizable as e-book formats, perhaps because these formats have additional copy protection (for example, digital rights management).<sup>10</sup>

In the book publishing industry, copyright protection involves two major steps: requesting that search engines de-list the link to the infringing content, and demanding that the pirated content be taken down directly through the site host. The two actions target different types of pirates. Consumers who are deterred by changes to search results are likely “casual” pirates, while more serious pirates are less likely to rely on search engines to find content. Sivan et al. (2014) show that the presence of piracy links in search results strongly affects the piracy behavior of consumers in the movie industry.

This paper follows titles that become protected by Digimarc, one of the industry’s largest anti-piracy services.<sup>11</sup> Its piracy protection strategy includes an automated crawling process to find suspected pirated content. The sites with confirmed pirated content are subject to two types of treatment: the company 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 prevents “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 - those who know where to find pirated content anyway. Protection continues as long as the title is under contract.

Digimarc has added most of the major U.S. publishing houses as clients over the past few years. The publishing companies include HarperCollins (February 2010), Macmillan (October 2010,

---

<sup>9</sup>Peer-to-peer networks are created when two or more computers are connected and share resources without going through a separate server computer.

<sup>10</sup>It is not obvious why pirates offer infringing content for books as offering this content is not typically monetized. Many suppliers of pirated content list fun, a sense community, and fame as the main motivators for offering such content. See <http://www.makeuseof.com/tag/motivates-people-record-upload-pirated-movies-music/>.

<sup>11</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.

although Macmillan now uses another service), Simon & Schuster (October 2011), and Random House (June 2013, now Penguin-Random House).<sup>12</sup> In addition, it has added several smaller publishers.<sup>13</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>14</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. While most of Rosetta’s titles were originally published more than a decade ago, the publisher also carries some original titles that are available only in electronic format.

Protection for Rosetta’s full catalog was due to begin in June 2011. Although the entire catalog began protection on the same date, the actual setup happened in shifts, and the timing was not coordinated with the publisher. When asked why titles became protected in shifts, an employee of the anti-piracy company relayed that the timing with which titles were protected was related to capacity constraints, rather than strategic decisions.<sup>15</sup> In addition, according to this account, the publisher was not aware of any capacity constraints at the time.<sup>16</sup>

### 3 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>17</sup> Most of the works were originally published several years ago,

---

<sup>12</sup>[http://www.mediabistro.com/galleycat/harpercollins-and-harvard-business-school-publishing-will-use-attribute-piracy-protection-program\\_b11178](http://www.mediabistro.com/galleycat/harpercollins-and-harvard-business-school-publishing-will-use-attribute-piracy-protection-program_b11178), <http://www.marketwired.com/press-release/Attributor-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-attribute.html>, and <http://randomnotes.randomhouse.com/anti-piracy-reporting-tool-added-to-author-portal/>.

<sup>13</sup>See <http://www.digimarc.com/guardian/customers> for a (non-comprehensive) list of Digimarc’s customers.

<sup>14</sup>While electronic editions can only be published by Rosetta, physical editions are published by other publishers. See <http://go-to-hellman.blogspot.com/2011/11/creative-commons-media-neutrality-and.html> for more information.

<sup>15</sup>In an e-mail correspondence with the author, July 28, 2015, the employee stated: “I am afraid we do not have concrete reasons why titles were protected in shifts. [...] It took some doing before we had a regular ONIX feed set up and it appears that the global catalog was added in shifts, where large blocks were ingested in batches rather than the entire catalog being added in one shot. I think we actually got a flat file or two before the full feed, and then deltas following the initial large feed with anything that wasn’t on the first list. Why it was broken down that way isn’t something we have much explanation for, apart from saying this is how it was supplied to us, and it took a while before it was complete.”

<sup>16</sup>The author is especially grateful to the recommendations of an anonymous referee who suggested a number of econometric tests to establish exogeneity of the timing beyond this email. These tests and the affirmative results are described below in Sections 4 and 5.

<sup>17</sup>See <http://www.rosettabooks.com/books/books-a-z/> for Rosetta’s catalog.

going back as far as the first half of the twentieth century, although all titles are still protected by copyright. There also are a few titles that have been published in the past fifteen years. The analysis includes the subset of 251 titles for which pirated content was found 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. While these sales data date back to 2002, I focus on the time period from January 2010 to December 2014. These data 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. On average, editions sell close to 3,000 copies in a year, although this varies by format, with trade paperbacks and e-books as the most popular formats.

On the piracy dimension, I obtain detailed information through the anti-piracy company. 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 24 different protection start dates between August 2011 and November 2013. Titles are searched with three different frequency levels, depending on the amount of piracy that is found. The search frequency of a title can vary over time, although it does not do so often.

Table 1 shows summary statistics on the number of sites found per title, the number of notices sent per site, and measures of the company's success, for all titles that were protected at some or all points during the observed time period. The anti-piracy service appears to have some success in reducing file-sharing, and this success comes reasonably quickly. Over 50% (10,417) of all infringing sites are taken down within one week of being found. Yet, it is possible that not all websites are found, and that taking down one website will simply lead to another one showing up to offer the same title.<sup>18</sup> The fact that the company finds close to 88 sites per title on average, and that it successfully takes down over 83% of those, indicates that the company affects the online book piracy landscape, at least for the titles that Rosetta Books offers. The relatively high success rate may be

---

<sup>18</sup>See Aguiar et al. (2015) and Danaher et al. (2015) for studies on the emergence of new piracy sites when one is shut down.



due to the nature of piracy in book publishing. Because book files are small, there are more small sites with infringing material, and the operators of such sites may be more easily intimidated by the take down notices.<sup>19</sup> Moreover, since offering pirated content is not usually monetized, pirates have no direct incentives to continue offering the content.

The number of detected piracy sites that have not been taken down increases on average over time because not all take down notices are successful. Figure 1 illustrates the cumulative number of sites that are detected, 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>20</sup> The number of active sites (dotted line) decreases at times. For example, while there is a spike in piracy in late December 2011 with up to 36 infringing sites (December 21), some of these have been successfully taken down by January 2012, with only 23 active sites by January 1. In early 2014, though, fifty infringing sites can be found. This variation allows 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 the anti-piracy service 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, although the de-listing typically happens within hours.

