

**Is there an Access/Incentive Tradeoff?**  
**Copyright's impact on Book Sales, Prices, and**  
**Availability**

**Stan J. Liebowitz**  
**liebowit@utdallas.edu**

**School of Management**  
**University of Texas at Dallas**  
**Preliminary September 2022**

Abstract: This paper uses an unusually rich 21<sup>st</sup> century data set to compare two sets of vintage bestsellers from the early 20<sup>th</sup> century that, by a circuitous path of copyright law alterations, came to have different copyright treatments. The most striking result is that copyrighted vintage bestsellers sell almost four times as many copies as public domain vintage bestsellers and this result holds throughout the sales distribution. This result is contrary to the expectation that copyright would restrict sales by allowing the exercise of monopoly power, and instead points to factors such as post-creation investment by publishers as being much more important than previously recognized. These greater sales occur despite a price premium that we find for copyrighted works, which itself is of a size similar to royalty payments typically paid to authors. We also find that copyrighted titles are slightly more likely to be sold in the market than are works in the public domain. These results comport with a view that copyright is socially beneficial, that retroactive copyright extensions are socially advantageous, and that indefinitely renewable copyright is likely to be an optimal policy.

Copyright, by providing ownership over a creative work, is the system that market economies generally adopt to provide creators with an incentive to author new artistic works.<sup>1</sup> Copyright, and intellectual property in general, is typically thought to provide a balance, perhaps lopsided, between the positive inducement of producing new creative works and the negative restriction on the consumption of protected works, both brought about by allowing a creator to prevent unauthorized copying of the work (often called the copyright “monopoly”). This balance is often referred to as the “access/incentive” tradeoff.

Since copyright restricts competition within a title (but not between titles) it has typically been thought likely to decrease sales of a title while increasing the price, as would be expected from a textbook monopoly. The greater profits due to copyright is thought necessary for the publisher to be able to pay the author of the work, since unfettered competition within the market for copies of a particular title would be expected to drive the economic profit to zero, leaving no revenue with which to pay the author. Copyright, therefore, is expected to increase the number of works since many authors require payment to create works.

The focus of the current analysis is on the consequences of the copyright system after the work is created, specifically copyright’s impact on the sales of books, the availability of books, and the price of books. I take advantage of a natural experiment whereby, due to a quirk in the serial alterations of America’s copyright law, titles written prior to 1923 lost their copyrights while other titles, almost as old, were able to retain their copyrights.

One of the novelties of my analysis is the use of a data set (Nielsen BookScan) that contains information about the national sales of individual titles, whereas previous analyses, few though they may be, mostly use Bowker’s Books in Print (BiP) which contains no sales data, or, if sales data are used, they tend to come from a single retailer. Using BookScan sales data reveals the surprising finding that copyrighted (CR) titles sell many more copies than public domain (PD) titles, with sales three or four times larger. This noteworthy sales advantage exists throughout the sales distribution, including both high and low selling titles.

The greater sales of individual copyrighted works found in my examination, in contrast to the predicted lower monopoly sales, suggests that basic monopoly model may neglect

---

<sup>1</sup> Economists and others (e.g., Boldrin and Levine, 2008) commonly refer to copyright as providing a governmentally authorized monopoly to creators, although Kitch (2000) argues that copyright merely provides ownership, not monopoly, and Liebowitz (2016b) points out that ownership can always be viewed as a tautological monopoly though usually one with no monopoly power.

important aspects of copyright, implying that there may be benefits to copyright beyond just providing authors a greater financial incentive to produce new works.

One explanation for why sales quantities might be enhanced by copyright is simply that ownership, as embodied by copyright, allows a publisher to internalize all the returns from its marketing investments (Landes and Posner, 2003, Adilov and Waldman, 2013).<sup>2</sup> Without copyright, a publisher trying to invest in fostering market demand for a book title would likely have to share any returns with other publishers also selling the same book title, reducing the incentives to make such investments and thus reducing the overall sales of the title. Copyright's advantage in this case would be another example of the tragedy of the commons, where the lack of property rights for public domain works fails to provide the efficient incentives for investment, leading to inefficiency.<sup>3</sup>

The finding of higher sales for CR works opens the possibility that copyright might be socially beneficial even if it has no impact on the creation of new works per se. If so, the consequences of copyright would be quite different than the standard analysis. Instead of balancing the harm from monopoly against the benefit of new creations, the "monopoly" would increase sales of individual titles while also providing incentives for the production of new titles. Both production and consumption would be enhanced by copyright, and its economic impact would be clearly beneficial as far as the sales of copies of the work to the public.<sup>4</sup>

My analysis also finds that CR works are sold at higher prices than PD works, although the price difference is in the vicinity of royalty rates that publishers pay authors, making the copyright premium somewhat lower than has been found by other authors (Heald 2008, Reimers 2019, Li, MacGarvie, and Moser 2018). Additionally, my analysis finds that copyrighted titles are slightly more likely to be sold than are public domain titles, in contrast to comparisons based on BiP measures of "in-print" availability.

---

<sup>2</sup> The term "marketing" here is meant to convey prosaic activities such as getting books into bookstores, on school systems' reading lists, getting mentions in popular articles, direct advertising, and so forth.

<sup>3</sup> Artistic works, being non-rivalrous, cannot be overused. But the resources used for marketing are rivalrous and can be underused due to free riding in the creative commons of public domain.

<sup>4</sup> This analysis ignores copyright's impact on the creators of follow-on works to obtain the permission of the copyright owner. Thus, copyright could reduce the production of new follow-on works, and since the elements of the original work that might be "borrowed" are nonrivalrous, any exclusion would be inefficient. Because it is unlikely that competing creators of follow-on works could engage in arbitrage, the copyright owner should be able to engage in price discrimination and price discrimination can cure imperfections from market production of nonrivalrous goods. Nevertheless, this is a complex topic best handled at another time and I am willing to stipulate that my analysis here is limited to the primary market for copies of the work sold to consumers.

When performing this analysis, a specific nomenclature is required to keep track of the units of observation. Individual creations, such as Steinbeck's *Of Mice and Men*, are referred to as titles. There can be multiple variations of a single title, such as hardcover or paperback versions, or versions from different publishers if the title is not copyrighted. The various versions of a title are called "editions" and there can be dozens or even hundreds of contemporaneous editions of a single title, particularly for popular public domain titles.

## I. Copyright and the Public Domain

Copyright owners are granted the exclusive right to make reproductions of their work (title) and that right has been extended to various forms of 'reproduction' including public performance, derivative works, electronic transmissions to the public and so forth, but it does not provide any protection from independently created competing works.

Book publishing is the oldest industry relying on copyright. After Gutenberg invented the printing press in the mid-1400s, various government-granted printing monopolies existed in various countries, such as the monopoly given to the Stationer's Company (essentially a guild of printers) in England, chartered in 1557. The Statute of Anne in 1709, considered the first modern copyright law, provided a copyright term of 14 years, renewable for an equal second term.

The original American copyright law of 1790 was modelled on England's Statute of Anne and provided 14 years of protection for American authors, followed by another round of 14 years if the copyright owner renewed the protection. In 1831, the length of the first term was increased from 14 years to 28 years. In 1909 both terms were set to 28 years. In 1964 an additional 19 years was added to that second term for all then-current and future copyrighted (CR) works. Thus, a title first published in 1922 (and renewed in 1950) would have been expected to enter the public domain 75 years later (28+28+19) in 1997, and similarly, a title published in in 1923 would have been expected to enter the public domain in 1998.<sup>5</sup> In 1998, however, Congress added 20 years protection for works that were then currently under copyright. Hence, a title published in 1923 (or later) and then renewed, would have continued to be protected until 2018 whereas a title published in 1922 (or before) would have lost protection in 1997.

