Institute for Economic Studies, Keio University

Keio-IES Discussion Paper Series

書籍海賊版の売り上げへの影響：日本のマンガをケースとして

田中辰雄

2016 年 12 月 29 日
DP2016-027
https://ies.keio.ac.jp/publications/7121/

Institute for Economic Studies, Keio University
2-15-45 Mita, Minato-ku, Tokyo 108-8345, Japan
ies-office@adst.keio.ac.jp
29 December, 2016

書籍海賊版の売り上げへの影響：日本のマンガをケースとして

田中辰雄

IES Keio DP 2016-027
2016 年 12 月 29 日
JEL Classification: D12; L82; M3; O34
キーワード: 著作権; 海賊版; 電子書籍; 差の差; マンガ

【要旨】

インターネット上の書籍の海賊版が正規版の売上に与える影響を、日本の漫画を事例として分析した。海賊版の流布程度を個別タイトル単位で調査し、それが正規版の売上に与える影響をみるとともに、出版社が共同で行った海賊版ファイルの削除作業を自然実験として、差の差による分析も試みた。二つの分析方法の結果は一致しており、次の事実が確認できた。海賊版は現在連載中の作品、特にその最新版の売り上げを減少させる。その一方、すでに完結した作品の売り上げを増加させる。前者の効果は海賊版が正規版を置き換える通常の被害である。後者の効果は、海賊版が、宣伝されずに忘れさられている過去作品を読者に思い出させる効果（remind effect)と考えられる。

田中辰雄
慶應義塾大学経済学部
〒108-8345
東京都港区三田2-15-45
tatsuo@econ.keio.ac.jp

The Effects of Internet Book Piracy: The Case of Japanese Comics

First version: October 18, 2016

Tatsuo TANAKA
Faculty of Economics, Keio University
tatsuo@econ.keio.ac.jp

Abstract

In this study, the effects of internet book piracy in the case of the Japanese comic book market were examined using direct measurement of product level piracy ratio and a massive deletion project as a natural experiment. Panel regression and difference-in-difference analysis consistently indicated that the effect of piracy is heterogeneous: piracy decreased the legitimate sales of ongoing comics, whereas the legitimate sales of completed comics increased. The latter result is interpreted as follows: piracy reminds consumers of past comics and stimulates sales in that market.

Keywords: copyright, piracy, e-book, comic, Japan, difference-in-difference
JEL: D12,L82,M3,O34

I Introduction

Whether piracy reduces legitimate sales has been a hot issue since the internet became the common infrastructure and provided users with new piracy options such as file sharing. Natural intuition tells us that piracy displaces legitimate sales because consumers are entertained by the same product without paying a fee. However, if piracy has some kind of advertisement effect such as sampling, network effect, or market-enlarging through word-of-mouth, and that overrides the displacement effect, legitimate sales may not be decreased (Takeyama, 1994; Peitz & Waelbroeck, 2008). Since copyright theory indicates that both the displacement effect and advertisement effect are possible, the question is left for empirical analysis.

There have been many econometric studies on the impact of piracy in the recent two decades with respect to the music and motion picture industries. Accumulated studies are surveyed in Liebowitz (2006), Oberholzer and Strumpf (2010), and Danaher, Smith and Telang (2013). However, there are very limited econometric studies on piracy in the book industry. One of the reasons is that piracy of books has been limited compared with music and movies because digitization of books lagged behind music and movies. Additionally, e-book readers such as Kindles are specialized devices, not general purpose devices like PCs, making them relatively free from piracy. Piracy can be suppressed on specialized devices more easily due to the limited software working on it. Nevertheless, digitization of books has momentum in this decade to push up the share of e-books to over 40% of publishers' total net revenue in the US and over 10% of sales in Japan in 2014.[1] As people get accustomed to reading e-books on various devices, including general purpose devices such as PCs or smartphones, piracy of books may become a serious concern for publishers in the near future.

The intent of this paper is to examine the displacement and/or advertisement effect of piracy in the case of comics in the Japanese book industry. There are three reasons we chose comics in Japan as a research target for book piracy. Firstly, piracy in comics is already very common compared with other book categories. One can find pirated illegal copies of almost all popular comics on the internet, and, as soon as a new volume is released from publishers, a pirated copy is uploaded onto piracy sites very quickly. Japanese comic publishers belong to an anti-piracy association named CODA (Content Overseas Distribution Association), which insists that loss caused by overseas piracy is estimated to be double the overseas legal revenue from 2014.

[1] Source: American Association of Publishers for US and Zenkoku Shuppan Kyoukai for Japan

Secondly, CODA carried out a massive deletion project from July of 2015 till March of 2016, which provided us with a natural experiment to test the impact of piracy. Publishers chose some of their comic titles as the target of this project, and CODA sent repeated warnings to piracy sites to delete all illegal files from July of 2015. This measure resulted in the decrease of illegal files of target comics to a small but significant extent. If we compare the sales of target comics (treatment group) with non-target comics (control group), we can evaluate the effect of this deletion project. In other words, we can apply difference-in-difference method to the comics market piracy case.

Thirdly, Japanese comics consist of multiple volumes, which reduces the endogeneity problem. Since the early volumes were published several years ago, consumers know the style and quality of the comic very well. Thus, we can assume that the popularity of the comics is established and fixed during the research periods. If this assumption holds, fixed effect term of panel regression removes the effect of the comics' popularity, which is the source of the endogeneity problem annoying piracy researchers.

We measured piracy ratio by the number of piracy sites that have access to the comics. We counted the number of piracy sites with respect to 3,360 volumes of 484 comic titles for eight months and applied ordinary panel regression and a difference-in-difference model. Both approaches consistently showed that the effects of piracy on legitimate sales are heterogeneous depending on whether the comics are ongoing or completed. Piracy decreases sales of ongoing comics, but it increases sales of completed comics. To put this another way, displacement effect is dominant for ongoing comics, and advertisement effect is dominant for completed comics. Since completed comics series have already ended, and publishers no longer do any promotion for them, consumers almost forget completed comics. We can interpret that piracy *reminds* consumers of past comics and stimulates sales.

The remainder of the paper is organized as follows. Section II contains a brief summary of the literature of studies on piracy, and, in section III, the Japanese comic industry is described. In sections IV and V, we explain the data collection procedure and the model specifications, respectively. Section VI contains a report of the results of panel regression using direct measurement of piracy, while section VII features the result of difference-in-difference analysis. Section VIII consists of the summary and discussion.

II Literature

Although early studies on the effects of piracy exploited internet penetration ratio as a proxy of piracy (Liebowitz, 2008; Zentner, 2005), this ratio is not a good measure of piracy because internet usage shifts consumer demand from old style entertainment media such as CDs or DVDs to new style media such as online music services with fixed fees, YouTube videos, smartphone applications, SNS chatting, basic web surfing etc. Therefore, CDs and DVDs are crowded out because of consumers' budget and time constraints, and sales may decrease even if there is no piracy.

A more correct approach is to measure each product's piracy ratio directly and look at the correlation between piracy and sales. In this approach, endogeneity is a key issue to be solved because, when content becomes popular, both sales and piracy simultaneously increase, resulting in a spurious positive correlation between sales and piracy. Another improved approach is to exploit an exogenous shock as a natural experiment to estimate the causality of piracy. If an exogenous shock, such as legal enforcement, occurs to part of the sample, we can compare the consequences between the sample with the shock (treatment) and the sample without the shock (control).

The results of these two approaches, unfortunately, seem to be mixed so far. Studies using a product level piracy ratio tend to indicate that piracy does not reduce legitimate sales. Oberholzer-Gee and Strumpf (2007) exploited file sharing data from a P2P server and found that piracy did not reduce music CD sales, using a German school holiday as an instrument variable. Martikainen (2011), Smith and Telang (2009), and Tanaka (2004) also measured the piracy ratio directly and reported that piracy did not harm legitimate sales.

Studies using exogenous shock as a natural experiment, on the other hand, tend to indicate that piracy is harmful to legitimate sales. Danaher, Smith, Telang, and Chen (2014) reported that French anti-piracy law, Hadopi (the so-called three strikes law), increased French iTunes music sales by 22-25%. This was shown via difference-in-difference method using music sales of other EU countries as the control variables. Adermon and Liang (2014) reported similar results in the case of Sweden's copyright law enforcement in 2009, and Danaher and Smith (2014) found the revenue of major movie studios increased after the shutdown of famous cyber locker site, MegaUpload. These natural experiments suggest the displacement effect is dominant against piracy on the internet.

In the case of the book industry, to our knowledge, there is no study that either directly measures piracy ratio or exploits a natural experiment. One excellent

econometric study is a field experiment done by Hardy, Krawczyk, and Tyrowicz (2014). They chose 94 pairs of similar books, from which one was randomly assigned to be the treatment and the other the control. They intensively deleted unauthorized copies of treatment books during one year, leaving control books to be pirated freely. Comparison of the sales of pairs showed that removal of unauthorized copies had no bearing on legitimate sales. Other literature on the effects of book piracy on the internet is anecdotal. For example, Hilton (2011) reported ten weeks free exposure of eight books online resulted in a 26% increase in legal sales. O'Leary (2009) also claimed there was little connection between file sharing and book sales after putting eight books on free distribution. In spite of this anecdotal evidence, in each case the sample size was too small to evaluate statistically

This study is the first to examine the effect of internet piracy on the book industry by using a large sample — albeit one that is limited to the specific category of comic books — and applying both a direct measuring approach and a natural experiment approach. We obtained consistent results with these two approaches, which have often shown opposite results in the case of the music and movie industries.