I use two more sources to collect additional information on each title's interest level (popularity) and quality. To measure each title's popularity at any given time, independent of its piracy level, I collect the Google search history for all titles and authors from January 2011 to December 2013 through Google Trends. While Google Trends does not provide absolute counts of search queries, it allows me to compare the relative interest levels across time and across search queries.<sup>21</sup> The information on title "quality" is collected through the crowd-sourced online book database Goodreads. This website provides over 10 million book reviews across 700,000 titles.<sup>22</sup> The mean (median) number of reviews for the titles in my analysis is 4,855 (483) for ebooks, and 20,821 (1,500) for

---

<sup>19</sup>Overall, 69.8% of infringing content is found on cyberlockers (Internet hosting services specifically designed to host user files), most of which in html and pdf formats, while only 18% were operated through peer-to-peer networks. Peer-to-peer networks are notoriously difficult to contain, with a success rate of only 65%, compared to 87% among cyberlockers.

<sup>20</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. I instead estimate the effect of protection itself.

<sup>21</sup>I use the variation on those dimensions to support my claims of exogeneity of protection in section 5.4.

<sup>22</sup>See <http://www.goodreads.com/api>.

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.

Finally, I use information on site traffic from Alexa Internet, Inc. to approximate the actual level of piracy in robustness checks.

## 4 Empirical Strategy and Identification

In an ideal experiment that determines the effect of piracy protection, the econometrician assigns titles randomly across groups and protects one set of titles. I observe a quasi experiment in which the *timing* of protection adoption across titles is uncorrelated with the returns to protection. Although protection for all of the publisher’s titles was contracted for at the same time, the order in which the protection was applied was determined by when each title was submitted, and the order of submission was independent of title characteristics.<sup>23</sup> The marginal effect of online piracy protection on book sales is then the change in sales of the treated titles as compared to the change in sales of the group of titles that was not yet treated. The empirical strategy takes advantage of the change in protection in a difference-in-differences analysis. The log of the (observed) unit sales  $q_{it}$  of title  $i$  in time period  $t$  is a function of the work’s observable characteristics and its protection status in that time period.

The effect of piracy protection is identified because I observe sales of titles when they are protected and when they are not protected. Since adoption of 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 directly affect legal demand for another title (although it might affect the level of *piracy* for other titles). I estimate the effect of piracy protection as

$$\log(q_{it}^k) = X_{it}'\beta + \alpha \cdot \text{protection}_{it} + \delta_i^k + \mu_t^k + \epsilon_{it}, \quad (1)$$

where  $q_{it}^k$  denotes the unit sales of title  $i$  in week  $t$  in format  $k$ , with  $k \in \{\text{hardcover, trade paperback, mass market paperback, audio, and e-books}\}$ ,  $X_{it}$  includes time-varying title characteristics including recent editions, title age, average sales prices, and the title’s search volume on Google

---

<sup>23</sup>I provide evidence of this independence in Section 5.4.

(the reader interest for the title).<sup>24</sup> Finally,  $protection_{it}$  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>25</sup> The log specification assumes that piracy changes sales by a common percentage, rather than by a common absolute amount. This controls for differences in absolute sales levels across titles.

I also make use of the fact that I observe unit sales of over 120 titles (depending on format), over an extended time period (170 weeks for physical books, or 30 months for e-books), by controlling for title  $i$  and format  $k$  ( $\delta_i^k$ ), and by adding time 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 that of the genre and the author. 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.

In further specifications, I use both the overall protection and a measure for the effectiveness of take down notices as my independent variables of interest to identify the leading mechanism behind the success of piracy protection. I also estimate the role of intensity of protection: the effect of different frequencies of search for infringing content. I then 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 benefit from protection.

## 5 Results

I follow unit sales for different formats of the titles from January 2010 to January 2014 (physical sales), and from July 2011 to December 2013 (e-book sales). I estimate several versions of equation (1) for different formats and sets of titles.

### 5.1 Baseline: Piracy Protection

Table 2 shows the baseline estimation results for all titles that moved into protection at some point during the observed time period, with at least two months on and off protection, respectively. The

---

<sup>24</sup>I control for reader interest instead of using an instrumental variables approach in which I would estimate a title's protection status as a function of reader interest in the first stage because reader interest is not a good predictor of a title's protection status.

<sup>25</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. The bias understates the true effect but goes to zero quickly (the bias is  $\frac{\Delta q}{q(q+0.0001)}$ ). The estimation of e-book sales does not require this adjustment as I do not observe time periods with zero sales.

independent variable of interest is an indicator variable that is one if the title is under contract with the anti-piracy company. I treat a title as protected in 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 varies across formats because some titles are not available in all formats.

The table shows that the effect of piracy protection on legal book sales depends on the edition’s format. 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 10%) on sales of e-books. Moving a title into piracy protection increases regular e-book sales by 15.5% ( $= e^{0.144} - 1$ ).<sup>26</sup> Physical editions of a title, on the other hand, 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 (including audio books), and even negative (but highly insignificant) for hardcovers and trade paperbacks. More generally, the effect on physical formats is estimated very imprecisely, which leads me to focus much of my analysis on e-books.

## 5.2 Additional Controls

Of course, Table 2 ignores several factors which may have contributed to an increase in e-book sales around the time that a title becomes protected. It could be that a move into protection coincided with the publication of a new edition of that title, with title and age-related demand fluctuations, or with price promotions.

A new edition of a title can direct attention to the title and spur demand. An increase in regular sales is expected even if we also expect more piracy. But if the publisher publishes a new hardcover edition, this edition increases competition for the other formats. This would have a negative effect on sales of those formats. 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 1 if a new electronic, print, or audio 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 equation (1). I further account for title-specific demand fluctuations by controlling for the title’s and author’s Google Trends search volume in the estimation, as well as for the total

---

<sup>26</sup>The remainder of the paper reports the extrapolated percent change in sales, while the tables report the estimated coefficients. Note that I cluster standard errors by title to account for common group effects and to reduce the potential for overstating significance due to serial correlation (see Bertrand et al. (2004)). The number of titles is “large”, so that any non-normality of the errors is not a big concern (see Donald and Lang (2007)). Note also that this positive effect of protection is robust to different definitions of protection. See appendix section A.1.

number of active detected piracy sites.

Finally, the publisher can adjust a book's price. Oestreicher-Singer and Sundararajan (2010) show that digital rights management has an economically significant effect on pricing decisions of digital content, and it is possible that piracy protection does as well. Since preventing piracy effectively limits the level of competition in the market, theory predicts a resulting increase in prices. If this is the case, 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>27</sup> An unreported difference-in-differences analysis with e-book prices as the dependent variable indicates that e-book prices decreased by 5.3% as a result of a move into piracy protection. This effect is significant at the 10% level. A negative correlation between prices and protection would indicate that the results above overestimate the effect of piracy protection. I therefore control for prices in the analysis as well.