This quite arbitrary 1922/23 dividing line between titles with and without copyright provides a natural experiment after 1997 that we will use to identify the consequences of copyright protection on the price and sales of books. The logic is quite simple. Titles that

---

<sup>5</sup> Although there was an important change to copyright law in 1976 that altered the term of copyright for new works, it did not affect the copyright duration of the old works discussed here.

are still being published a century or so after being written have all proven the ability to retain the interest of generations of readers. Whether a title was written in 1922 or 1923, however, is nothing more than a random accident that has nothing to do with the titles themselves or their copyright status in the early 21<sup>st</sup> century. Those titles published in 1923 or later have, in years prior to 2018, received the “treatment” of copyright, whereas those written prior to that date were not so treated.

## II. Prior Literature

Until the last decade and a half, there were virtually no empirical studies examining the impact of copyright on book prices, sales, or availability. Two studies undertaken at about the same time, Heald (2008) and Liebowitz (2009), looked at how the 21<sup>st</sup> century prices of titles that were first published in the period surrounding the 1923 copyright cutoff, compared to one another. The evidence in those studies is based mainly on the BiP data set which provides information (e.g., edition year, price, pages, binding, publisher) about editions of book titles. If an edition is listed as “in-print” it is treated as available to the public.

Heald (2008), chose old titles from three samples. The main sample was based on 20 years of bestsellers around the 1923 cutoff, but he supplemented the sample with 40 titles written during that period which were not necessarily bestsellers at the time but were popular in recent decades. He used data taken from the 2006 online edition of BiP (and offline BiP editions from previous years and decades) to compare public domain and copyrighted titles. He was most interested in whether PD titles were more likely or less likely to be available to consumers than CR titles, but he also tried to determine whether there was a price difference between PD and CR titles. When a title had multiple editions he chose the lowest price among the various editions, and he compared the average of these “lowest price” variants for his sample of CR and PD titles. For his complete sample of 287 titles still in-print, he found that CR works had the same average price as PD works. For his supplemental sample of 40 recently popular durable titles (that need not have been bestsellers when published), however, he found that the lowest-priced editions of those titles covered by copyright had considerably higher prices than titles that had lost their copyright, with the copyright premium ranging from 41% to 81%. Because he used the lowest priced edition as the price for a title instead of some sort of average price of editions, and he did not control for number of pages or the type of book binding, it is not clear how useful his results are. A perhaps more compelling but less general result was his comparison of prices per page for Penguin Classic paperback titles, which implicitly controls for format, pages, and publisher. He found that the copyrighted editions were 56% more expensive per page.

Liebowitz (2009), in a preliminary version of this work, focused on the price differential that copyright provided. Using the hardcover 2004-5 BiP, the 1923 copyright cutoff, and

including titles from 1895-1940, he used a regression analysis with the list price as the dependent variable and independent variables such as number of pages, type of binding, genre, publisher type, copyright status and a sales proxy (Amazon ranking) that we now know to be unreliable (Liebowitz and Zentner, 2021). Like Heald, he found that copyright had no effect on price for the full sample. He also found that limiting results to major publishers did not alter his results.

A decade later Reimers (2019) used a similar methodology, comparing the prices of editions of former best-sellers published between 1910 and 1936 and using the 1923 copyright demarcation and a regression design. She describes this method as a regression discontinuity. She examined copyright's impact on the Amazon price for these editions, controlling for binding, edition age, and type of publisher. Using Amazon prices seems to be an unusual choice since the publisher doesn't set the price that retailers charge.<sup>6</sup> Another questionable decision was including both new and used books in her analysis, a consequential decision for her results as we will see below. Nevertheless, she concludes that CR editions are priced 27% higher overall (although, as discussed below, her inclusion of used books and her choice of explanatory variables makes her measurement questionable). Additionally, she finds no support for the suggestion that CR might increase a publishers' post-creation investment in a title, and that PD titles had a larger number of editions by a very large factor of seven. Finally, and the main focus of her analysis, was her conclusion that a retroactive copyright extension, such as the 1998 Copyright Act, was generally harmful to social welfare if the incentive impact of copyright were ignored.

Note that Reimers and Liebowitz both have data based on works that were successful long ago, based on bestseller lists. Heald has former bestsellers and currently popular durable (old) titles. This latter set of titles consists of classic works that are popular now, irrespective of how popular they were upon initial publication. This would tend to tilt his sample toward publications that speak better to modern audiences than to contemporaneous audiences, implying that his sample is likely to be less homogenous than the samples use by Liebowitz or Reimers.

Finally, using historical data, Li, MacGarvie, and Moser (2018) examine how an 1814 change in English copyright law altered the relative price of editions depending on whether authors died or survived during the initial CR period, with the latter group

---

<sup>6</sup> The determination of consumer prices came to a head in *United States v. Apple Inc.*, 952 F. Supp. 2d 638 (S.D.N.Y. 2013) whereby leading publishers had tried to switch to an "agency model" allowing the publishers to determine the price to customers and Apple would take a percentage of revenue, as it does in its app store. If Amazon did not switch to this model, the publishers threatened to withhold their books. The Court found this behavior to be an antitrust violation. Similarly, publishers do not control the price that bookstores charge to customers.

discovered as having a smaller increase in CR duration from the changed law. This is not a comparison of the price of titles with or without CR, but a comparison of prices based on the remaining copyright *length* of the title. Prices are affected if publishers price discriminate during the life of a copyrighted title by starting with a high initial price and lowering it as the remaining time under CR decreases. Li, MacGarvie, and Moser conclude that the greater extension of copyright for works of dead authors raised the price of those books by about 43% relative to the works with a shorter increased CR length. It is not clear whether these values are at all comparable with comparisons of CR vs PD in current markets since the market organization was vastly different in early nineteenth century England, with book customers being mainly for-profit lending libraries, not individuals.

### III. The book publishing industry

Book publishing can be decomposed into several major categories (Greco, 2005) with “trade” books (books purchased by individuals) comprising the largest share. Trade books are further classified as adult fiction, adult nonfiction, and juvenile. Both BookScan and Books in Print (BiP) provide information on new physical titles (versus used books) in each of these three categories. We ignore the used book market because copyright owners have no control over that market and generate no direct revenue from it.

There are many book publishers, with over 2200 listed in the 2004 BookScan data. Imprints, which are smaller publishing units, often components of major publishers, number almost 7500.<sup>7</sup> Nevertheless, the largest publishers tend to dominate the market. For the 2004-2016 period, the top 4 publishers generated more than 70% of revenue in the market for physical works of fiction and the top 20 publishers generated more than 97% of the revenue. Comparable values for imprints are 15% and 44%.

The business model of printing books has changed in some important ways over the last few decades. Offset printing, also known as lithography, has been the primary form of book printing during the last century, with the disadvantage of relatively high setup costs but the advantage of low variable costs. Print on demand (POD) digital technology is a newer printing method with lower fixed costs but higher variable costs (and essentially zero inventory costs because inventories are not needed). Book editions with low expected demand often use POD whereas editions with high expected demand usually use offset. Although another “printing” change in the 21<sup>st</sup> century has been the introduction of eBooks, which are not included in BookScan data, such books did not exist in 2004, the focal year of this study (because that is the year for which I have BiP data).

---

<sup>7</sup> These numbers overate the number of independent publishers because oftentimes two publishers under the same ownership are listed as separate in the BookScan data although I tried to merge leading publishers under single ownership when calculating the market share values.