III Industry

The Japanese comics industry consists of comic books and comic magazines, the total sales of which were 4.046 billion dollars in 2014. Of these two categories, we focus on comic books, whose corresponding category in the US is comic and graphic novels. The size of the comic book market of these two countres is shown in Table 1 for comparison. In 2014, the total sale of comic books in Japan was 2.853 billion dollars — almost three times more than US sales of 0.935 billion. When we consider the difference of total book market size between Japan and the US, the difference in comics market size is impressive. If we divide the sales of comics by total book market size, we find that the share of comics in the total book market is 36.1% in Japan, as compared to 2.6% in the US. This means the size of the comics market in Japan is, relatively speaking, over ten times bigger than its counterpart in the US. In Japan, approximately one third of all books are comics. When we divide the sales of comics by total population, we obtain per capita expenditure on comics. This figure amounted to $2.90/person in the US and $22.40/person in Japan. The latter number increases to $31.80/person if we include comic magazines. This implies that the propensity of buying comics is ten times higher in Japan than in the US. Comics (called "manga" in Japan) is a major category in the Japanese book industry.

Table 1　Size of comics market in Japan and the US (2014)

	comic books	all books	share of comics	expenditure per capita	
	(million$)	(million$)	(%)	($/person)	
Japan	2,853	7,898	36.1%	22.4	(31.8)*
US	935	35,995	2.6%	2.9	

(excluding magazines)

* when magazine is included

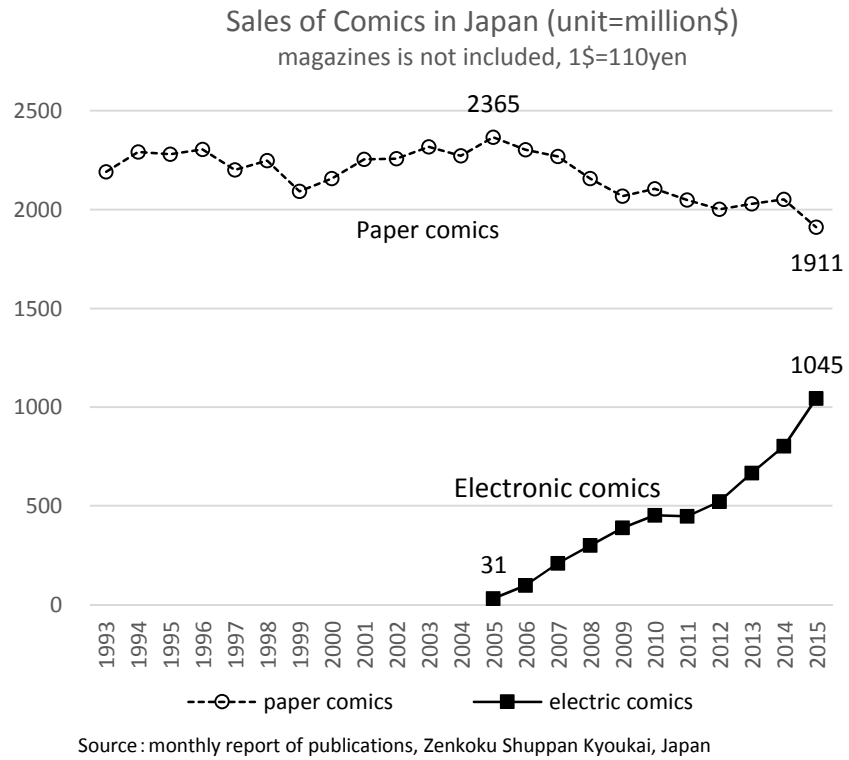
source: ICv2.com

Annual Report and 2014 Global Publishing Statistics, International Publishers Association, monthly report of publications, Zenkoku Shuppan Kyoukai, Japan

The market structure of the comics publishing industry is oligopoly, both in the US and Japan. In the US, two dominant publishers, Marvel Comics and DC Comics, had 100% share in the top 20 ranking and 96% of top 50 ranking in July 2016. There are four major publishers in Japan: Shuei-sha, Kodan-sha, Shougakkan, and KADOKAWA, which held 98% of the top 20 comics ranking in July 2016. (Ranking data is based on ICv2.com for the US and ORICON for Japan).

Figure 1 shows the sale of comics in Japan from 1993 to 2015. Paper comics peaked at 2,365 million dollars in 2005 and have gradually declined to 1,911 million dollars in 2015. Some publishers claimed that this decline was caused at least partly by piracy. However, since sales of electronic comics started to increase from 31 million dollars in 2005 to 1,045 million dollars in 2015, the sales decline of paper comics may be due to just a demand shift from paper to electronics, as is often the case in the digital age. Whether comic piracy harms legal sales or not is still an empirical question.

Figure 1



Source : monthly report of publications, Zenkoku Shuppan Kyoukai, Japan

Unlike the case of the music industry, the main mode of comic piracy is not file sharing, but piracy sites, which consist of cyber locker sites and reach sites. Illegal files are uploaded to the server of cyber locker sites, and reach sites provide users with a search service to find links to the files. Users search for a comic name on the reach site and obtain the link to the file on the cyber locker site. Then, they download the pirated comics from the cyber locker sites. Cyber locker and reach sites are managed by different firms.

The main reason for this division of labor is to avoid legal prosecution. It is illegal to keep pirated files on a server when it is known that they are pirated files. To put it another way, if the individuals who run the cyber locker site are unaware that the file is pirated, they do not face legal punishment. When a copyright holder requests the cyber locker site delete the file, the site administrator must do so because, at that time, he has learned that it is an illegal file (so-called “notice-and-takedown” policy). Reach sites have no legal obligation because they list only URL links. If the two sites were integrated, it would be clear that the administrator of the integrated site keeps the file on the server with the knowledge that it is illegal, because the site provides a comic search service. Thus, the integrated site would be shut down.

IV Data

In July of 2015, the Content Overseas Distribution Association in Japan (CODA) started the Manga Anime Guardian Project, which requested cyber locker sites to delete the pirated comics files. This was a large project funded by the Ministry of Economy, Trade, and Industry, and most comic publishers (22 publishers) in Japan joined it. In this study, we used the data of four major publishers, Shuei-sha, Kodan-sha, Shougakkan, and KADOKAWA, whose total share in the top 30 comic sales in annual ranking during 2010-2014 was almost 95%.

Deleting pirated files is a cat-and-mouse game since deleted files are soon restored. However, this project still had some effect because it covered many publishers and sent requests repeatedly. The actual requesting operation was carried out by a firm hired by CODA, which crawled webs to find pirated files and sent warning messages continuously until the files were deleted. Reportedly, over 90% of piracy files were deleted by these requests, though new files were soon restored with new URLs. In section VII, we show via difference-in-difference method that this project had a small but significant effect on piracy level.

As mentioned in the previous section, there are two types of piracy sites: reach sites and cyber locker sites. Cyber locker sites were the target of this deletion project. For the purpose of our research, however, reach sites were more important than cyber locker sites because, from the view of users, reach sites are the places for they find pirated comics. Let us assume that one cyber locker site is shut down, but all reach sites keep active links to pirated files. This assumption is reasonable because reach sites usually list links to multiple locker sites, and/or they can add a new link to another cyber locker site. If the reach sites maintain active links, there is no change to users because they can access pirated files as before. However, if one reach site is shut down, it causes an inconvenience to the users of the site because they have to find another reach site to access pirated comics. Therefore, we focused on the activities of reach sites in our research.

Unfortunately, the number of downloads of each comic title is not open to the public. Thus, we had to use a proxy of the degree of piracy for each comic. For this purpose, we used the number of reach sites that had active links. For example, when there were ten reach sites with active links for comic A and five reach sites for comic B, we assumed that comic A was more pirated than comic B because more users were able to find the link to the pirated files of comic A than comic B. Hardy, Krawczyk, and Tyrowicz (2014) also used number of available piracy files on the internet as a proxy of

the degree of piracy.

There are many reach sites on the internet. Of these reach sites, we chose 16 major Japanese active reach sites to be researched. "Active" means that the links have been updated recently. "Japanese" means that the pirated comic files have not been translated to non-Japanese language. Since the Manga Anime Guardian Project targeted Japanese sites only, we also limited our research scope to the Japanese site. These 16 Japanese sites are "major" sites in the sense that the pirated files are well-managed by the administrators and the sites keep complete volumes of many comics, whereas other minor sites are not managed by administrators and often lack complete volumes. The largest reach site of our 16 sites covers 81% of all targeted comics, and the smallest site covers 21%. Since coverage of other minor sites is probably below 20%, we suppose 16 reach sites cover most piracy users.