Table 3 shows that a recently published electronic edition increases e-book demand for a title, regardless of the specification. The effect of piracy protection on e-book sales remains similar in size and significance as I add more control variables. Even when the average sales price is controlled for (column 3), the effect of piracy protection remains similarly strong, and significant at the 5% level, leading to a 15.9% increase in unit sales. Column 3 also shows that consumers are relatively price elastic with a price elasticity of -1.48. I also estimate the impact of protection on e-book revenue, and I find that e-book revenue increases by 17% when a title becomes protected, as shown in column 4.

The remaining coefficients are sensible as well. One would expect e-book sales to increase when a new e-book edition is published, while a new print edition (a competing format) could move demand away from the electronic edition. Audio editions, on the other hand, could serve as complements to e-books.<sup>28</sup> Moreover, when a title's Google search volume is higher, it tends to sell more units. The negative coefficient on the number of active piracy sites confirms that these sites provide substitutes to the legal e-book sales. After adding year-month fixed effects, the fact that

---

<sup>27</sup>A decrease in prices as the level of competition decreases is 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. Alternatively, piracy protection increases the publisher's awareness of the zero-price competitor in the first place.

<sup>28</sup>For example, one might listen to an audio book in a car, and continue to read the e-book version at home.

the titles in the dataset become older over time does not seem to affect sales much. The positive coefficient on the squared title age suggests that more recent titles are more likely to lose demand in the time period of the analysis, while demand for older titles has leveled off even more by 2011.

### 5.3 Costs, Benefits, and Efficiency

Private copyright protection can indeed be effective, at least for the e-books in my dataset. Even more, piracy protection can be *efficient* in the sense that the benefits to the publisher, authors and protection service outweigh their costs. In 2010, Rosetta's total revenue was over \$1.5 million across over 300 books.<sup>29</sup> Since then, the number has likely increased significantly as e-books have become more popular and Rosetta has substantially increased its catalog. Taking the annual revenue of \$1.5 million as given, Rosetta's revenue would increase by \$255,000 if all works benefited equally from piracy protection. Of course, pirated content has been found for only 228 titles (out of 673), so that Rosetta's total revenue likely increased by less than that. Since the reported revenue of \$1.5 million corresponds to a catalog of approximately 300 titles, one might assume that the incremental revenue to the 228 titles for which piracy was found adds up to \$193,800 ( $= 255,000 \times 228/300$ ).<sup>30</sup>

This benefit needs to be compared to the costs of protection. To the publisher, this cost is the fee it pays to the anti-piracy company - a simple transfer of wealth. To the anti-piracy company, the marginal cost of protecting the publisher consists of several dimensions in addition to the fixed costs of running the company. The publisher-specific marginal costs of protection include the development of scraping scripts, manual discovery of infringing content (emulating the behavior of actual pirates), approval of the links that have been identified as infringing (for example, to negotiate fair use situations), and enforcement (sending notices, monitoring, legal escalation, and so on). I can make inferences on these publisher-specific costs as Rosetta's titles account for 0.26 percent of all links that Digimarc has found in the previous year. These 0.26 percent would amount to 823 hours of incremental operational work to protect this publisher.<sup>31</sup>

The revenue to the publisher and authors (as royalties) increases by approximately \$194,000. At an hourly wage at the US average of around \$25, the incremental cost to the anti-piracy company of protecting the publisher would amount to \$20,575 - not much more than 10% of the increase

---

<sup>29</sup>See <http://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/47933-rosetta-books-comes-of-age.html>.

<sup>30</sup>This is a conservative estimate as the titles with piracy are likely more popular than those that are not pirated.

<sup>31</sup>Of course, this measure does not include the overhead and fixed costs of operating the anti-piracy company. The true amount of time needed to protect the publisher would be larger.

in revenue.<sup>32</sup> At the minimum wage, this cost would of course be even lower. These calculations suggest that the benefits to the publisher and authors outweigh the marginal costs to the anti-piracy company by a large margin. Whether these benefits also outweigh the fixed (overhead) costs of running the anti-piracy company and the potential losses in consumer surplus due to the removal of the zero-price pirated option is beyond the scope of the available data, although it is reasonable to assume that the overhead costs associated with a surveillance operation of this sort are small, especially when they are distributed over several publisher clients.

#### 5.4 Evidence Supporting the Exogeneity of Protection

The identification strategy assumes that the timing of piracy protection is uncorrelated with factors that determine the outcome of interest - in this case, log-unit sales. Of course, a publisher may be more interested in protecting those titles that have gained popularity lately than in protecting those that do not 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.

This paper deals with this problem in several ways. First, in addition to controlling for a large list of demand shifters, I use a list of titles that are "old" - most of them were published before 2000 - so that it is unlikely that a change in demand is caused by title specific market shocks. Second, conversations with both the publisher and the anti-piracy company suggest that the timing of protection was not determined by the publisher. Third, I use Google Trends data to provide two types of evidence that the timing of protection is not correlated with the variation in title interest across titles or within each title over time, implicitly assuming that a title's search volume is indicative of demand shocks but is not affected by piracy protection. I find no correlation across titles between the overall search volume and the timing of protection. I also find no correlation across time in the search volume and the time of protection.

To test whether the decision to have a title protected early is a result of the title's overall popularity, I plot the titles' mean monthly popularity over the observed time frame against the month in which the title's protection began. Figure 2 shows that those titles that were protected earliest (between September and December 2011) had a mean search volume over the observed

---

<sup>32</sup>See <http://www.bls.gov/news.release/empsit.t19.htm> for information on mean hourly wages in the private sector.

time period that was not higher than that of titles whose protection status was changed later. In addition, the unconditional correlation between the titles' average monthly search volume and the month in which they become protected is positive and close to zero at 0.05. This suggests that the publisher did not strategically move those titles into protection first for which the publisher expected to extract the highest revenue.

To test whether the decision to protect a title is driven by sudden spikes in the interest for a title, I regress each title's search volume in a given month on a quadratic function of the time (in months) before and after protection starts. One would expect the coefficient on the linear term to be negative simply because reader interest likely decreases over time. The coefficient on the quadratic term, on the other hand, provides information on the reader interest level around the start of copyright protection. If reader interest for a title spikes around the start date of protection, one would expect the coefficient on the quadratic term to be negative, but if there is no spike in reader interest, the coefficient will be close to zero or even positive.