POD has allowed for new business models where small publishers can have vast libraries of titles with no inventory and virtually no fixed printing costs since the book is not printed until a copy is ordered. Some of the on-demand publishers, such as *IndyPublish*, *Books on Demand*, *BiblioBazaar*, and *Kessinger*, often list many more editions and ISBNs than do more established publishers with much larger sales.<sup>8</sup>

For example, Publishers Weekly in 2010 reported:

...many in the industry were stunned to see an unfamiliar company name, BiblioBazaar, leading a surging new segment of “non-traditional” publishing stats with a whopping 272,930 titles produced in 2009--almost as many titles [as] the entire “traditional” publishing business cranked out last year. Could it be? Could one little-known company really produce so much volume?<sup>9</sup>

...“If by ‘produce’ you mean create a cover file that will print at multiple POD vendors, a book block that will print at multiple POD vendors, and metadata to sell that book in global sales channels, then yes, we did produce that many titles,” said Mitchell Davis, president of BiblioLife, parent company of BiblioBazaar.<sup>10</sup>

These small “nontraditional” publishers pose a danger to researchers examining titles in the public domain. I use the term “danger” because many editions from these publishers, perhaps most, do not appear to sell any copies in a year or even a decade, making them more like phantom editions than real editions. This is not merely a hypothetical concern: 480 out of 774 BiP listed editions in my soon to be described vintage bestseller data, fail to sell any copies in 2004, and 311 of these editions fail to sell any units in the full 13 years of my data. A large portion of the zero-selling editions in my sample is due to the 310 Kessinger Publishing Company editions, all of which have zero sales in 2004 (and 163 of

---

<sup>8</sup> Most published editions of titles in the US receive an ISBN (International Standard Book Number) number purchased from Bowker Identifier Services. An ISBN represents a single format or a single edition of a title. Some self-published authors choose not to purchase or use an ISBN number.

<sup>9</sup> Publishers Weekly also reported that the second and third highest production of titles was from Books LLC and Kessinger Publishing LLC which produced 224,460 and 190,175 titles respectively. See <https://www.publishersweekly.com/pw/by-topic/industry-news/publishing-and-marketing/article/42826-self-published-titles-topped-764-000-in-2009-as-traditional-output-dipped.html>

<sup>10</sup> See <https://www.publishersweekly.com/pw/by-topic/industry-news/publisher-news/article/42850-bibliobazaar-how-a-company-produces-272-930-books-a-year.html>

which are zero-sellers over the full 13-year interval of my data). The data used by Reimers are also strongly influenced by these types of editions.<sup>11</sup>

Treating these editions as viable could have powerful misleading effects on conclusions that researchers might draw. Small publishers such as these might have unusual pricing strategies, as Ellison and Ellison (2018) discuss for sellers of used books. For example, in my bestseller data, the publisher Kessinger is listed in BiP as providing 117 different paperback editions at an abnormally low price of \$1.99 (see Table 1 below for typical paperback prices)<sup>12</sup> and Reimers' data include 68 such editions. Because Heald (2008) uses the lowest price edition (whether sold or not) as the representative price for a title, if any of these Kessinger editions were in his data, his methodology would conclude that the selling price for these titles was much lower than the prices consumers actually paid when they purchased editions of these titles.<sup>13</sup>

Another publisher specializing in old titles is Reprint Services Corporation. Unlike the Kessinger editions previously mentioned, however, their titles tend to have unusually high prices, averaging \$96, and most of their titles in my sample (23 out of 26) do not sell any units in 2004 and 17 sell zero units in the full 13 years. As was the case with Kessinger editions, an analysis that treats these editions as equivalent to editions that actually are sold in markets is likely to distort results.

## IV. The Data

My sample consists of the top 10 fiction best-sellers from Bowker's *Publishers Weekly* yearly best-seller lists in each year from 1895-1950,<sup>14</sup> for a seeming total of 560 possible titles. The actual number of titles differs from this value, however, because individual titles are often listed in more than one year's best-sellers list, and sometimes a yearly tie in ranking leads to 11 titles being listed in a year.

---

<sup>11</sup> For example, among the 2,862 active hardcover and paperback editions listed in Reimers's (BiP) data, 1043 were from BiblioBazaar and 590 were from Kessinger. Furthermore, only 493 of these 1633 editions (~30%) sold any copies over the 13-year period of my BookScan data.

<sup>12</sup> None of these 117 editions sell any copies during the entire 13-year period of the BookScan data even though some of these editions covered well known titles with high yearly sales.

<sup>13</sup> This is consistent with the large increase in CR vs PD price differential he found after excluding minor publishers.

<sup>14</sup> Titles published between 1895 and 1909 originally had a copyright term of 28 years plus an additional 14 if renewed, as opposed to the 28+28 terms available during and after 1910. Those pre-1910 titles might, in theory, have been expected to be slightly stronger works on the margin since to cover a fixed opportunity cost of writing, the expected payout per year on a marginal title would need to be higher with fewer years of payout. Given my focus on bestsellers, however, I would not expect this minor difference to play a role in the analysis. If it did, however, it would advantage works in the public domain.

Once this set of former best-selling titles was created, 21<sup>st</sup> century information was collected for each edition of each title using the 2004 hardcopy version of Bowker's "Books in Print" database which does not contain sales data. Because I only had BiP data for 2004, that became the focal year for the analysis. I also had access to Nielsen (now NPD) BookScan data, covering the years 2004-2016, for physical copies of works of fiction that were sold by retailers. BookScan is normally considered the gold standard for data on physical book sales. The quality of the BookScan data is attested to by the fact that BookScan's subscribers are mainly book publishers wishing to keep track of how their titles are doing in the retail market. Publishers are willing to pay for this information since they do not know the actual sales of their titles until retailers finish returning the copies that do not sell, often many months after the initial delivery. Further evidence of the regard in which the BookScan data set is held is the fact that Amazon, wishing to provide authors with detailed information about their book sales, gives them access to BookScan sales data.<sup>15</sup> Nonetheless, some important channels, such as book clubs, do not have their sales counted by BookScan and BookScan is limited to enumerating sales physical copies.

The BookScan data set contains information for all physical editions sold in BookScan's large panel of retailers, which, given its high coverage of the largest retailers, is likely to include just about every edition (but not every unit) sold by retailers in the U.S.<sup>16</sup> BookScan data includes, in addition to sales, edition information similar to what is found in BiP. I found BookScan data to sometimes contain editions of old best-sellers that were not listed (or not found by my RAs) in BiP. I used BookScan data when BookScan and BiP overlapped in their coverage of variables as I deemed electronic data from BookScan likely to be more reliable than hand copied (very small print) numbers from the hardcopy BiP. Although I had BookScan data for page numbers for the small number of vintage bestsellers, I had to use page data from the WorldCat database for the large number of more typical books when examining the entire fiction sector.

Many of the former best-sellers were no longer in-print, according to BiP. For those that were in-print, my RAs examined titles published after 1922 and determined whether their copyright was renewed 28 years later.<sup>17</sup> All post-1922 titles that were renewed were

---

<sup>15</sup> This practice began in 2011 and continues to this day. See <https://latimesblogs.latimes.com/jacketcopy/2010/12/amazon-gives-nielsen-bookscan-to-authors.html>

<sup>16</sup> BookScan derives its data from transaction information (e.g., checkout scanners) reported by many retailers thought to represent about 85% of the market, including online retailers such as Amazon. Given that BookScan includes retailers with large catalogs of titles, it is likely that almost all titles/editions that are sold in the market are included in the data. As not every retailer is included, it is likely that the BookScan data will be short on the sales of each title/edition.

<sup>17</sup> Examining renewals required checking each post-1922 title against the *Catalog of Copyright Entries*, and an online database of that information can be found at <https://exhibits.stanford.edu/copyrightrenewals>.

deemed copyrighted, all other titles in the sample were deemed to be in the public domain.

These former best-sellers, to which I will hereafter apply the term “vintage,” are currently published using the same printing technologies as recently written titles. In addition, these vintage titles are often published by the same publishers who mainly publish recent titles. Thus, we can use information from the very large number of recently written titles to better gauge the book characteristics that determine the price of books currently sold.