Almost all comics series consist of multiple volumes. Of these volumes, we chose the first, middle, and most recent two volumes as samples. If the comic consists of ten volumes, for example, we count pirate links of volume 1, 5, 9, and 10, respectively. The reason for this choice is that readers of these volumes have different characteristics. The readers who buy recent volumes are more likely fans of the comic and have purchased previous volumes already, whereas the ones who buy the first or middle volume have noticed the comic recently and started reading. This heterogeneity may influence the effects of piracy. In order to take this heterogeneity into consideration, we count the number of reach sites where the first, middle, and most recent two volumes were linked.

Both sales and number of reach sites are monthly data. Sales data starts in March of 2015 and ends in February of 2016. Counting the number of reach sites began in June of 2015, a month before the massive deletion project started, and ended in January of 2016.

To make the data structure clear, we show the hypothetical case of a comic in Table 2. This comic is assumed to consist of ten volumes, volumes 9 and 10 of which were published in August and December of 2015, respectively. Volumes 1 and 5 were published several years ago. Distribution of sales of newly published volumes 9 and 10 was skewed because sales were highly concentrated to the initial few months. The number of reach sites that had active links to volume 1 and 5 slightly decreased from six sites in June to approximately five sites after July, thanks to the massive deletion project. The decrease in the number of reach sites was small because of the cat-and-mouse game structure of deletion. Regarding volume 9, links to pirated files do not appear on the reach site in the published month of August, but they appear on five

sites a month later. As for volume 10, three reach sites began listing links in the publishing month. This paper's research question is whether the number of reach sites reduced comic sales and whether the massive deletion project increased sales.

Table 2　Data Structure (example)

	year	2015										2016	
	month	3	4	5	6	7	8	9	10	11	12	1	2
sale	vol1	100	100	100	100	100	400	200	100	100	500	200	100
	vol5	100	100	100	100	100	300	200	100	100	300	100	100
	vol9						4000	1000	800	500	400	400	400
	vol10										5000	1500	1000
number of reach sites	vol1				6	5	5	4	5	5	6	5	
	vol5				6	5	5	4	5	6	6	5	
	vol9						0	5	5	6	6	5	
	vol10										3	5	

↑ deletion start in July

↑ counting start in June

Covered in this research are comics published by four major publishers (Shuei-sha, Kodan-sha, Shogakkan, and KADOKAWA). These four publishers listed 322 comic titles as the targets of the massive deletion project. As control data for difference-in-difference regression, we added another 162 comics published by these four publishers that were not the target of the massive deletion project. Therefore, the total number of comics we counted on the reach sites was 484 (=322+162). Almost 70% of comics are digitized; therefore, sales data of the electronic versions was also available. We treated sales of the paper version and electronic version separately because electronic books and paper books are not interchangeable (Hu & Smith, 2013), and substitutability with pirated files may be different between electronic and paper comics. Adding the paper versions and electronic versions together, the total sales data to be analyzed increased to 840, 551 of which belonged to the treatment group and 289 of which belonged to the control group. Since we chose four volumes from each comic, the total number of volumes was 840*4=3360, which is the maximum number of groups of the panel regression. With approximately ten months of data for each volume, as shown in Table 1, the total data size was over 30,000.

Of our 840 comics, 601 were ongoing comics in the research periods, meaning that they ended in 2015 or continued into 2016. The remaining 239 were completed comics that　lasted until 2014. These categories, ongoing and completed, are important in this

paper.

Counting the number of reach sites was done manually rather than via web scraping software because URL expression of links on reach sites changed over time, and they sometimes became inactive. As a result of the massive deletion project, piracy files on the cyber locker sites were frequently deleted and soon restored. Therefore, the URL links on the reach sites often became inactive or expressed differently. To obtain accurate information, we visited every reach site monthly and checked to see if the links were active or not by clicking the URLs and exiting just before the download process began. In other words, we checked 16 reach sites every month with regards to 3,360 volumes.

Book sales data is not shipping-based but actual sales based which was offered by the four publishers.

V Model Specifications

We used two models to examine the effect of piracy on legal sales. The first was ordinary panel regression, which correlated comic sales to the number of reach sites controlling the comics' popularity as fixed effect.

$$\log(Yit) = a_i + b*\log(Site_{it}+1) + c*Z_{it} + e_{it} \qquad (1)$$

Y_{it} represents the sales of comics i in month t, whose unit is number of books. $Site_{it}$ is the number of reach sites that have active links to comic i in month t. Log transformation was applied assuming nonlinear effect of reach site on the sales. Z_{it} is a vector of other exogenous variables that affect the sales of comics, such as media mix, TV promotion, animation etc. The error term e_{it} is an idiosyncratic random variable. The coefficient b represents the effect of piracy — either displacement effect ($b<0$) or advertisement effect ($b>=0$).

The popularity of individual comics, which is a source of endogeneity, was assumed to be controlled by fixed effect a_i. This assumption is valid if popularity was constant within the research periods. The assumption that popularity did not change during the research periods seems unrealistic and misleading when compared with the music industry studies. However, unlike the music industry, there were two reasons to adopt this assumption in our research. First, in the case of comics, the quality of product was already known to consumers because all comics consisted of multiple volumes, and the early volumes were published several years ago. In the case of the music industry, the

quality of CD albums is not known to consumers at the time when the albums are just released. Since the quality of CD albums changes from album to album even if the artist is the same, consumers gather information about the quality of the album through reviews in music magazines, comments on music web sites, word-of-mouth, trial listening etc. If consumers find the quality of the album to be excellent, they begin to purchase CDs or download its pirated files. Thus, in the music industry studies, CD sales and downloading piracy naturally increased simultaneously in the research periods, which inevitably caused an endogeneity problem. That is the reason why most studies on music CD piracy have exploited instrumental variables.

However, in the comic case, early volumes were published several years ago. The median starting year in our data is 2009. Thus, half of the comics in our study started over six years ago. Seventy five percent of comics started before 2012, and 95% started at least one year prior to the research period. What this means is that information about the comics such as characters, storyline, and reviews are already open to public. Consumers can make their subjective evaluations of the comic before publication of new volumes. In other words, consumers know the quality of the comic already because each volume is a part of the same content that started years before. Hence, it is a reasonable assumption that the popularity of the comics was established and constant during the research periods. The main factors varying the sales with time in the research periods were time trends and promotions like TV animation or electronic promotion, which were represented by controlled variable Z.

The second reason that we assumed fixed effect could control endogeneity is that the explanatory variable, the number of reach sites, is determined by site administrators and not by consumers. In case of music piracy, both the number of downloads and CD sales are determined by the same consumer; thus, correlation between them is almost certain and exactly simultaneous. In our research, the explanatory variable was not the number of downloads but the number of sites with active links to the comic, and whether reach sites list links to the comic or not is determined by site administrators who do not have information about the changing popularity of the comic among consumers.

How can administrators simultaneously learn about a change in the popularity of comics? Although product price is a signal of demand change in the ordinary market transaction, there is no price of piracy. The price of paper books cannot be a signal because the retail book price is fixed by publishers in Japan, owing to a resale price maintenance system. Comic ranking in ORICON is useless because it shows only the top 50 sales weekly; hence, sales of the remaining titles, which make up over 90% of

the total, are unknown. If the reach site already had the link to the comic, the administrator could learn the change in popularity by looking at the consumer click count of the link. However, since we are discussing whether the administrator lists the link on the site or not, the site does not have the link yet. There is no systematic way for administrators to simultaneously know the change in popularity of comics they have not listed yet. Therefore, there will be considerable lag for administrators to learn the change in popularity and add new links to their reach sites, which make simultaneous correlations between changes in popularity and the number of reach sites very low or almost zero. If popularity change does not correlate with explanatory variables, we can include it in the error term e_{it} without endogeneity bias.

Considering these two reasons, it is not an unrealistic assumption that fixed effect a_i can control the endogeneity of the popularity of comics. Therefore, as the first approach, we adopted the ordinary panel regression, assuming that fixed effect a_i controlled the popularity of the comic.

Although the assumption of constant popularity might be probable, it is not certain. Changes in ongoing comics such as a shift in storyline or the death of a popular character can affect sales even within a research period. Also, exogenous shocks other than Z (tv animation or promotion) impact popularity. For example, if a big sports event such as the summer Olympics is held in the research periods, sports comics may be sold more during the event periods, and administrators may notice this change and quickly added new related links to their sites. Thus, popularity can change during a research period to some extent, which causes an endogeneity problem.

To cope with this problem, as a second approach, we applied standard difference-in-difference regression using the massive deletion project that started in July 2015 as a natural experiment.

$$Log(Yit) = a_i + b_0 * DafterJuly_t + b * Deletion_i * DafterJuly_t + c * Z_{it} + e_{it} \quad (2)$$

$DafterJuly_t$ is a dummy variable equal to one if t is after July 2015, and $Deletion_i$ is a dummy variable equal to one if comic i is a target of the massive deletion project. If coefficient b is positive, we can say that the massive deletion project increased the sales of comics because targeted comics were sold more than non-targeted comics after the deletion project, which suggests that piracy decreased legal sales. This regression can control the endogeneity problem because it does not use number of reach sites as an explanatory variable.