Table 4 shows that the coefficient of interest is indeed close to zero, or even positive, suggesting that reader interest is higher in those months that are farther removed from protection. This effect is not a result of recent editions being published or title-specific trends as the regression includes title and month fixed effects and indicators that are one if an edition of the title has recently been published. The pattern holds for title search volume (columns 1 through 3) as well as for author search volume (column 4), and it holds when limiting the range of months to six months before and after the start date (column 1). There is no evidence of spikes in reader interest around the start date of protection.

## 6 The Process of Protection

The above results establish that the informal, targeted type of copyright protection that is commonly used in the book publishing industry can increase legal sales by more than the cost of protection. In what follows, I examine the mechanisms which make this protection effective.

### 6.1 Measuring the Speed of the Impact of Protection

To analyze the promptness with which piracy protection takes effect, I employ a regression discontinuity strategy in which I use the time (in months) before and after protection starts (around the



cutoff) as the explanatory variable of interest as in equation (2):

$$\log(q_{it}) = X'_{it}\beta + \left( \sum_{m \in M} \alpha_m \cdot \mathbb{1}[t - \textit{protection\_start} = m] \right) + \delta_i^k + f(t) + \epsilon_{it}, \quad (2)$$

where *protection\_start* is the first month of protection,  $X_{it}$  includes the title’s age, recent editions, title and author search volume, and the average sales price,  $\delta_i^k$  are title-format indicators, and  $f(t)$  represents a linear time trend.

Figure 3 shows that piracy protection takes around five months to have an effect, indicating that it takes quite some time to find and deactivate “enough” infringing content. This also suggests that the age and popularity of a title plays an important role when determining the effectiveness of protection. Piracy is likely to re-emerge more slowly for older titles, so the anti-piracy efforts have a chance to “catch up” with pirates. Piracy protection is most effective for those titles whose demand has leveled off, and which are less likely to encourage new illegal content to emerge.

Figure 3 provides additional evidence of the exogeneity of piracy protection. If piracy protection were endogenous, demand for the title would likely be unusually large in the time periods leading up to piracy protection. This does not appear to be the case for the e-books (or other formats) in my data set as the coefficient on log-unit sales leading up to protection is close to zero, and does not vary significantly before protection starts.<sup>33</sup>

## 6.2 Separating the Effects of De-listings and Take Downs

The above estimates show the combined effect of two actions: de-listing links from search engines and taking down sites with infringing content. In order to know which of these two actions makes piracy protection effective, 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 would indicate that more piracy sites emerged than were removed. In that case, the only “successful” component of protection is the de-listing of links from search engines. I further control for the overall piracy “supply” by including the number of detected sites

---

<sup>33</sup>To alleviate remaining concerns, I repeat this analysis for titles which were originally published more than 20 years ago - titles for which a sudden spike in demand is less likely - and obtain almost identical results. I also estimate the regression with year-month indicators instead of a linear time trend, and the results remain unchanged. In addition, I add linear pre- and post-protection trends, and I add title-specific linear time trends in other regressions. The effect of protection remains robust in direction, magnitude, and statistical significance. Finally, I run a Placebo test in which I artificially set the protection start date several months before the actual start date and show that the effect of “protection” disappears in appendix section A.2.

that have not yet been taken down for each title in addition to the controls used in section 5.2.

The relative effect of take down notices and de-listings can be inferred from the relative sizes and significance of the two coefficients of interest. When a title is under contract but more sites emerge than are taken down, any effect we see must be due to the de-listing of links. If more sites are successfully removed than emerge, the estimated effect is a combination of both the de-listings and the take downs. Table 5 reports the effects of each action: being under contract (link de-listings), and successful take downs. It suggests that de-listings have a stronger effect on e-book sales than take downs do. The coefficient on de-listings remains almost unchanged compared to the results in Table 3, increasing e-book sales by 15.8%. On the other hand, a successful attempt to take down infringing sites (a decrease in the number of sites for a title) leads to a small and imprecisely estimated additional increase (1.4%) in e-book sales. These results are not surprising as only 8.7% of the title-month combinations under protection see successful take down notices.

The relative success of de-listings suggests that casual pirates can be more easily deterred, and a copyright protection strategy that targets these casual pirates may be more effective than one that targets more serious pirates. The difference between the coefficients is insignificant at the 10% level, however, and the interaction of both actions may be critical for the success of e-book piracy protection. Again, neither type of protection significantly affects physical book sales.<sup>34</sup>

### 6.3 Intensity of Piracy Protection

The anti-piracy company protects titles with three levels of priority as guided by the observed level of piracy. Titles that are assigned a low priority are crawled approximately once every month, while medium priority protection corresponds to crawling about once per week, and high priority means the title is searched for every day. One might assume that high priority titles are about seven times as costly to protect as medium priority titles. Do those differences in costs outweigh 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 50.3%. The effect for medium priority protection is an order of magnitude smaller. The trend holds true for mass market paperbacks - a format that can be seen as a close substitute to online piracy - and to some extent for the other formats, as well. High-intensity protection is

---

<sup>34</sup>The findings are robust to different definitions of successful take downs. I show in appendix section A.3 that the results remain almost identical when I use a measure of site traffic instead of the number of infringing sites to determine whether take down notices are successful.

at least the most effective, and it might be the most (cost) efficient as well.

Of course, the decision to assign a certain priority level to a book in a given time period introduces another type of endogeneity. The more a title's copyright is infringed, the more regularly it will be searched for by the anti-piracy company, and the more effort the title's protection requires. In the analysis, I control for the level of piracy (the active sites) and for the overall level of reader interest for the title and author (Google search volume) in addition to the controls for recent editions and title age. The fact that the difference is large even after controlling for the different measures of piracy and overall interest suggests that high-frequency crawling indeed is more effective than low-frequency crawling. Titles that are protected more intensely benefit more from protection than those that are crawled less frequently.

## 6.4 Different Types of Titles

The above results show the overall effects of piracy protection on unit sales and the mechanisms behind such effects, but book piracy can impact different titles in different ways. The effectiveness of piracy protection depends on the title's popularity as well as the readers it attracts. I explore these differences here. I first analyze in which ways the effect of piracy protection depends on the title's popularity. After that, I explore the role of a title's genre.

Piracy affects regular sales through two counteracting channels. The illegal versions can steer consumers away from legal options (displacement effect), but the consumer may also find it easier to hear about a product if free versions are more readily available, potentially increasing legal sales (promotional effect). 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 could actually spur demand for the title by making it accessible to more people. My results support this idea in part, although the displacement effect outweighs the promotional effect for all types of works.