The most important influence on price is usually the book binding. The distinction between hardcover and paperback bindings is well-known but there are three major bindings used in this market with the third category being mass market paperback. Mass market paperbacks are cheap small books that are made for portability.<sup>18</sup> While other obscure types of bindings exist in this market, they represent a very small portion of the market and are removed from the analysis.<sup>19</sup>

Since the price of books is an important element in our analysis, it is worth noting that there are some editions that have unusual prices.<sup>20</sup> For example, the best-selling author John Grisham has leather-bound and autographed editions of many of his novels, and they sell for the very high price of \$250. In other instances, publishers might have a strategy to capture low information consumers while charging very high prices. I label all unusually high-priced editions as “collectibles” (defined as having a price greater than \$65)<sup>21</sup> and remove them from the analysis. There were only 902 collectable editions out of more than 125,000 editions sold in 2004. Similarly, I remove editions with the price of zero which I take to indicate an error in the data.<sup>22</sup>

---

They found that 3 of the 180 post-1922 titles had lost their copyright in this manner. These were *Black Oxen*, *Mistress Wilding*, and *The Plastic Age*, published in 1923, 1924, and 1924 respectively.

<sup>18</sup> MM paperbacks play a diminished role over time and play almost no role in juvenile or nonfiction markets.

<sup>19</sup> Other format binding categories include “library,” “boxed set,” and “board books,” among others. I merge library with hardcover titles and remove editions in the other categories.

<sup>20</sup> The highest prices for editions sold in 2004 (excluding boxed sets) are \$1845, \$900, and \$750. It is possible that some of these values are data errors.

<sup>21</sup> The 99<sup>th</sup> percentile price was \$51 for the entire market. Vintage editions, when ranked by price, had a large jump from \$56 to \$67. Given this, the \$65 value seemed like a reasonable cutoff. None of the main results in the paper are affected by these choices since most of the high-priced vintage titles had no sales and titles with prices greater than \$65 never surpassed sales of 10 units.

<sup>22</sup> Because BookScan data cover physical books that are costly to produce and are sold in (online and offline) retail stores, it is very unlikely that the price would be zero since that would impose potentially large losses on the publisher. There were only 322 instances where 2004 fiction titles had a price of zero (out of 141,042).

Table 1: 2004 Binding Shares, Revenues and Prices for New and for Vintage Editions Sold

	All editions with +sales & Price<65, >0			Vintage editions with + sales & Price<65		
	Share of Editions	Share of Revenue	Avg Price	Share of Editions	Share of Revenue	Avg Price
MM paper	23.4%	29.0%	\$6.52	12.0%	22.8%	\$6.24
Trade paper	53.8%	35.2%	\$16.13	56.2%	73.3%	\$17.51
Hardcover	22.8%	35.8%	\$25.34	31.9%	3.9%	\$26.14
Total	126,244	\$2,423,907,707		552	\$15,056,849	
Revenue is calculated as the product of unit sales and list price.						

Table 1 compares vintage titles to mainstream fiction titles. The left portion of the table, representing the entire adult fiction market in 2004, shows for each binding type the share of market by edition and revenue, along with average prices. The right portion of the table shows similar statistics for the much smaller number of vintage bestsellers. This comparison reveals that revenues are fairly evenly split between formats for the market as a whole, whereas the vintage titles generate almost no revenue from hardcover editions (in spite of the large number of available hardcover titles and their higher prices) and almost three quarters from trade paperbacks.<sup>23</sup> With regards to price, the vintage editions have prices that are quite similar to the mainstream editions being sold: hardcover editions are about \$10 more expensive than trade paperbacks, which in turn are about \$10 more than mass market paperbacks which have prices of about \$6.

Table 2 compares characteristics of copyrighted and public domain vintage editions. The characteristics for CR and PD editions differ in some dimensions. Copyrighted editions have prices that are 16% higher the prices of PD titles. But CR titles also have more pages, and are more likely to be published in hardcover, both of which tend to raise prices. On the other side of the ledger, CR titles are more likely to be published by major publishers, and have older editions, each of which tends to lower price.

Table 2: Characteristics of 2004 Sold, P<65, Editions of Vintage Titles

		mean	t-test Diff	Obs.
Price	PD	\$17.86	3.17	342
	CR	\$20.63		210
# of Pages	PD	328.4	4.9	338
	CR	396.1		198
Major Publisher	PD	22.8%	3.3	342
	CR	35.7%		210

<sup>23</sup> Revenue is calculated as the product of list price and quantity sold. Since most retailers sell non-best sellers at list price these revenue values should be reasonably accurate.

Edition Year	PD	1997.5	4.9	338
	CR	1992.4		209
Paperback	PD	59.4%	1.9	342
	CR	51.0%		210
MM Paperback	PD	11.7%	0.2	342
	CR	12.4%		210
Hardcover	PD	28.9%	1.9	342
	CR	36.7%		210

## V. The Identification Strategy

The broad outlines of the methodology have already been given. Ideally, we would like to have two similar and randomly drawn sets of titles, providing copyright protection to one group and denying copyright protection to the other. We would want to limit variations in the physical quality of the books, so we might require that book formats in the experiment be standardized and that book publishers use the same set of printers. We might insist that all books in the experiment be of a similar number of pages, a similar quality of binding, paper, ink and the use of illustrations. We could then see how the two groups of titles differed in terms of price, sales, and availability.

But even under terms of this ideal experiment, there might be possible problems. First, titles might have different prices even if they were physically identical. Different titles appeal to different types of readers, implying that consumers' reservation prices could well differ across different titles. Clerides (2002), however, has found that for a publisher who gave him access to financial records, book prices were affected mainly by cost shifters, not demand shifters, so demand variation may not be an important factor in book pricing.

Our actual identification strategy tries to follow this basic logic. Authors writing books in the decades surrounding 1923 are likely to have thought that copyright would have long expired on their creations at the dawn of the 21<sup>st</sup> century. There is no reason, therefore, to think that the nature of their creations should differ by whether they wrote their books before or after the 1923 cutoff. The 21<sup>st</sup> century copyright assignment should be unrelated to author efforts and product, and this provides the basis for identifying the impact of copyright on price, sales, and availability.

We also try to control for book characteristics that might influence prices and sales. By using only works of fiction, we reduce the variability in reading audiences and book characteristics compared to using all categories of books (e.g., textbooks). We can control for some of the physical characteristics of books since they are reported in our data, such as the number of pages, the style of binding, or the use of illustrations. Other characteristics are more difficult to control for, such as the quality of the paper, ink, and

binding. For these latter characteristics we use the fixed effects for the imprint producing the title, in the belief or hope that imprints tend to publish books of a similar physical quality. Supporting this belief that imprints might tend to ‘specialize’ in production quality is evidence that imprints tend to ‘specialize’ in publishing editions of particular binding-types, compared to the market as a whole.<sup>24</sup>

Nevertheless, there are two potential problems with this identification strategy as regards the sales of titles. One is due to the fact that the copyrighted vintage titles in our sample are not randomly assigned with regard to the age of the title but in fact are newer than the public domain vintage titles. Characteristics of bestsellers might change over time in terms of English usage, morals, stereotypes and so forth, such that the age of a title (its “recency”), which is related to its copyright status, may tend to affect its current sales or availability in the market.

The other potential problem is that the definition of “book” has changed. Although books have traditionally been printed exclusively on paper, in the last few decades digital books (audio and digital print) have entered the market. This creates two possible issues for our analysis. First, the BookScan data do not include digital editions, so we are ignorant about part of the market. Second, because there is essentially no cost in producing additional digital copies, digital PD titles can be priced much lower (at zero) when competing with physical PD titles, possibly reducing the sales of physical PD titles.

These issues are addressed as we go through the empirics.