To apply this approach, two key variables, $DafterJuly_t$ and $Deletion_i$, had to be

randomly chosen. We can suppose that DafterJuly was randomly chosen because July was selected by the technical parameters of this project, and most consumers were unfamiliar with this project. , However, $Deletion_i$ was not randomly chosen because the target comic list of this project was offered by publishers. It is likely that publishers chose popular comics or comics that they wanted to promote for specific reasons. In order to address this bias, we had to choose control comics with similar characteristics to the target comics. We examine this similarity later.

As other exogenous variables Z_{it}, we adopted the following six variables and monthly dummies.

time$_t$: Since all entertainment content gradually loses popularity after the initial boost, sales of comics also tend to decrease slowly with time. The variable *time,* therefore, is a trend variable starting from one to twelve during the research periods in order to remove this general tendency.

log(trend$_{it}$*)*: Sales distribution of newly published volumes is quite different from other volumes because sales in the first month are typically extremely high and then decrease very rapidly. To represent this skewed distribution, we adopted the variable *log(trend*$_{it}$*)*, trend, which starts from one at the month when the volume was published in the research periods. Log transformation was done to present the skewed pattern of declining sale. This variable was applied to volumes published in the research periods — that is, the recent two volumes of ongoing comics.

newpub1$_{it}$*, newpub2*$_{it}$: When a new volume is published, sales of the first and middle volumes also increase because of the advertisement effect of the new volume. These two variables are dummy variables for this advertisement effect, equal to one after the most recent volumes of the comic are published. Newpub1 is equal to one after for the second recent volume, and newpub2 is for the most recent volume. These variables were applied to ongoing comics only because completed comics do not have newly published volumes.

anime$_{it}$: When the comic becomes an animated TV series, the sale of the comic book gains momentum. Variable *anime*$_{it}$ is equal to one if the TV animation series of the comic i was on air at the month t.

e_promotion$_{it}$: As for the electronic version of the comic, sales heavily depend on promotion by the electronic shop sites. If the electronic shops put the comic's advertisement on the banner of the top page, the comic sells well. Also, electronic shops sometimes set several comics' prices down by 50% or more since the marginal cost of supplying the electronic files is almost zero, resulting in a boost in sales.

$E\_promotion_{it}$ is a dummy variable equal to one if the comic was advertised or discounted on the electronic shopping sites.

$Monthly\ dummies_t$: Monthly dummies were included to remove seasonal variation and temporary monthly common shocks.

Price was not included as a control variable because book price is fixed by publishers in Japan. Retail book stores in Japan are not allowed to change book retail prices; thus, book prices did not change in the research period, with the exception of promotion in electronic shops. The expected sign was negative for "time" and "ltrend", and positive for "anime", "newpub1", "newpub2", and "e_promotion".

Descriptive statistics are shown in Table 3. Mean sales were 1,165 books/month, which was far different from their median 70 books/month, indicating that distribution of sales was skewed (skewness = 41.75). Log-transformation of sales made the mean and median almost the same number, 4.2, and the distribution became not skewed (skewness = 0.106). The mean number of reach sites was 7.163, with a distribution similar to normal distribution except for many zero points. One fourth of the number of sites were zero because it takes some time for new links to appear on the reach sites after a new volume is published. In that case, the number of sites continues to be zero until pirated files are uploaded and links created.

Table 3 Descriptive Statistics

Variable	unit	Obs	Mean	Median	Std Dev	Min	Max
sale	#	31167	1165	70	11353	0	999826
log(sale)		31167	4.202	4.263	2.321	0	13.815
site	#	26880	7.163	8	5.083	0	16
time		40320	6.5	6.5	3.452	1	12
ltrend		14021	0.977	1.099	0.920	0	2.565
newpub1	dummy	20160	0.500	1	0.500	0	1
newpub2	dummy	20160	0.322	0	0.467	0	1
anime	dummy	31167	0.041	0	0.198	0	1
e_promotion	dummy	31167	0.036	0	0.186	0	1

VI Panel regression

Table 4 shows the result of the panel regression of model (1). Column (1) in Table 2 uses all comics. Before examining the effect of reach sites, let us see other control variables to check the validity of the estimation. Coefficients of time and log(trend) were negative, as expected, which indicates that initial sales rapidly decreased at a rate of power $t^{-1.242}$ for newly published volumes and gradually declined at 3.32% per

month after the initial boosts. Other exogenous variables, newpub1, newpub2, anime, and e_promotion were positive, as expected. Most recently published volumes of ongoing comics increased sales of former volumes by 29.1%. TV animation series boosted comic sales by 87.8%, and electronic promotion increased sales by 51.7%. These estimated figures are plausible from the view of Japanese comic business according to my interview survey to the four publishers, which supports the validity of this estimation.

The coefficient of reach sites is negative in column (1), indicating that piracy reduced legal sales in the research period. However, when we divide the comics into ongoing comics (column (2)) and completed comics (column (3)), the coefficients show opposite signs, though one of them is insignificant. As for ongoing comics, a 1% increase of reach sites *reduced* sales by 0.102%, whereas the effect on completed comics was not significant and slightly positive. In other words, the displacement effect was dominant in the case of ongoing comics, though some kind of advertisement effect worked in the case of completed comics and cancelled the displacement effect.

Table 4　Basic case: ongoing comics vs completed comics

	(1) All comics	(2) Ongoing comics	(3) Completed comics
VARIABLES	Log(sale)	Log(sale)	Log(sale)
log(site+1)	-0.0891***	-0.102***	0.132
	(0.0287)	(0.0305)	(0.104)
time	-0.0332***	-0.0178***	-0.0453***
	(0.00314)	(0.00566)	(0.00383)
log(trend)	-1.242***	-1.294***	
	(0.0637)	(0.0624)	
newpub1	0.0286	-0.0185	
	(0.0339)	(0.0369)	
newpub2	0.291***	0.255***	
	(0.0263)	(0.0307)	
anime	0.878***	0.879***	1.059***
	(0.0673)	(0.0705)	(0.224)
e_promotion	0.517***	0.483***	0.638***
	(0.0323)	(0.0364)	(0.0753)
Constant	4.976***	5.856***	3.152***
	(0.0625)	(0.0729)	(0.211)
Monthly dummies	included	included	included
Observations	24,562	15,754	7,356
R-squared	0.205	0.232	0.115
Number of group	3,372	2,256	932

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The reason that advertisement effect appears for completed comics is supposed to be that readers do not remember the completed comics because they ended several years ago, and publishers do not advertise them. Piracy "reminds" consumers of the existence of the completed comics of years ago, and handfuls of downloaders of pirated files find them interesting. A small percentage of them, or those informed by word-of-mouth, purchase the comic. In contrast, ongoing comics are well-known because of serial publication in weekly comic magazines and publishers' advertisements in the book store. Therefore, the "remind effect" of piracy has relatively little effect on ongoing comics.

To test the validity of this “remind effect” hypothesis, we divided the data into the first and middle volumes and recent two volumes. As for ongoing comics, publishers and book stores advertise recent volumes in-store more than first and middle volumes. Thus, the benefit of remind effect is minimal in the case of recent volumes of ongoing comics since they are already advertised very well. Relatively speaking, then, the recent volumes suffer more from piracy than the first and middle volumes because recent volumes obtain less benefit from the remind effect. Regarding completed comics, people who notice the comic by remind effect and have interest in it will start to buy the comics from the first volume, not from the most recent volume, and some readers may stop reading before reaching the final volume.[2] Therefore, the remind effect works better for the first and middle volumes than recent volumes in the case of completed comics. In summary, if the remind effect is working, recent volumes suffer from piracy more for ongoing comics, and first and middle volumes benefit more from piracy when it comes to completed comics.

To investigate this, we ran regressions of columns (2) and (3) of Table 4, dividing the data into recent two volumes and first and middle volumes. Table 5 shows the results. In the case of ongoing comics, the elasticity of the recent two volumes (-0.0967) was two times higher than the first and middle volumes (-0.0397), showing that recent volumes suffer more from piracy than the first and middle volumes. Regarding the completed comics, the coefficient of the first and middle volumes was significant and very high (0.522). These results are consistent with the expected pattern discussed in the previous paragraph, which supports the remind effect hypothesis.

[2] It is common fact in the book market that when the book consists of two volumes volume 1 sells more than volume 2. The reason is that there are some readers who buy volume 1 and quit buying volume 2, whereas there is no reader who would buy volume 2 only.