I use the number of Goodreads reviews to proxy for how well-known a title is. 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 reviews.<sup>35</sup> I use this variation across titles to

---

<sup>35</sup>Number of reviews as of August 2, 2014.

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). I treat a title as well-known if it has more than 500 reviews, and as obscure if it has fewer reviews.<sup>36</sup>

The size and direction of this promotional effect depend 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. I interact a measure of title quality with the protection variables for the obscure titles in a triple-differences analysis. I treat a title as "good" if its Goodreads rating is above 3.8 - the mean of the titles' ratings across all formats.

Table 7 summarizes the effect of piracy protection on well-known titles and on those that are more obscure. As expected, the effect on e-book sales of popular titles is larger than that on more obscure titles, although the difference is not significant at the 10% level. Sales of popular titles increase by 20.4%, while sales of the more obscure titles increase by ten percent, although this effect is estimated less precisely. This difference supports the notion that there may be a promotional effect for obscure titles, while well known titles mostly face a displacement effect of piracy. The work's quality, on the other hand, does not significantly change the effect of piracy protection, perhaps because information on a work's quality has become readily available through the Internet (for instance through Goodreads). The results are robust to different cutoff points for popularity and quality.

Finally, while a title's popularity is correlated with the priority with which the work is protected, these two measures are not perfectly correlated. 25% of the low-priority searches are for well-known titles (according to my measure), and 11% of the high-priority searches are for obscure titles. This suggests that the intensities in section 6.3 capture more than just systematic differences across titles due to title popularity. Moreover, comparing the differences in the effectiveness of protection in section 6.3 and this section, one can see that high-frequency protection increases sales beyond what we would expect simply from being a more popular title. Given the large differences, it seems plausible that daily protection is the most efficient.

The effect of piracy protection further depends on the types of consumers that the work attracts. Some genres benefit more from word-of-mouth advertising than others. In the extreme case, the demand for textbooks likely remains constant over time (as long as the number of students remains

---

<sup>36</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.

unchanged), so that piracy mostly displaces book sales without creating much additional demand. On the other extreme, the appeal of a suspense novel by an unknown author is very unclear ex ante.

I observe several fiction and nonfiction books.<sup>37</sup> If demand for the nonfiction books is more certain than that for fiction books, one would expect the effect of piracy protection to be larger 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 those of fiction titles for all formats, although the difference is statistically insignificant at the 10% level throughout.

## 7 Conclusion and Policy 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 media industries. As a result, making the protection of a copyright in such industries effective and efficient has been a focus among academics and policy makers.

The effort to protect a work from piracy can take on different forms. Most efforts have been legislative or at least formal and broad in nature, with mixed reception and success. Many governmental (legislative) attempts have faced significant opposition among the public, with several bills being shut down either before their introduction (for example, SOPA/PIPA in the United States) or shortly thereafter (for example, HADOPI in France). Although legal efforts have been less controversial, they can be quite costly, as many efforts attempted to protect entire industries (for example, the shutdown of Megaupload), and most involved expensive law suits.

The book publishing industry showcases a rather different approach - one in which private companies provide surveillance of individual works. This type of protection is less likely to be controversial, and costly courts can largely be avoided. Whether piracy protection is effective in this setting can indicate whether we can rely on private action to prevent copyright infringement. Even more, if we know what types of works are most likely to benefit from copyright protection,

---

<sup>37</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.

policy makers and enforcers can make protection more cost efficient. This paper offers insights into the cost effectiveness of protection on three dimensions: the types of pirates offering and consuming these works, the effort with which to protect a work from piracy, and the types of works that are protected.

First, the book publishing industry may be particularly well-suited for piracy protection. The small file sizes of digital books compared to other media products make it particularly easy both to offer and to consume pirated content. As a result, one might expect more casual pirates both on the demand and the supply side of piracy. Casual suppliers are more likely to be intimidated into taking down their content, as can be seen in the large success rate of take down notices for content on cyberlockers and general websites compared to peer-to-peer networks. On the demand side, much of the effectiveness of piracy protection can be attributed to the removal of links to infringing sites from search engines - presumably a channel of exploration that is most often used by casual pirates. Since a request to de-list a link from search engines can be sent relatively easily, a more efficient copyright protection strategy could target these channels, rather than attempt to take down the content itself, especially since such attempts can become quite costly if they lead to legal escalation.

Second, protection has the largest effect when protection of a work is “intense” and infringing content is searched, de-listed and taken down frequently. The differences in effectiveness may outweigh differences in costs, although a direct comparison is beyond the scope of this study.

Finally, for well-known works and those by popular authors, online piracy mainly poses a threat to regular book sales, while authors who are less established could see a promotional effect in addition to the displacement effect (see Peitz and Waelbroeck, 2006b). My results support this idea, although the displacement effect of piracy outweighs the promotional effect regardless of the type of work.<sup>38</sup>

Note here though that I analyze a particular set of titles: those which were originally published several years ago and which 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. The time it takes for protection to see an effect is especially long when considering the sales patterns of recently published works, as titles typically are most successful right after their publication. Piracy protection may be most effective for works whose sales have already leveled off,

---

<sup>38</sup>Nevertheless, the promotional effect of free online versions has already been assumed by several emerging authors who have turned to offering excerpts of their titles for free on their own websites. This strategy allows authors to control the level of sampling and promotion, whereas they have little to no control over online piracy.

and where new piracy is less likely to emerge quickly. Future research should therefore examine whether copyright protection can also be effective for more recent works.

## References

- Aguiar, L., Claussen, J., and Peukert, C. (2015). Online copyright enforcement, consumer behavior, and market structure. *Consumer Behavior, and Market Structure (March 30, 2015)*.
- Bertrand, M., Duflo, E., Mullainathan, S., et al. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, 119(1):249–275.
- Bhattacharjee, S., Gopal, R. D., Lertwachara, K., and Marsden, J. R. (2006). Impact of legal threats on online music sharing activity: An analysis of music industry legal actions\*. *Journal of law and economics*, 49(1):91–114.
- Danaher, B. and Smith, M. D. (2014). Gone in 60 seconds: the impact of the megaupload shutdown on movie sales. *International Journal of Industrial Organization*, 33:1–8.
- Danaher, B., Smith, M. D., and Telang, R. (2015). The effect of piracy website blocking on consumer behavior. *Available at SSRN 2612063*.
- Danaher, B., Smith, M. D., Telang, R., and Chen, S. (2014). The effect of graduated response anti-piracy laws on music sales: evidence from an event study in france. *The Journal of Industrial Economics*, 62(3):541–553.
- Donald, S. G. and Lang, K. (2007). Inference with difference-in-differences and other panel data. *The review of Economics and Statistics*, 89(2):221–233.
- Hardy, W., Krawczyk, M., and Tyrowicz, J. (2014). Internet piracy and book sales: a field experiment. Technical report.
- Liebowitz, S. J. (2008). Research note-testing file sharing’s impact on music album sales in cities. *Management Science*, 54(4):852–859.
- Luo, H. and Mortimer, J. H. (2015). Copyright enforcement: Evidence from two field experiments.
- Oberholzer-Gee, F. and Strumpf, K. (2007). The effect of file sharing on record sales: An empirical analysis. *Journal of Political Economy*, 115(1):1–42.