## **VI. Copyright and Quantity Sold**

Although the main justification for copyright has been the expectation that it enhances the creation of new works, there have been some theoretic suggestions (with anecdotal evidence) that copyright also may induce post-creation investment in the title to keep the “franchise” going at an efficient level (Landes and Posner, 2003, Adilov and Waldman, 2013). The argument has been that the ownership provided by copyright allows a publisher of a title to eliminate the likely free riding on post-creation investment that would be expected for a title in the public domain (also inhibiting behavior that would decrease franchise value). This ability to fully harvest the returns from post-creation

---

<sup>24</sup> If a Herfindahl index is created for the market shares of the three binding categories (trade paper, mm paper and hardcover) for the entire market, as found in the left-hand side of Table 1, the index value is 3367. Creating a Herfindahl index for the binding shares for individual imprints, however, the value is 7641 for imprints producing more than 100 editions (with a similar value for imprints publishing more than 500 editions). The higher concentration of binding types for imprints implies that they tend to specialize in the binding types they publish. Given this, it seems reasonable that they would also tend to specialize in the other physical qualities of their editions, although we cannot be certain.

investment should increase sales, surplus and profitability for CR titles. Both Reimers and Heald, however, believe that they have tested this theory and found it wanting.<sup>25</sup>

Largely absent in these discussions, however, is the question of whether the fruits of greater post-creation investment would be found in the extensive margin of how many titles are in the market (as Heald presumed) or the intensive margin related to the sales levels of the titles already in the market. With current POD printing technology and the minimal post-creation investment needed to keep such editions in-print, the superior ability of CR title owners to capture post-creation investment might not matter very much to a title's "availability," especially as measured with BiP.

In contrast, post-creation investment might be an important component of sales *quantity* since most publishers do not just put titles in-print and hope that some sell. They have sales agents and marketing departments to help get titles in front of consumers and into bookstores.

### **A. Measuring the Sales Difference**

To determine whether copyright affects the sales of titles we use BookScan's sales data for our vintage titles.<sup>26</sup> A comparison of CR to PD sales in 2004 is found in Table 3. It shows a remarkably large sales advantage for copyrighted title relative to public domain titles. The third row of Table 3 indicates that the average copyrighted title sells almost 4 times as many copies as the average public domain title, and the median CR title sells more than 23 times as many units as the median PD title. These are exceptionally large differences.

*Table 3: 2004 Title Sales by Copyright Status*

	obs	Mean sales	Median sales
Copyright	110	8,799	178
Public Domain	108	2,320	7.5
Ratio CR to PD		3.8	23.7

For anyone expecting copyright to embody textbook monopoly behavior, leading to lower sales, higher prices, and the attendant deadweight losses, the numbers in Table 3 should be quite shocking. If these numbers accurately reflect the impact of copyright, something other than a copyright-induced exercise of monopoly power must be occurring in this

---

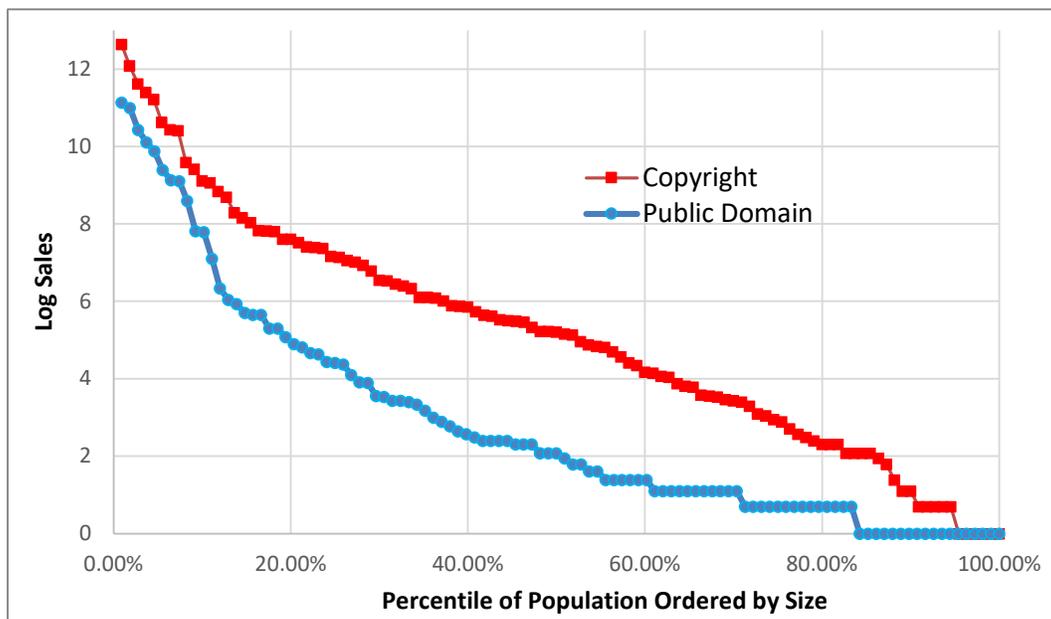
<sup>25</sup> Heald believe that his finding that CR vintage titles are not more abundant than their PD counterparts is evidence against this post-creation investment theory, and Reimers believes that a coefficient value in her estimation argues against this hypothesis.

<sup>26</sup> Although Reimers had access to the BookScan data set, she only uses it to gauge the relative size of her "measured" sales of vintage titles on Amazon compared to sales of those titles in the overall market as measured by BookScan.

market. The rest of this section explores these numbers in more detail and then examines some alternative explanations for the greater sales of copyrighted works.

These larger average sales for CR titles could be due to just one or two top selling titles, since sales of vintage best-sellers tend to be dominated by a relatively small number of top sellers, as is the case for the entire book market.<sup>27</sup> To determine how CR sales differ from PD sales throughout the samples, Figure 1 compares the (logged) unit sales for each PD and CR title with the same ranking percentile.<sup>28</sup> Because the number of PD and CR titles is almost the same (110 CR, 108 PD), the figure effectively compares titles one by one. The figure makes clear the rather remarkable result that each and every ranked CR title outsells its PD counterpart throughout the distributions, until sales converge on 1 unit [ $\ln(1)=0$ ]. Thus, all CR vintage titles sell better than their PD vintage title counterparts. In fact, the greatest CR sales advantage is in the range of the 40<sup>th</sup> to 60<sup>th</sup> percentile, where CR titles tend to outsell PD titles by about 20:1. That is why the difference in median values is so large in Table 3.

*Figure 1: Title Unit Sales by Sample Percentile, from Largest to Smallest, 2004*



The distributions of sales for both PD and CR titles are highly skewed, have a very large range, are greatly impacted by a small number of outlying values, and thus are not

---

<sup>27</sup> The top 5% of vintage titles generate 84% of unit sales whereas for the full sample of 2004 titles the top 5% generate 86% of sales.

<sup>28</sup> Thus, if there were 100 PD titles and 50 CR titles, the 20th percentile would compare the unit sales of the 20th PD title and the 10th CR title.

normal. Because each sample has slightly above 100 observations, parametric estimation should be viable, but we will also provide nonparametric testing of the sales differences between the samples. Using logged sales to weaken the impact of outliers, a t-test comparing means of sales for CR and PD titles has a value of 5.66. Similarly, the nonparametric Mann-Whitney test, which is based on ranks and thus removes the outsized influence of outliers, provides a significant z value of 5.63. The uncompressed raw data provide less statistical significance for CR titles' greater sales, with the t-test measuring a statistically marginal 1.79, although we should expect somewhat more imprecision when there are only a handful of titles that dominate sales, as seen in Figure 3 below.

But this analysis is for 2004 alone. In Appendix Z, I reproduce this analysis for the years 2008, 2012, and 2016 and find results very similar those in Figure 1 and Table 3.<sup>29</sup> When I pool the data for these four years the t-test for differences in sales increases to 3.3, for logged sales it increases to 7.6, and the Mann-Whitney test provides a z-score of 7.8. With this more complete data the difference in sales, whether logged or not, is significant well beyond the 1% level.