Table 5　Comparison of recent two volumes and first & middle volumes

	(1) Ongoing	(2) Ongoing	(3) Completed	(4) Completed
	Recent two volumes	First & middle volumes	Recent two volumes	First & middle volumes
VARIABLES	Log(sale)	Log(sale)	Log(sale)	Log(sale)
log(site+1)	-0.0967**	-0.0397	-0.0401	0.522***
	(0.0375)	(0.0246)	(0.110)	(0.115)
time		-0.0684***	-0.0529***	-0.0377***
		(0.00466)	(0.00592)	(0.00481)
log(trend)	-1.578***			
	(0.0718)			
newpub1		0.168***		
		(0.0360)		
newpub2	0.357***	0.386***		
	(0.0630)	(0.0274)		
anime	0.643***	0.981***	1.167***	0.953***
	(0.155)	(0.0655)	(0.317)	(0.314)
e_promotion	0.354***	0.531***	0.609***	0.664***
	(0.0660)	(0.0428)	(0.107)	(0.106)
Constant	7.836***	4.670***	3.678***	2.179***
	(0.119)	(0.0589)	(0.221)	(0.247)
Monthly dummies	included	included		
Observations	6,478	9,276	3,663	3,693
R-squared	0.269	0.214	0.123	0.122
Number of group	1,088	1,168	467	465

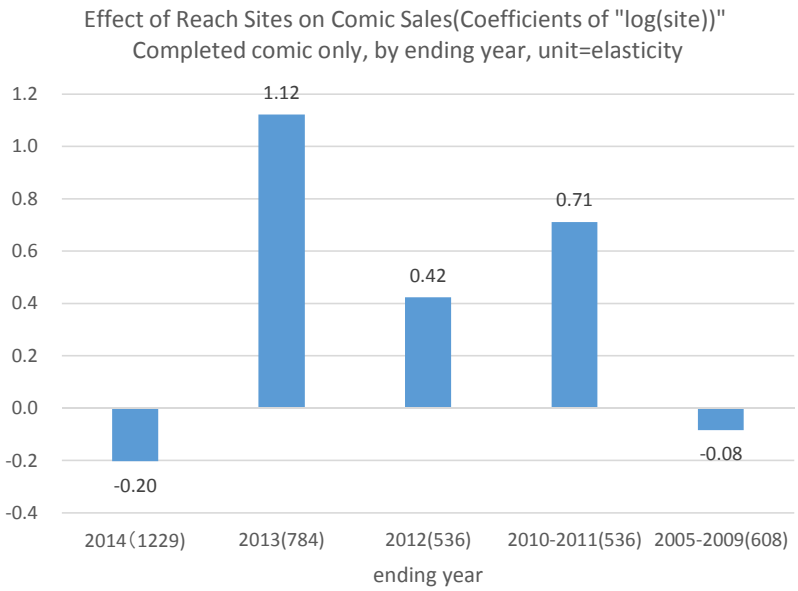
Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

How far into the past does remind effect extend? To answer to this question, we divided the completed comics by ending year and applied the same panel regression. Figure 2 shows the coefficients of log(site+1) depending on the ending year of comics, which is shown along the horizontal line. Regarding periods before 2010, several years were grouped into one category in order to keep the sample size at a similar level (sample size is in parentheses). Figure 2 shows that the elasticities of reach sites were positive for comics that ended in 2010-2013, thus suggesting that the remind effect was effective for this period.

Interestingly, elasticities were almost zero in the near past (in 2014) and long ago (during 2005-2009). The fact that remind effect was ineffective for 2014 can be explained as follows: consumers remembered the comic well because it had only been released one year prior. Note that this paper's research period is 2015-2016. If

consumers remember the comic well, there is no need for pirated files to stimulate consumers' memories to encourage them to purchase the comic. The reason that remind effect did not work before 2009 is uncertain. One of the possible explanations is that consumers lost interest in the comics of that time period because the style or quality of comics had changed.

Figure 2

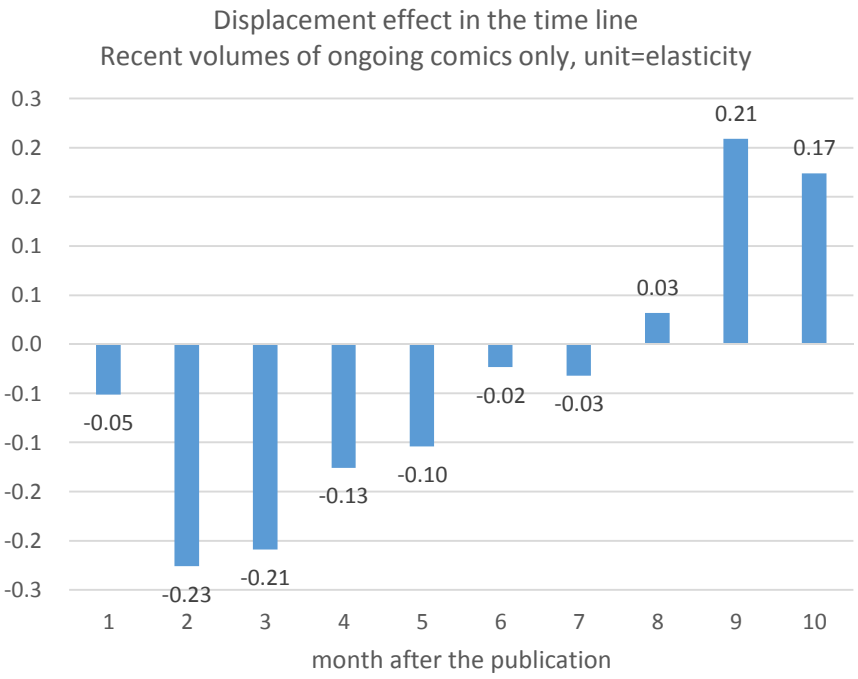


The displacement effect of ongoing comics may change over time. To demonstrate this, we broke down the coefficient of $\log(site_{it}+1)$ of the regression (1) in Table 3 by replacing the variable with cross term with dummy variables, $Dj*\log(site_{it}+1)$, where Dj is equal to one if it is j months after the publication month. In other words, we replaced $b*\log(site_{it}+1)$ with $b^j* Dj*\log(site_{it}+1)$. Coefficients of this cross-term, b^j , represent the timing and strength of the displacement effect.

Figure 3 shows the result. The vertical line represents elasticity, coefficient of $D_j*\log(site+1)$, and the horizontal line is j, months after the publication. This graph shows that the displacement effect worked mainly during the initial five months. Note that this graph shows relative strength, assuming that elasticity was zero when t=0, just published month; hence, the positive elasticities at t=9 and 10 do not necessarily mean displacement effect was reversed. We can safely say that the displacement effect was concentrated during the first several months and declined rapidly after the initial periods. This rapid decline of displacement effect of newly published volumes is also

the reason why the remind effect becomes dominant in the case of completed comics.

Figure 3

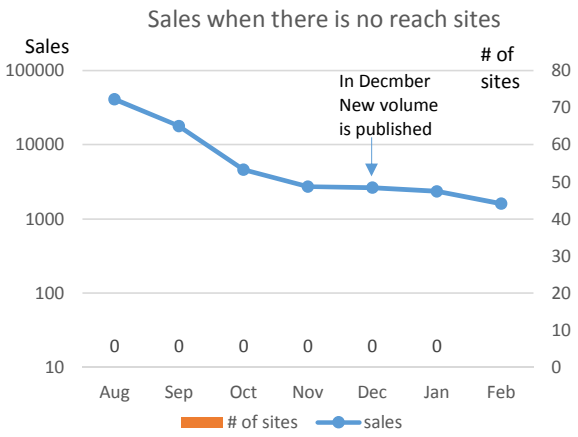


Since studies on the effects of piracy in the book industry are seldom seen, we will show a typical example for illustrative purposes. Figure 4 shows five cases. On the horizontal axis is month, and vertical lines are sales for the left axis and number of reach sites for the right axis. Figure 4(a) shows the case of newly published volumes when there was no reach site with links to this comic. Without reach sites, sales decreased continuously after publication. A slight increase in December was caused by another new publication of this comic.

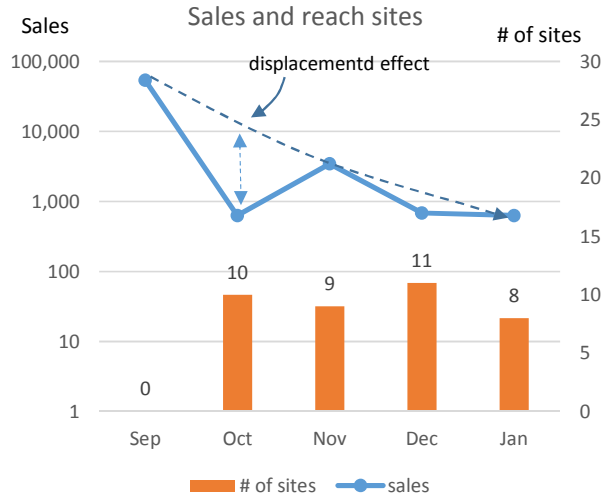
Figure 4(b) and (c) are the cases where reach sites listed links one month after the publication. The volumes were published in September, but there were no links to pirated files on the reach site at the month of publication. One month later, ten or eleven reach sites listed the link on the sites, and then the sales declined sharply, deviating from the trend. This deviation — that is, the sharp downshift of sales in October (and November) — is interpreted to be displacement effect.