- Oestreicher-Singer, G. and Sundararajan, A. (2010). Are digital rights valuable? theory and evidence from ebook pricing.
- Peitz, M. and Waelbroeck, P. (2006a). Piracy of digital products: A critical review of the theoretical literature. *Information Economics and Policy*, 18(4):449–476.
- Peitz, M. and Waelbroeck, P. (2006b). Why the music industry may gain from free downloading the role of sampling. *International Journal of Industrial Organization*, 24(5):907–913.
- Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy*, 110(4):705–729.
- Peukert, C., Claussen, J., and Kretschmer, T. (2013). Piracy and movie revenues: Evidence from megaupload: A tale of the long tail? *Available at SSRN 2176246*.
- Rob, R. and Waldfogel, J. (2006). Piracy on the high cs: Music downloading, sales displacement, and social welfare in a sample of college students\*. *Journal of Law and Economics*, 49(1):29–62.
- Sivan, L., Smith, M. D., and Telang, R. (2014). Do search engines influence media piracy? evidence from a randomized field study. *Available at SSRN 2495591*.
- Smith, M. D. and Telang, R. (2012). Assessing the academic literature regarding the impact of media piracy on sales. *Available at SSRN 2132153*.
- Waldfogel, J. (2010). Music file sharing and sales displacement in the itunes era. *Information economics and policy*, 22(4):306–314.
- Waldfogel, J. and Reimers, I. (2015). Storming the gatekeepers: Digital disintermediation in the market for books. *Information Economics and Policy*, 31:47–58.
- Zentner, A. (2006). Measuring the effect of file sharing on music purchases\*. *Journal of Law and Economics*, 49(1):63–90.
- Zhang, L. (2014). Intellectual property strategy and the long tail: evidence from the recorded music industry. *Available at SSRN*.



Table 1: 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 were successfully taken down

Table 2: Baseline Results: Piracy Protection

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Protection	-0.0649 (0.338)	-0.167 (0.192)	0.103 (0.136)	0.342 (0.415)	0.144* (0.0767)
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,104
R-squared	0.024	0.043	0.039	0.041	0.081
Number of Titles	60	88	59	43	126

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The dependent variables are log-unit sales in each format. The observations are on the weekly level for the physical formats, and on the monthly level for e-book sales. The percent change in unit sales as a result of protection is calculated as  $e^{\beta_{\text{Protection}}} - 1$ .

Table 3: Effect on E-book Sales (all controls)

VARIABLES	(1) E-book Sales	(2) E-book Sales	(3) E-book Sales	(4) E-book Revenue
Protection	0.153* (0.0775)	0.178** (0.0835)	0.147** (0.0617)	0.157** (0.0642)
Recent e-book	0.244** (0.104)	0.274*** (0.103)	0.388*** (0.0785)	0.351*** (0.0812)
Recent print	-0.366 (0.320)	-0.372 (0.335)	0.159 (0.176)	-0.0136 (0.199)
Recent audio	0.247 (0.249)	0.255 (0.239)	0.270* (0.157)	0.265 (0.174)
Title search volume		0.0152 (0.00925)	0.00903 (0.00709)	0.0110 (0.00769)
No title interest		-0.140 (0.359)	-0.110 (0.258)	-0.119 (0.287)
Author search volume		0.0000546 (0.000813)	-0.000662 (0.000862)	-0.000429 (0.000819)
No author interest		-0.203*** (0.0638)	-0.394*** (0.105)	-0.332*** (0.0863)
Active sites		-0.0198* (0.0113)	-0.0108 (0.00876)	-0.0137 (0.00851)
Title age		0.0229** (0.00871)	-0.00608 (0.00647)	0.00334 (0.00638)
(Title age) <sup>2</sup>		-0.00000859 (0.00000930)	0.0000263*** (0.00000720)	0.0000149** (0.00000747)
Log price			-1.483*** (0.155)	
Time FE	monthly	monthly	monthly	monthly
Title FE	yes	yes	yes	yes
Observations	3,104	3,041	3,041	3,041
R-squared	0.084	0.091	0.479	0.127
Number of Titles	126	120	120	120

Robust standard errors are clustered by title. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1. The dependent variables are log-e-book sales in columns 1 through 3, and log-e-book revenue in column 4. Recent editions are 1 if an edition was published in the previous two months. Title and author search volume are normalized such that Arthur C. Clarke's *2001: A Space Odyssey* has a search volume of 100 in January 2011. No title/author interest is 1 if there was not enough search volume for Google Trends to record a number. Title age is in months.

Table 4: Google Search volume around the protection start date

	(1)	(2)	(3)	(4)
	Title interest	Title interest	Title interest	Author interest
Timing	-0.0894 (0.0831)	-0.0326 (0.0263)	-0.0309 (0.0195)	-0.118*** (0.0287)
Timing <sup>2</sup>	0.00297 (0.00296)	0.00169 (0.00164)	0.00127 (0.00124)	0.00704*** (0.00260)
Recent editions	✓	✓	✓	✓
Title FE	✓	✓	✓	✓
Year/month FE	✓	✓		✓
Observations	2234	3315	3315	3315
R-squared	0.034	0.023	0.014	0.035
Number of Titles	126	126	126	126

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . Columns 1 through 3 look at the Google search volume for each title, and column 4 looks at the search volume for the author. Column 1 restricts the sample to the 6 months before and after the protection start date. The other columns include all observations.