## **B. *Testing Potential Identification Problems***

Before we can accept these values as evidence that copyright increases sales, however, we need to address the two potential identification problems mentioned earlier. The first issue is the possibility that sales of physical PD titles are reduced due to direct competition from freely available digital copies of identical titles on sites such as Project Gutenberg, Google Books and Amazon, whereas copyright prevents CR works from encountering similar competition.<sup>30</sup>

Although competition from free digital downloads should reduce sales of physical copies to some extent, whether the impact is large or small depends on how good a substitute those free digital copies are. The current analysis is based on data in 2004. Google Books was not officially unveiled to the public until October of 2004 and the Kindle eReader (as well as the smartphone) was not launched until 2007.<sup>31</sup> Project Gutenberg, therefore, appears to be the dominant free competitor to physical books in 2004, and its electronic

---

<sup>29</sup> For each of these years every CR title sells more than the similarly ranked PD title, as in Figure 1, except for 3 ties in 2016.

<sup>30</sup> Reimers assumed that competition from free digital books fully explained the lower PD sales. She states on page 263 "editions of protected [copyrighted] titles are sold more often on Amazon than editions of public domain titles, likely due to zero-price competition from Project Gutenberg and Google Books."

<sup>31</sup> The Sony Librie, the first commercial reader to use E-ink (eliminating the need for a backlight) was introduced in 2004, but only in Japan.

titles were going to be read mainly on desktop computer screens<sup>32</sup> (unless they were printed out, which would lessen or eliminate any cost advantage relative to physical books). The free digital download reading experience in 2004, therefore, would seem to be quite inferior to using a physical book, and quite inferior to the not yet available smart phones, tablets, or eReaders, since desktop computer screens in 2004 were not portable, required sitting down, and tended to be somewhat hard on the eyes. Thus, the substitutability should have increased dramatically after the introduction of eBook readers, smart phones, and tablets. Because we have data from prior to the eBook phenomenon and through its peak, we can test whether and by how much this much improved substitutability of free digital downloads led to the predicted increase in the CR/PD sales ratio over time.

*Table 4: Change in CR to PD Title Sales Over Time*

	2004	2008	2012	2016
CR/PD Average sales	3.79	3.21	3.17	3.87
CR/PD Median Sales	23.73	7.39	5.86	3.68
eBook Share of Trade Units	0%	<1%	17.9%	14.7%

The first two rows of Table 4, which reproduces information found in Appendix Z, display the ratio of CR to PD sales for four years, measured by average and median values. The third row presents the market share of eBooks over time, as a proxy for the ease and familiarity users are likely to have with downloading digital books. Contrary to expectations, the CR/PD sales ratio (whether median or average) does not increase after the growth in familiarity and utility from digital books. The ratios of CR to PD sales in 2012 and 2016, when digital sales were large, are about the same as what we find in 2004 and 2008 when smart phones, Kindles, and commercial digital books were mainly nonexistent. This failure to find support for the free digital book substitution thesis for PD titles does not mean that there might not be a small effect. But it appears that we can rule out this thesis as a major explanation of the large 2004 CR/PD sales differential. I conclude that competition from free digital downloads is too weak to cause any but a minor portion of the large CR/PD sales difference.

We next turn to the suggestion that newer vintage titles will sell better than older vintage titles merely because they are newer, even though the newest titles would have been 54 years old in 2004. While it is clearly true that in the time period near the date of publication, the recency of the publication matters a great deal in the sales of the title, it is less clear that this would be so after a handful of decades have gone by. If it were

---

<sup>32</sup> Desktop computers made up nearly three quarters of the flow of new personal computers in 2004 and thus made up an even greater share of the stock of personal computers in use according to <https://www.zdnet.com/article/73-3-of-pcs-sold-in-2004-were-desktops-24-laptops/>.

generally true that more recent titles sell in larger quantities, we should expect to see a positive relationship between year of publication and sales. To avoid conflating the impact of copyright from the impact of recency, we must examine the relationship between recency and sales separately for CR and PD titles.

Figure 2: Titles with Logged 2004 sales and Publication Year

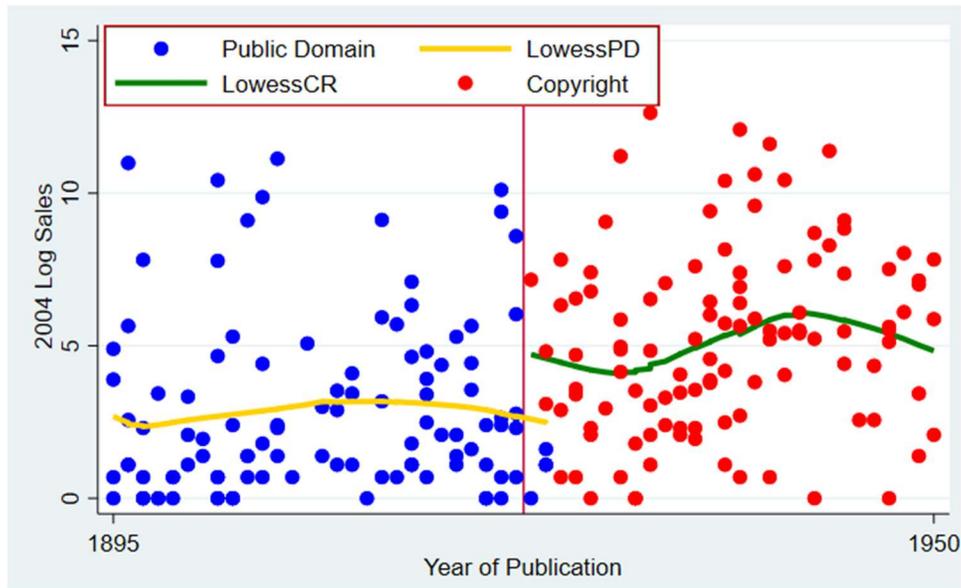


Figure 2 shows the relationship between logged sales and publication year for every title, with the 1922/23 cutoff shown by the vertical line.<sup>33</sup> In addition to the data points, we can see two lowess smoothed curves, one for CR titles and one for PD titles.<sup>34</sup> Public domain titles (left side) show no relationship between year of publication and sales, as the lowess curve, after a very brief decline, rises slowly and then falls slowly. A linear regression line (not shown) has a very slightly upward slope that is not statistically significant.<sup>35</sup> Figure 3 represents data identical to Figure 2 except that the sales values are not logged. The lowess curve and regression line (not shown) for PD titles both show very slight *declines* in sales as PD titles become more recent.

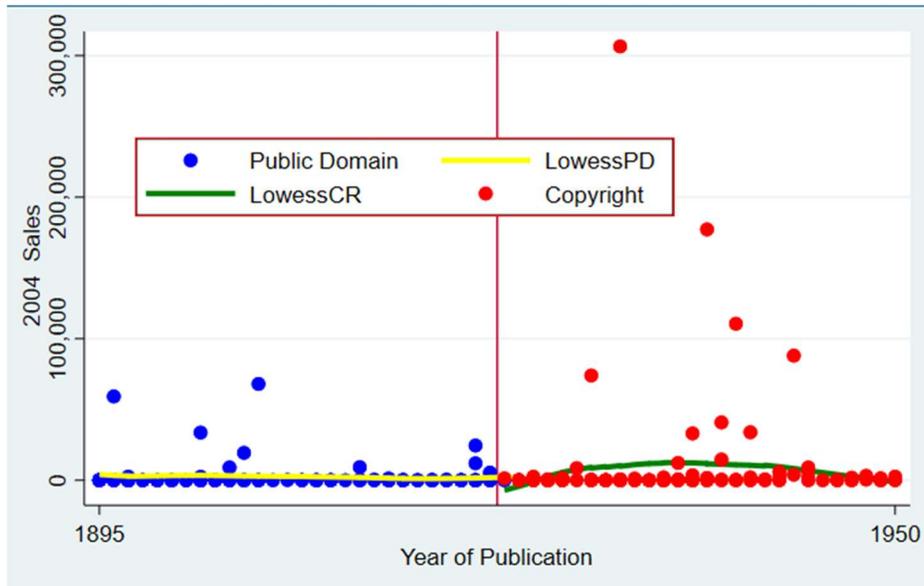
<sup>33</sup> The three public domain points that are mixed in with points representing copyrighted works are titles published after 1922 that were not renewed 28 years later and thus entered the public domain.