Figure 4

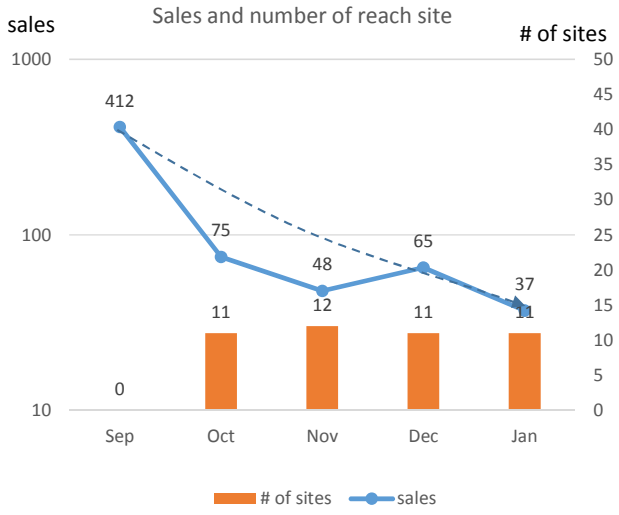
(a) no piracy case: a comic for girls, second recent volume



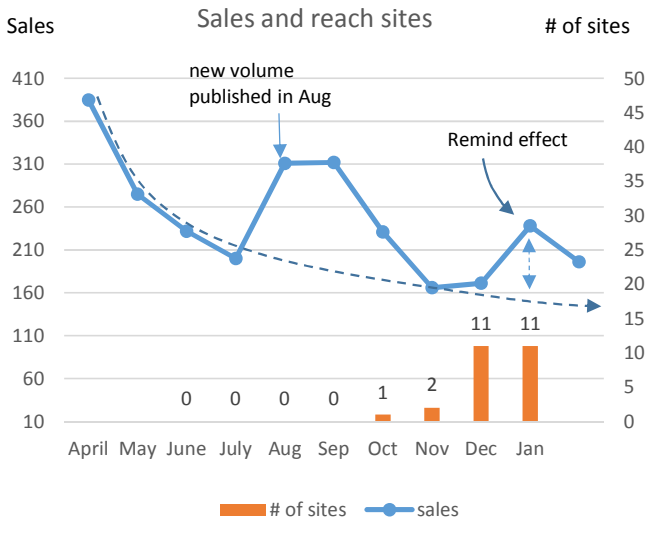
(b) Displacement effect case1
a comic for girls, most recent volume volume



(c) Displacement effect case2
a comic for boys, most recent



(d) Remind effect case 1
a comic for men, the first volume



(e) Remind effect case 2
a comic for men, middle volume

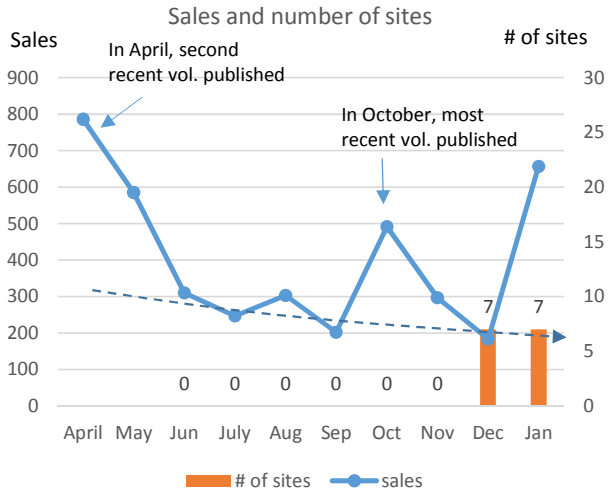


Figure 4(d) shows the case when pirated files appeared on the reach site long after publication. This comic started in 2012, and this graph represents the sales of the first volume. Since three years passed after the publication, sales were low and gradually decreasing (Note that the left axis is scaled not by log term but number term). A sudden rise of sales in August was caused by the publication of a new volume. The number of reach sites that had the link to this volume was zero until September and began to increase slightly in October and November. The number of reach sites jumped to 11 in December, which is interpreted to have raised the sales in January, deviating from the trend. This deviation from the trend is supposed to be the result of remind effect. Figure4(d) is a case of a middle volume, the sales of which jumped after piracy files appeared in December. We can interpret this jump as remind effect: consumers read this comic via piracy files and found it interesting, and some of them or others who heard about it through word-of-mouth purchased it.

VII DID regression

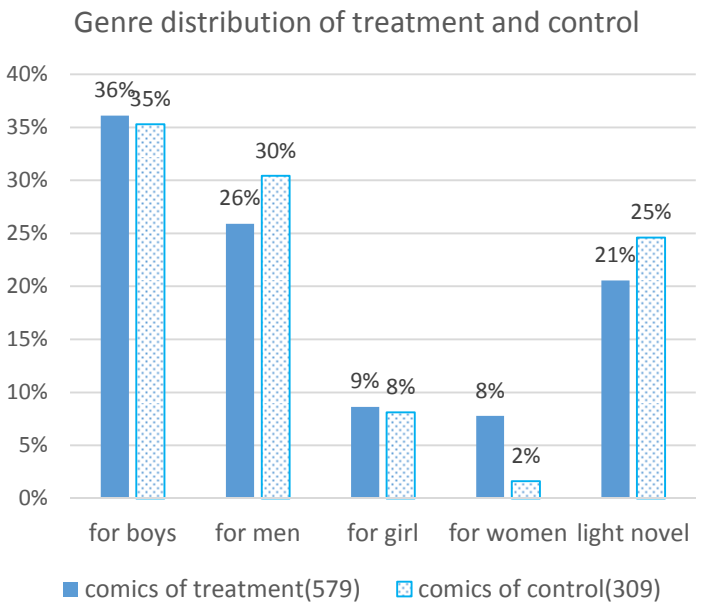
Fixed effect cannot cope with endogeneity if the comic's popularity changes during the research periods and site administrators notice it quickly. If a comic suddenly becomes popular in the middle of the research period, both the sales and the number of reach sites (or links to pirated files) will increase simultaneously, which make the number of sites endogenous. Difference-in-difference analysis can cope with this endogeneity problem because it does not use number of sites as an explanatory variable.

We applied the DID model using the massive deletion project, which started in the middle of research periods as a natural experiment to measure the net effect of deleting pirated files on comic sales. A total of 551 comics were targeted by this deletion, and these made up the treatment group. For control data, we chose 289 comics that were not targets of the massive deletion project.

For difference-in-difference regression, treatment data and control data should have the same characteristics except for the exogenous shock. Figure 5 shows the genre distribution of the target comics (treatment) and non-target comics (control). Japanese comics are ordinarily categorized into four genres: for boys, for men, for girls, and for women. Light novel refers to comic-like novels in Japan. As Figure 5 shows, the distribution pattern was similar between the treatment and control groups. Starting year and ending year of the comic were also similar, as shown in the attached sheet in

Figure 5. However, sales level was somewhat different; the table shows that total annual sales of the first volume were 5,791 books for the treatment group and 3,220 books for the control group. This is because publishers chose big titles for this massive deletion project. This bias was considered later in the regression analysis.

Figure 5



	comics of treatment (579)	comics of control(309)
starting year	2007.6	2007.8
ending year	2014.1	2013.3
total sales of the first volume	5791	3220

Unfortunately, the actual effect of this massive deletion project was limited since deleted files are quickly restored to cyber locker sites, and reach sites rewrite the links. It is often the case that policy intervention against piracy does not affect the level of piracy. (Ushiyama, 2009; Bhattacharjee, Gopal, Lertwachara, & Marsden, 2006). Thus, before estimating the effect of deletion, we had to examine the actual effect of this massive deletion project. According to difference-in-difference approach,

$$Site_{it} = a_i + b0*DafterJuly_t + b*Deletion_i*DafterJuly_t + c*Z_{it} + e_{it} \quad (3)$$

Dependent variable $Site_{it}$ represents the number of reach sites of comic i at month t. $DafterJuly_t$ is equal to one if month t is after July, and $Deletion_i$ is equal to one if comic i is a target of the massive deletion project. The cross-term coefficient of $DafterJuly_t$ and $Deletion_i$ is an effect of the massive deletion project. Control variables Z_{it} are general time trends (time) and nonlinear trends ln(trend) for newly published volumes.

Since some cyber locker sites set expiration dates to the files on their server, pirated files tend to disappear gradually even without the deletion request from copyright holders. The time trend variable was introduced to represent this general tendency. Regarding newly published volumes, the number of sites is usually initially zero and increases nonlinearly. Ln(trend) represents this nonlinear trend.

Table 6 shows the result. As column (2) shows, the cross-term coefficient of completed comics, -0.619, was small but significant. Therefore, we can say the massive deletion project reduced the number of reach sites successfully for completed comics. Regarding ongoing comics, however, as column (1) shows, the cross-term coefficient was not significant. Ongoing comics are more popular than completed comics because their episodes are released weekly in serial comic magazines and often advertised by publishers. Thus, it is natural that pirated files of ongoing comics would be restored to piracy sites more quickly than already completed comics. Quick and frequent re-uploading offset the effect of the massive deletion project. Since the deletion of files is uncertain regarding ongoing comics, we focus on the case of completed comics hereafter.

Table 6　Effect of Deletion Project

	(1) Ongoing comics	(2) Completed comics
VARIABLES	Site	Site
Djuly	0.0928	0.466***
	(0.133)	(0.107)
Deletion*Djuly	-0.141	-0.619***
	(0.176)	(0.139)
time	-0.0713***	-0.0166
	(0.0179)	(0.0151)
log(trend)	3.691***	
	(0.249)	
Constant	6.109***	6.924***
	(0.170)	(0.105)
Observations	9,369	4,480
R-squared	0.133	0.009
Number of group	1,345	560

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7 shows the result of difference-in-difference regression. The first two columns represent the basic case. Column (1) is the case of ongoing comics, and column (2) is the case of completed comics. Regarding ongoing comics, the cross-term coefficient of DafterJuly and Deletion was positive (0.703) but not significant. This insignificance corresponds to the insignificance of the deletion effect seen in Table 6. The positive sign suggests that sales might increase with deletion, which is consistent with the result of panel regression seen in section VI.