Table 5: De-listings and take downs

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Link de-listings	0.130 (0.352)	-0.153 (0.182)	0.155 (0.144)	0.228 (0.400)	0.147** (0.0617)
Takedown success	0.0485 (0.117)	0.00569 (0.0834)	0.115 (0.105)	0.0569 (0.162)	0.0135 (0.0738)
Log price					-1.482*** (0.155)
Recent edition controls	✓	✓	✓	✓	✓
Title age controls	✓	✓	✓	✓	✓
Active site controls	✓	✓	✓	✓	✓
Title search volume controls					✓
Author search volume controls					✓
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,041
R-squared	0.034	0.046	0.047	0.059	0.480
Number of Titles	60	88	59	43	120

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The dependent variables are log-unit sales in each format. Controls for search volume are only included in the e-book format because the search volume information is only available on the monthly level for many books.

Table 6: Piracy Protection - Priority Levels

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Low priority	-0.363 (0.241)	-0.997** (0.397)	-0.975 (0.932)	-0.0270 (0.521)	0.0750 (0.0837)
Medium priority	0.242 (0.628)	-0.297 (0.339)	-1.001 (0.775)	0.911* (0.500)	0.0147 (0.0834)
High priority	0.0425 (0.261)	-0.0583 (0.154)	0.455*** (0.158)	-0.252 (0.405)	0.407*** (0.150)
Log price					-1.475*** (0.155)
Recent edition controls	✓	✓	✓	✓	✓
Title age controls	✓	✓	✓	✓	✓
Active site controls	✓	✓	✓	✓	✓
Title search volume controls					✓
Author search volume controls					✓
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,041
R-squared	0.035	0.051	0.062	0.063	0.48
Number of Titles	60	88	59	43	120

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The dependent variables are log-unit sales in each format. Low priority corresponds to titles being searched for once a month, medium priority titles are searched for once a week, and high priority titles are searched for daily.

Table 7: Piracy Protection - Displacement and Promotion

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Popular	0.238 (0.429)	-0.152 (0.217)	0.415* (0.222)	0.418 (0.444)	0.185* (0.107)
Obscure × “good”	-0.182 (0.378)	-0.166 (0.353)	0.885*** (0.160)	-0.221 (0.382)	0.103 (0.0856)
Obscure × “bad”	-0.171 (0.401)	-0.138 (0.922)	-1.933* (0.992)	-0.879 (1.045)	0.131 (0.0818)
Log price					-1.479*** (0.153)
Recent edition controls	✓	✓	✓	✓	✓
Title age controls	✓	✓	✓	✓	✓
Active site controls	✓	✓	✓	✓	✓
Title search volume controls					✓
Author search volume controls					✓
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,041
R-squared	0.035	0.046	0.059	0.061	0.480
Number of Titles	60	88	59	43	120

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The dependent variables are log-unit sales in each format. A title is classified as popular if it has more than 500 reviews on Goodreads, and it is considered “good” if its average rating is above 3.8.

Table 8: Effect of Protection: Fiction and Nonfiction Titles

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Fiction	-0.223 (0.283)	-0.239 (0.251)	0.113 (0.146)	0.283 (0.442)	0.139** (0.0663)
Nonfiction	0.820 (0.744)	0.0551 (0.223)	0.485* (0.277)	0.360 (0.897)	0.217** (0.0108)
Log price					-1.482*** (0.155)
Recent edition controls	✓	✓	✓	✓	✓
Title age controls	✓	✓	✓	✓	✓
Active site controls	✓	✓	✓	✓	✓
Title search volume controls					✓
Author search volume controls					✓
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,041
R-squared	0.036	0.047	0.047	0.058	0.480
Number of Titles	60	88	59	43	120

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The dependent variables are log-unit sales in each format.



Figure 1: Active piracy sites - I Am Legend

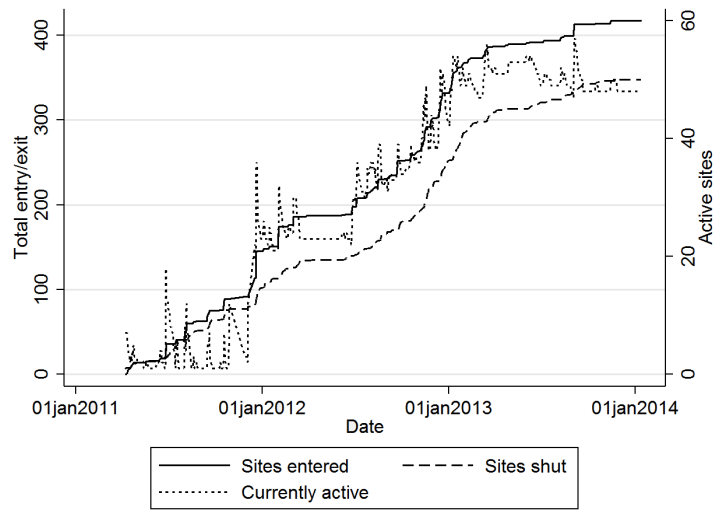
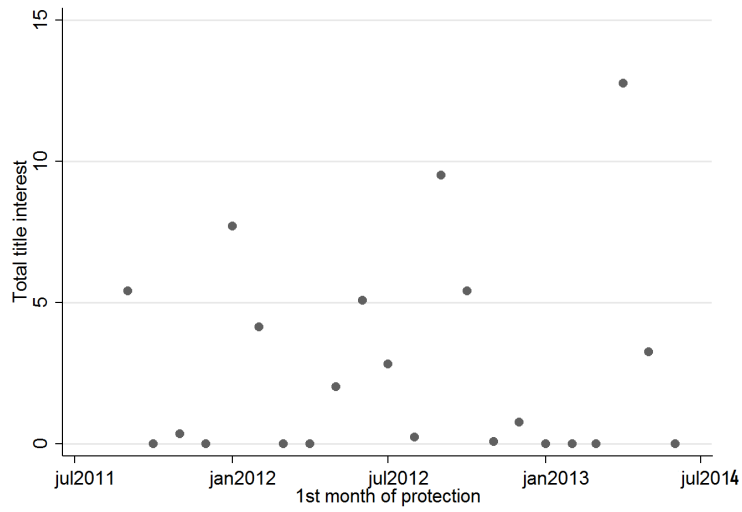
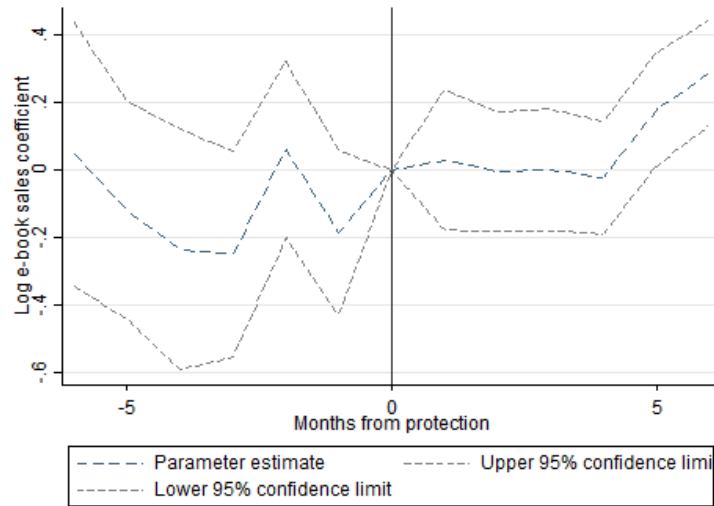