<sup>34</sup> The bandwidth parameter used in the lowess smoothing is .8. I also tried smaller bandwidths (down to .1) to make the curve more localized, but it did not change the basic relationship of the two lines.

<sup>35</sup> The coefficient on logged PD sales is 0.02 with a t statistic of 0.65 and for uncompressed sales it is -97.7 with a t of -1.0.

We conclude that there is no evidence that age of a title affects sales for PD titles, whether we used raw or compressed (logged) data. Figure 3 also makes apparent the large difference between top selling titles and more ordinary titles and the obvious outsized influence that top selling titles are going to have on average values in the uncompressed data.

Figure 3: Title Sales by Publication Year



The nature of the relationship between recentness and sales is somewhat less clearcut for copyrighted titles than it was for PD titles. The lowess curve for logged sales (Figure 2) falls at first (from a higher sales level than the PD titles of similar vintage) but then begins to rise after about 1930 or so, peaks in about 1940 and then falls. A linear regression suggests a significant positive relationship between recency of publication and logged sales for copyrighted titles, but a regression on unlogged sales has the opposite sign, while being insignificant.<sup>36</sup>

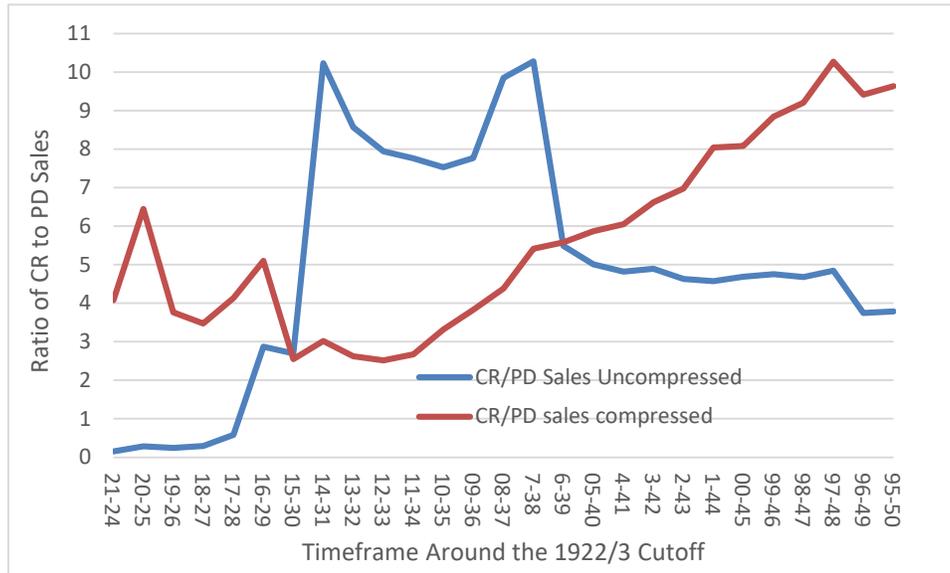
How should we interpret these somewhat divergent pieces of information? Given the complete lack of evidence that recency affects sales for PD works, to believe that recency affects sales for CR titles would require an ad hoc thesis such as that recency has no effect until it is within, say, seventy-five years or so. That could be the case, but it seems rather strained as well as being inconsistent with the last ten or fifteen years of the data.

---

<sup>36</sup> The coefficients for sales and logsales are -42.2 and 0.09 respectively, with (robust) t-statistics of -0.13 and 2.34

Another approach to test this recency hypothesis is to reduce the age difference between CR and PD titles by reducing the interval around the cutoff. Unfortunately, doing so also entails reducing the size of the sample. Figure 4 shows the relative size of CR sales to PD sales for compressed and uncompressed sales values for different intervals around the cutoff.

Figure 4: Comparison of CR and PD 2004 Title Sales by Time Period



Altering the number of years around the cutoff provides two slightly different sets of results, depending on whether we used compressed data or uncompressed data, although in both cases CR titles generally outsell PD titles by a considerable margin. The sales differential for compressed data, represented by the red line in Figure 4, provides simple and consistent results. CR titles always outsell PD titles (the CR/PD sales ratio is always greater than 1)<sup>37</sup>, for any timeframe and, as shown in Appendix Z, the difference is always statistically significant for any timeframe consisting of 7 or more years (x2) around the 1922/23 copyright cutoff (at which point there are 59 titles in the sample, 34 PD and 25 CR). This evidence from compressed data indicates that the sales improvement of CR titles over PD titles holds even for short timeframes around the cutoff. Therefore, the sales advantage of CR titles is unlikely to be due to recency because the newer titles are hardly newer at all.

Using uncompressed data provides a similar result in most cases, but there is less precision due to sales outliers. Because outliers only occur every few years, using short

<sup>37</sup> These values were calculated by finding the average logged sales by copyright status and then exponentiating the average values and forming the ratios, which range from a low of about 2.5 to a high of 10.3.

time periods runs the risk of results being dominated by just one or two titles, and that happens here. The ratio of CR to PD average uncompressed sales is shown as the blue line in Figure 4. Contrary to the general result, PD titles outsell CR titles for the first 6 years around the cutoff (the ratio on the axis is less than 1)<sup>38</sup>, but for any wider period, CR titles outsell PD titles.

Examining Figure 3 makes it apparent that there are some clearly above average selling PD titles within one or two years of the demarcation, but no such CR titles, which is why PD sales values are higher for the periods within 6 years of the demarcation. Once some big CR sellers are in the sample, however, CR titles sell between 4 and 10 times as many copies as PD titles. In Appendix Z I present evidence that the uncompressed sales advantage for CR works is of borderline significance (5%-10%) starting with periods 16 years (x2) or greater around the 1922/23 cutoff and continuing until the full sample of titles is included.

Using narrower periods to weaken the impact of recency therefore does not change the conclusion that CR titles sell more than PD titles except in the instance of uncompressed data and periods less than 7 years (x2), and this is clearly due to a small number of better selling PD titles that are not balanced by any similar CR titles in those years. The more general result is that using smaller time periods to weaken the impact of recency does not alter our finding that CR titles widely outsell PD titles.

But even if recency were related to additional sales for CR titles, however, it is not clear whether this would change the conclusion that copyright increases sales of titles.

First, it appears that copyrighted titles sell in larger quantities than PD titles even before any serious effect of recency could kick in for CR titles. Comparing the lowest values for compressed [logged] sales of PD titles in 1922 and CR titles in 1923 leads to the conclusion that CR titles in 1923 sell 6.4 times as many copies per title as PD titles in 1922 (converting the average compressed values into uncompressed sales).<sup>39</sup> Performing the same calculations, but for each of the two full periods, reveals that copyrighted titles sell 9.6 times as many copies as PD titles,<sup>40</sup> so that about two thirds (6.4/9.6) of copyright's advantage in sales would be unrelated to the effect of recentness on CR titles' sales. Similarly, when we use linear regression on compressed CR data and use the resulting significant coefficient to predict 1923 compressed sales, we find that CR titles sell 3.1 times as many units as average PD titles (which are not related to recency). This indicates

---

<sup>38</sup> In the 1917-1928 period there are 48 titles, 29 PD and 19 CR.

<sup>39</sup> We take the compressed lowest estimates for 1922 and 1923 and convert them to uncompressed values using exponentiation, and then compare the uncompressed values.