As for completed comics, the coefficient is significant and -0.830, meaning that the massive deletion project reduced the sale of completed comics by 8.3%. This result implies that pirated files increase the sales of completed comics, which is also consistent with the result of panel regression, confirming the validity of the "remind effect".

Table 7　Difference-in-Difference model

	(1) Ongoing comics	(2) Completed comics	(3) Ongoing comics	(4) Completed comics	(5) Ongoing comics	(6) Completed comics
	Basic case	Basic case	With monthly dummies	With monthly dummies	100<sale <5000	100<sale <5000
VARIABLES	Log(sale)	Log(sale)	Log(sale)	Log(sale)	Log(sale)	Log(sale)
DafterJuly	0.0837*	0.0480			-0.0547	0.0562
	(0.0439)	(0.0357)			(0.0456)	(0.0405)
Deletion*DafterJuly	0.0703	-0.0830**	0.0672	-0.0800**	0.154***	-0.111**
	(0.0536)	(0.0398)	(0.0537)	(0.0398)	(0.0559)	(0.0469)
time	-0.0584***	-0.0600***	-0.0654***	-0.0653***	-0.0609***	-0.0684***
	(0.00499)	(0.00373)	(0.00630)	(0.00515)	(0.00583)	(0.00544)
log(trend)	-0.722***		-0.724***		-0.410***	
	(0.0569)		(0.0565)		(0.0720)	
newpub1	-0.00776		-0.0294		0.0412	
	(0.0318)		(0.0325)		(0.0384)	
newpub2	0.112***		0.121***		0.102***	
	(0.0312)		(0.0318)		(0.0371)	
anime	0.976***	0.862***	0.976***	0.859***	0.864***	0.604***
	(0.0751)	(0.187)	(0.0750)	(0.188)	(0.106)	(0.179)
e_promotion	0.508***	0.689***	0.408***	0.656***	0.671***	0.875***
	(0.0362)	(0.0747)	(0.0370)	(0.0762)	(0.0412)	(0.104)
Constant	5.390***	3.446***	5.434***	3.488***	4.689***	4.301***
	(0.0425)	(0.0203)	(0.0610)	(0.0458)	(0.0499)	(0.0278)
Monthly dummies			included	included		
Observations	21,074	10,093	21,074	10,093	12,787	5,493
R-squared	0.166	0.141	0.175	0.146	0.158	0.191
Number of group	2,279	932	2,279	932	1,298	517

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Columns (3)–(6) in Table 7 show a verification of the robustness of this result. Monthly dummies are included in columns (3) and (4) to remove temporary shocks. Columns (5) and (6) limit the data to comics whose sales (measured by annual total sales of the first volume) were between 100 and 5000 in order to consider the difference of total sales level between the treatment and control groups shown in Figure 5. In both cases, cross-term coefficients were positive for ongoing comics and negative for completed comics. The results of the basic case were maintained

Regarding completed comics, we checked robustness further. We focused on completed comics because the number of sites were significantly decreased by the massive deletion project, and the fact that piracy increases legal sales is a controversial topic in copyright issues. The robustness of regression (2) is verified in Table 6, a basic case of completed comics.

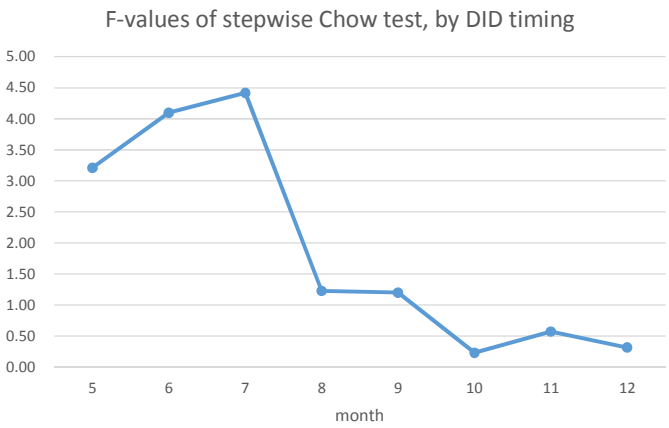
Although a placebo test is as common as a robustness check for difference-in-difference approach, we could not apply it here because the sample size in the time series was too small (12 months). Instead, we divided the sample into subgroups and examined the correlation between deletion and sales. If correlation among subgroups was consistent, this suggests that the result was not caused by specific small data, but reflects an overall tendency. This subgrouping check is the approach adopted by Danaher and Smith (2014) to investigate the effects of the shutdown of major cyber locker site, MegaUpload, on DVD sales. Since the penetration of MegaUpload was different among countries, they correlated the countries' sales change after the shutdown with the penetration ratio of MegaUpload and found a positive correlation. In other words, they divided the worldwide consumers into country subgroups and checked the correlation between the degree of intervention (penetration of MegaUpload) and outcome (DVD sales) in the difference-in-difference model. We applied this approach as well.

Before making subgroups, we applied an ordinary stepwise Chow test to see whether July was the proper timing for applying DID regression. The null hypothesis was that coefficients of both DafterJuly and DafterJuly*Deletion were zero in the regression (2) of Table 7. Figure 6(a) shows the F-value of Chow test changing the hypothetical starting point of massive deletion from May to December. Clearly, July is the most probable point at which structural change occurred, which suggests the July was the proper point to apply difference-in-difference regression.

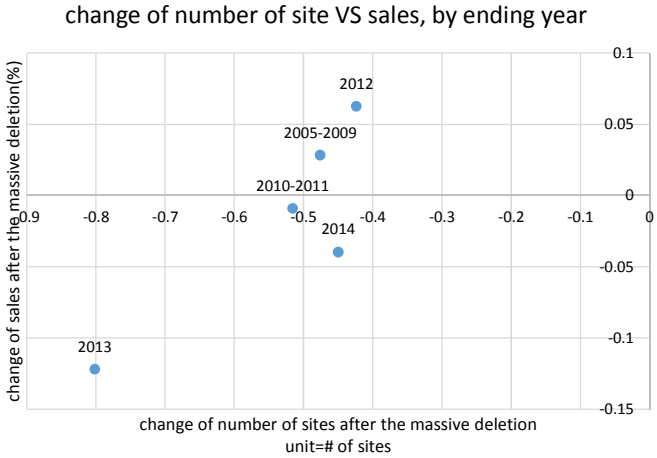
Firstly, we divided the completed comics by ending year into five groups and ran DID regressions of sites and sales. We then checked correlation of cross-term coefficients. Let b_{site}^{y} be coefficients of cross term, DafterJuly*Deletion, of the regression of the reach site (model (3)) and let $b_{log(sale)}^{y}$ be cross-term coefficients of the regression of the sales (model(2)) when the data is limited to comics ending in year y. Both terms measure changes caused by the massive deletion project, and b_{site}^{y} represents the change in number of comics sites ending in year y, and $b_{log(sale)}^{y}$ represents the percentage change of sales of comics ending in year y. If piracy reduces legal sales, b_{site}^{y} and $b_{log(sale)}^{y}$ have a negative correlation in terms of y because, when more sites are reduced for the comic group y, then more sales increase for the same comic group y. To the contrary, if piracy increases legal sales, they correlate positively. Figure 6(b) shows the result. The horizontal line is b_{site}^{y}, change in number of sites, and the vertical line is $b_{log(sale)}^{y}$, change in sales. In our findings, we observed weak positive correlation, which means that comics deleted more are sold less in terms of ending year grouping. This result suggests that piracy increases legal sales.

Figure 6 Robustness check of remind effect of completed comics

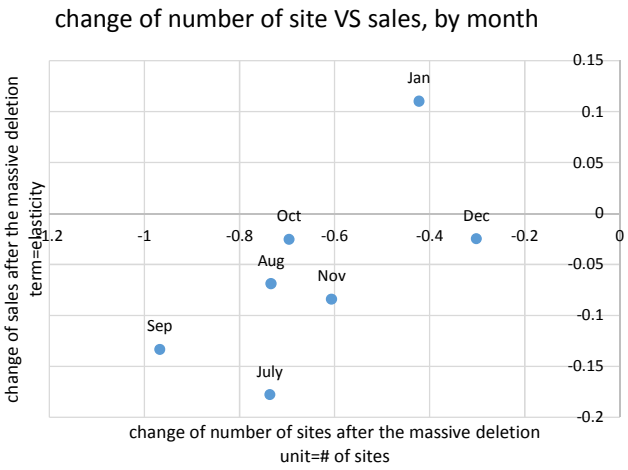
(a)