Figure 2: Total title search volume by protection start month



Total title interest is normalized average monthly Google Search volume by title. Google Search volume is 100 for Arthur C. Clarke's *2001: A Space Odyssey* in January 2011. The plotted points are the averages of all titles' search volumes which were moved into protection in a given month.

Figure 3: Protection and e-Book Sales



Month 0 corresponds to the month immediately before protection starts, and month 1 is the first month of protection. The figure shows a statistically significant increase in sales in months 5 and 6 of protection, and it remains positive thereafter, albeit less precisely estimated as the coefficients further from the protection start date are identified off fewer titles.

## A Appendix

The main results 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 have different interpretations of protection and its success, as well as a placebo test on the timing of protection. The effectiveness result of protection is robust to these alternative specifications.

### A.1 Robustness Check: Protection When Sites are Removed

The main analysis assumes that a title is constantly protected. That is, the anti-piracy company 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 9 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, the effect on e-books is smaller than it is under the assumption of constant protection. The effect remains largely significant, however, indicating again that the estimates in the main analysis are consistent. Since titles move in and out of protection, there is less concern about path dependency of the treatment variable.

### A.2 Robustness Check: Placebo Test

Despite the controls in the main estimation, there may be concern that the results pick up additional effects that are not accounted for and remain unobservable to the econometrician. Figure 3 suggests that the unit sales right before protection begins are not significantly higher than the units sales in the preceding months although the coefficients are estimated quite imprecisely.

Here I address these concerns more directly by testing whether those books that are about to

move into protection already see an increasing sales trend compared to those books that go into protection later. I drop all data points after the title went into protection, and I set the new “protection” date between 2 and 6 months before the actual protection date. The analysis then is the same as in equation 1. Table 10 shows that while the “protection” coefficient is positive for most artificial start dates, it is much less so than in the main specification, and never statistically significant. Even if this estimation were to suggest an upward trend in sales prior to the protection start date, this trend does not explain the large, statistically significant effect that is estimated in the main specification. Protection remains cost efficient when comparing this net effect to the costs of protection.

### A.3 Robustness Check: Site Traffic as a Measure of Success

Section 6.2 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 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>39</sup> Note that the sites reported on Alexa.com are the host sites to the infringing links, meaning that I do not observe traffic to the infringing content itself.

I use the inverse of the host 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 4 shows that traffic to infringing sites is not perfectly correlated with the number of active infringing sites. Here I 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, rather than the number of infringing sites.

Table 11 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 again appears to be much stronger, although the difference remains statistically insignificant at the 10% level.

---

<sup>39</sup><http://www.alexa.com/about>, retrieved September 25, 2014.

Table 9: “Sporadic” Piracy Protection

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Protection	0.101 (0.184)	0.00764 (0.0810)	0.201** (0.0845)	-0.0482 (0.190)	0.133*** (0.0447)
Log price					-1.482*** (0.156)
Recent edition controls	✓	✓	✓	✓	✓
Title age controls	✓	✓	✓	✓	✓
Active site controls					✓
Title search volume controls					✓
Author search volume controls					✓
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,041
R-squared	0.033	0.046	0.047	0.057	0.480
Number of Titles	60	88	59	43	120

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The dependent variables are log-unit sales in each format. Here, “protection” is turned on if at least one infringing site is taken down.

Table 10: Placebo Test: E-book sales

	(1)	(2)	(3)	(4)	(5)
	E-books	E-books	E-books	E-books	E-books
Protection	0.0878 (0.101)	0.141 (0.0974)	0.0420 (0.0862)	-0.0178 (0.0873)	0.0348 (0.102)
Log price	-1.182*** (0.164)	-1.176*** (0.162)	-1.185*** (0.166)	-1.339*** (0.256)	-1.709*** (0.283)
Recent edition controls	✓	✓	✓	✓	✓
Title age controls	✓	✓	✓	✓	✓
Active site controls					✓
Title search volume controls					✓
Author search volume controls					✓
Months early	2	3	4	5	6
Observations	1229	1223	1215	1158	1111
R-squared	0.442	0.444	0.444	0.447	0.472
Number of Titles	101	99	97	83	75

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The dependent variables are log-unit sales in each format. The table drops all data points after the title went into protection and sets an artificial start date some months earlier, as specified by “Months early.” For example, if a title’s actual protection start date was July 2012, then column 1 sets it at May 2012, column 2 sets it at April 2012, and so on. The number of titles decreases as I drop those titles that now have no “pre-protection” period.

Table 11: De-listings and take downs: using traffic information

VARIABLES	(1) Hardcover	(2) Trade PB	(3) Mass Market	(4) Audio	(5) E-books
Under contract	0.0951 (0.350)	-0.137 (0.183)	0.182 (0.145)	0.249 (0.395)	0.135** (0.0601)
Successful takedown	0.0587 (0.126)	0.00624 (0.0762)	0.00966 (0.0918)	0.00321 (0.161)	0.0226 (0.0722)
Traffic on active sites	0.131 (0.421)	0.0899 (0.168)	0.337 (0.207)	-0.433 (0.444)	0.119 (0.188)
Log price					-1.483*** (0.155)
Recent edition controls	✓	✓	✓	✓	✓
Title age controls	✓	✓	✓	✓	✓
Time FE	weekly	weekly	weekly	weekly	monthly
Title FE	yes	yes	yes	yes	yes
Observations	9,350	13,506	8,714	6,630	3,041
R-squared	0.033	0.046	0.048	0.060	0.480
Number of Titles	60	88	59	43	120

Robust standard errors are clustered by title. \*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$ . The dependent variables are log-unit sales in each format. Take downs are considered successful if the imputed site traffic to the infringing content is lower at the end of this time period than at the end of the previous time period.



Figure 4: Active piracy sites and estimated traffic - I Am Legend