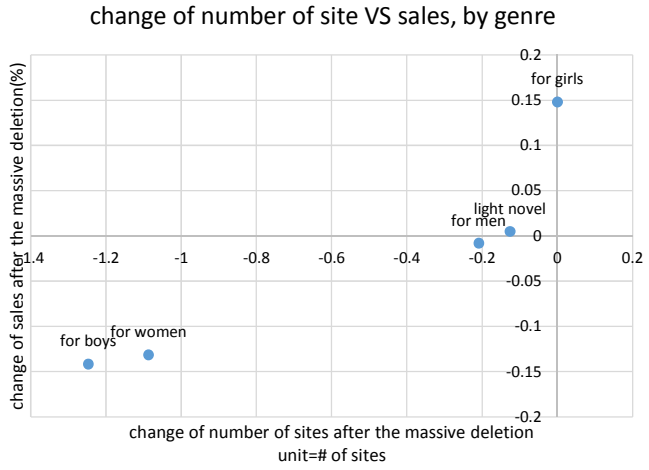
(b)



(c)



(d)



Secondly, we broke down the cross-term coefficient into monthly terms by replacing DafterJulyt with $Djuly_t$, $Daug_t$, $Dsep_t$, …, $Ddec_t$, which are equal to one if t is equal to the current month, as in Figure 3. Since the performance of the massive deletion project may change month to month, the degree of reduction of reach sites also may change over time. If the sales of comics positively correlate with the change in site reduction by monthly terms, this is supporting evidence of the result. Thus, we replaced $DafterJuly_t$ in the site regression (model (3)) and sales regression (model(2)) with a set of $Djuly_t$, $Daug_t$, $Dsep_t$,…$Ddec_t$, and plotted the cross-term coefficients, $b_{site}{}^j$

and $b_{log(sale)^j}$. The results are shown Figure 6(c), indicating positive correlation, which means that sales decreased more in the month when reach sites decreased more. This result again suggests piracy increases the legal sale of completed comics.

Thirdly, we let the coefficients change depending on comic genres. As previously mentioned, the Japanese comic industry has five genres: for boys, for men, for girls, for women, and light novels. Genre dummies Dg were introduced to correspond to these genres, and cross-terms of models (2) and (3) were replaced with $\Sigma b_{log(sale)^g}$ $Dg*DafterJuly_t*Deletion_i$, and $\Sigma b_{site^g}*Dg*DafterJuly_t*Deletion_i$, respectively. Coefficients $b_{log(sale)^g}$ represent sales change of comics in genre g, and b_{site^g} represent site change of comics in genre g. Correlation of these coefficients is shown in Figure 6(d), which indicates positive correlation again. Therefore, if comics of a genre are deleted more than other genres, the sales of comics of the genre decrease more than other genres. This result suggests that piracy increases legal sales with regard to completed comics.

In Figure 6(b)(c)(d), we try to break down cross-term coefficients of difference-in-difference regression by subgrouping data with the purpose of examining the consistency of degree of policy intervention (deletion) and its outcome (sales). All graphs show that correlation is positive; thus, they are consistent with the results in Tables 2 and Table 5. No matter which subgrouping is used, either by ending year, by timing, or by genre, more deletion leads to less sales, which means piracy stimulates sales as far as completed comics are concerned.

VIII Summary and Discussion

In this paper, we examined the effect of piracy on the comic book market in Japan using direct measurement of product level piracy ratio and a massive deletion project as a natural experiment. We found the piracy effect was heterogeneous depending on whether the comic series was ongoing or completed. Piracy decreased the legitimate sales of ongoing comics but stimulated legitimate sales of completed comics: displacement effect was dominant for ongoing content, and advertisement effect was dominant for completed content. We can interpret this result as an indication that the piracy reminds consumers of past comics that are no longer promoted by publishers.

An interesting question for further research is whether the total effect of piracy is negative or positive. Unfortunately, we were unable to calculate this because the sales share of ongoing comics and completed comics is unknown. The loss of displacement effect is far larger than the gain of remind effect per individual book base, since

ongoing comics usually sell over ten thousand copies in the newly published month, while most completed comics series sell only several hundred in a month. However, volumes of completed comics far outnumber newly published volumes of ongoing series. Which effect is dominant depends on the total sales share of ongoing and completed comics, and this figure is not available.

An important point we should stress here is the policy implication of the heterogeneity of piracy effect. If the effect of piracy is heterogeneous, it is not the best solution to shut down the piracy sites but to delete harmful piracy files selectively if possible. In this case, deleting piracy files of ongoing comics only is the first best strategy for publishers regardless of whether the total effect is positive or negative, because the availability of piracy files of completed comics is beneficial to both publishers and consumers.

Heterogeneity is not limited to the book industry. In 2007, Blackburn showed that file sharing decreased CD sales of big artists, whereas it increased CD sales of unknown minor artists. Bhattacharjee et al. (2007) reported that file sharing reduced survival time in the Billboard CD album ranking for low-ranked albums but did not hurt top-ranked albums. Mixed results in the literature of piracy studies so far may be caused by the heterogeneity of products. Product level research that takes this heterogeneity into consideration may shed some light on the effects of piracy. This paper's contribution is the addition of new dimension to the heterogeneity — present content and past content.

References

Adermon,Adrian, Che-Yuan Liang, 2014,"Piracy and music sales: The effects of an anti-piracy law", Journal of Economic Behavior & Organization 105　pp.90–106

Arnold, Michael, Eric Darmon, Sylvain Dejean and Thierry Penard 2014 "Graduated Response Policy and the Behavior of Digital Pirates: Evidence from the French Three-Strike (Hadopi) Law" Working paper NO. 2014-07m, Alfred Lerner College of Business & Economics, University of Delaware

Bhattacharjee, Sudip, Ram D. Gopal, Kaveepan Lertwachara, and James R. Marsden, 2006 "Impact of Legal Threats on Online Music Sharing Activity: An Analysis of Music Industry Legal Actions," The Journal of Law and Economics 49, no. 1 (April 2006): 91-114.

Bhattacharjee,Sudip,Ram Gopal, Kaveepan Lertwachara, James Marsden, Rahul Telang, 2007, "The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts", Management Science 53(9), pp1359-1374

Blackburn, David, 2007, "The Heterogenous Effects of Copying: The Case of Recorded Music", Working Paper, Harvard University, Cambridge.

Danaher, Brett, Michael D. Smith 2014 "Gone in 60 Seconds: The Impact Movie Sales," International Journal of Industrial Organization 33 (2014) 1–8

Danaher, Brett, Michael Smith, Rahul Telang, and Siwen Chen 2014 "The Effect of Graduated Response Anti-piracy Laws on Music Sales: Evidence from an Event Study in France," Journal of Industrial Economics 62(3) 541-553

Danaher, Brett, Michael Smith, Rahul Telang, 2013, "Piracy and Copyright Enforcement Mechanisms"　NBER Working paper 19150

Hardy, Wojciech, Michal Krawczyk and Joanna Tyrowicz, 2014, "Internet Piracy and book sales: a field experiment," Working papers NO.23, Faculty of Economic Science, University of Warsaw

Hilton, J., Wiley, D., 2011. "Free e-Books and print sales," Journal of Electronic Publishing 14 (1).
http://hdl.handle.net/2027/spo.3336451.0014.109

Hu, Y. J., Smith, M. D., 2013. The impact of ebook distribution on print sales: Analysis of a natural experiment. SSRN Scholarly Paper ID 1966115, Social Science Research Network.

Liebowitz, Stan J., 2006, "File Sharing: Creative Destruction or Just Plain Destruction?" Journal of Law and Economics 49(1) pp. 1-28

Liebowitz, S. (2008), "Testing File Sharing's Impact on Music Album Sales in Cities", Management Science 54, 852–9.

Martikainen, E., Jan. 2011. Does file-sharing reduce DVD sales? SSRN Scholarly Paper ID 1742443, Social Science Research Network.

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

Oberholzer-Gee, Felix; Strumpf, Koleman.,2010,"File Sharing and Copyright", Innovation Policy and the Economy 10, pp. 19-55.

O'Leary, B., May 2009. *Impact of P2P and Free Distribution on Book Sales*. O'Reilly Media, Inc.

Peitz, Martin and Patrick Waelbroeck,2008, "Piracy of Digital Products: A Critical Review of the Theoretical Literature," Recent Trends in the Economics of Copyright, 2008, pp. 125-52, The International Library of Critical Writings in Economics series ,Edward Elgar, Also available at SSRN: https://ssrn.com/abstract=889128 or http://dx.doi.org/10.2139/ssrn.889128

Smith,M.D.,R.Telang.2009. "Competing with free:The impact of movie broadcasts on DVD sales and Internet piracy." Management Information Systems Quarterly 33(2)321–338.

Takeyama, Lisa N..1994," The Welfare Implications of Unauthorized Reproduction of Intellectual Property in the Presence of Demand Network Externalities." Journal of Industrial Economics 42(2), pp. 155-66.

Tanaka, Tatsou. 2004. “Does File - Sharing Reduce CD Sales? A Case of Japan.” Paper prepared for Conference on IT Innovation, Hitotsubashi University, Tokyo.

Ushiyama, Atsushi, 2009, "Crime on the Incidence of Crime: A Comparison with Robbery," Japanese Economy, 36 (2), pp. 78-102

Zentner, A. (2005), “File Sharing and International Sales of Copyrighted Music: An Empirical Analysis with a Panel of Countries”, B. E. Journals in Economic Analysis and Policy: Topics in Economic Analysis and Policy 5(1), 1–15.