<sup>40</sup> This value is also found in Figure 4's red line for the complete time period of compressed values found on the right-hand edge of the figure.

that about 1/3 of the difference in sales of CR and PD titles was unrelated to the impact of recency on sales. It thus appears that a large portion of the difference between CR and PD sales is not related to recency even if one believes that recency affects the sales of CR titles.

Second, given that recency does not affect sales of PD titles, it is crucial to understand why recency might affect the sales of CR titles. It might just be that recency only matters within some timeframe that does not include years prior to 1923 in our 2004 data, as already mentioned. But it might also have something to do with copyright itself. If there is something about ownership that causes recency to affect sales, then copyright would still be responsible for the higher sales of more recent CR titles. Without a better understanding of the cause of recency's impact on CR titles, we cannot presume that a relationship between recency and sales for CR titles negates a claim that copyright is responsible for increases sales.

In spite of a few anomalies, the conclusion of this section is that copyright increases the sales of titles. The two potential alternative explanations, competition with free digital downloads and the recency of copyrighted works, have been examined and found to be unable to account for the large advantage in sales for CR works. This finding of copyright causing greater sales is consistent with a strong post-creation investment effect. Admittedly, this is only indirect evidence for post creation investment because we are not measuring post creation investment per se. Nevertheless, it is a positive sales effect that is brought about by some aspect of copyright, and it is not clear that it matters what aspect of copyright brings about the additional sales, although the post-creation investment hypothesis must be a leading candidate.

## **VII. Copyright and “Availability” of Vintage Titles**

One of the questions that has been examined in this literature is whether a copyright regime increases or decreases the availability of already published titles. Heald concludes that physical editions of PD titles, in more recent years, are just as if not more likely to be made available in the market (as evidenced by BiP) than are copyrighted works. He believes this runs counter to the expectation that competition could reduce the market viability of some titles which might only survive under single ownership. Looking at editions, not titles, Reimers and Heald both conclude that copyright reduces the number of editions available per title.

A question they did not answer is whether the greater availability indicated by BiP data translated into greater consumption variety. Heald understands the problem, when he says: “Arguably, comparative sales data would be a better measure of availability than in-print status; however, historical sales data are generally not publicly available” [p. 1038]. Since we have such sales data, we can measure consumption availability.

Table 5 provides the numbers for this type of “availability” in 2004. As already explained, my data consist of the top ten bestsellers for each year between 1895 and 1950. Because a bestselling title in one year is often also a bestselling title the next year, the actual number of first-time bestselling titles in a year is generally less than 10, as seen in row 1 of Table 5.<sup>41</sup> The number of possible first-time bestsellers is somewhat higher for PD titles, indicating that titles were more likely to appear multiple times in the bestseller lists prior to 1923 than after, although this difference is obviously unrelated to current copyright status.

*Table 5: 2004 Availability of Vintage Bestselling Titles*

	Total	Copyrighted	PD	t-test CR diff PD
1. # Potential Titles per Publication Year	8.84	8.68	9.00	
2. % Vintage Titles “in-print” (BiP)	65.0%	54.5%	75.6%	5.28
3. % Vintage Titles with Sales	44.2%	47.2%	41.1%	1.32

Using the hardcopy 2004 BiP, Row 2 shows that 65% of the potential titles in our sample were still in-print in 2004. PD titles were considerably more likely to be in-print than CR titles, with in-print values of 75.6% and 54.5% respectively, a statistically significant difference. This apparent finding that copyrighted vintage titles are less likely to be in-print than PD vintage titles is largely consistent with what Heald found.<sup>42</sup> The implication would seem to be that copyright decreases the availability of titles in spite of copyright’s advantageous elimination of free riding on post-creation investment. This has been taken by Heald to support a view that the removal of copyright does not lead to the underexploitation of already created works.

The BookScan data set allows us to go directly to the share of vintage titles that are sold each year, as opposed to merely being in-print. There are two sources of difference between BiP and BookScan data. First, there are some BookScan listed editions that are either not in BiP or not found in BiP.<sup>43</sup> Second, and probably more important, many BiP listed editions were not found to have any sales and thus do not appear in BookScan. The third row of Table 5 compares the shares of potential vintage titles that are sold in 2004, by copyright status. Contrary to the seemingly greater availability of PD titles in row 2, CR titles are somewhat *more* likely to be sold than are PD titles, although the latter difference

---

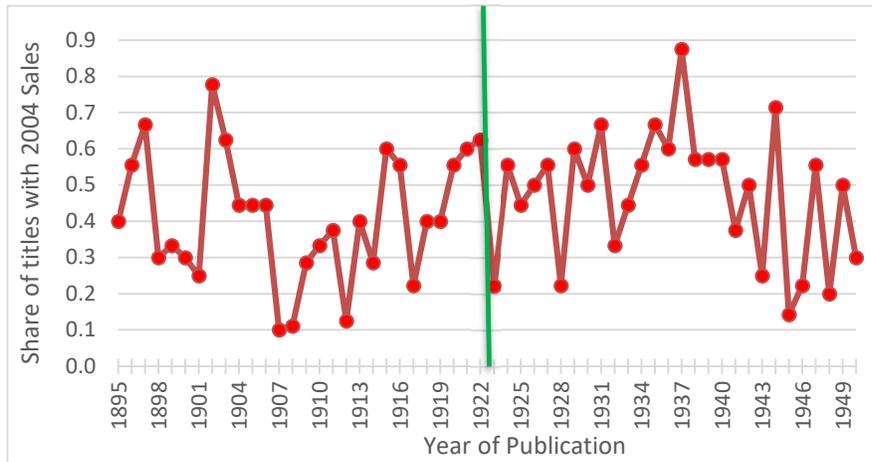
<sup>41</sup> In two years (1898 and 1922) there was a tie for the tenth-place bestseller, so that 11 “top ten” titles were listed.

<sup>42</sup> Heald found that copyrighted and PD vintage works were about equally likely to be available in in the 1980s and 1990s editions of BiP (see his Figure 1). But by 2004, PD works were considerably more likely to be found in BiP (his Figure 1, whose values I eyeballed at 55% and 86% for copyrighted and PD works respectively).

<sup>43</sup> BookScan has editions of titles, sometimes the highest selling editions, that Reimers did not find in the electronic version of BiP, as discussed in Appendix R.

is not statistically significant. This result is consistent with the claim that copyright’s prevention of free riding would enhance the number of titles “available” to consumers although the result is too weak to verify that claim.<sup>44</sup> Nevertheless, the “availability” of PD titles relative to CR titles is quite different depending on whether we use sales or in-print status to measure it.0000

Figure 5: 2004 Share of Vintage Titles with Sales by Pub Year



In addition to comparing average availability, we can check the trends of availability to make sure that the difference in means is not due to some underlying time trend. Figure 5 shows, for each publication year, what share of the yearly vintage best-sellers are sold in 2004, with the 1923 copyright cutoff shown as the vertical line. The data points indicate no apparent visual trend for PD or CR titles (a regression indicates .09 extra titles per PD decade and -.04 fewer titles per CR decade, although neither is statistically significant). We conclude from this that the values in Table 5 do not mask important trends.

Finally, both Reimers and Heald examine how copyright affects the availability of *editions per title*, although additional variants of a title would seem to provide far less novelty and value than additional titles, making the importance of more or fewer editions per title of questionable interest. Nevertheless, both researchers find that there are more editions for PD titles than for CR titles. Reimers finds that there are 43.7 editions for each PD title and only 7.4 editions for each CR title, a very large difference.<sup>45</sup> Heald limits editions to 1 per title for each publisher and finds a smaller advantage of 5.2 editions per PD title versus 3.2 editions for CR titles.

<sup>44</sup> None of this is to say that PD titles are not more likely to be available than PD titles *in some other form*, particularly since there are organizations such as Project Gutenberg or Google which make PD titles available as free digital downloads.

<sup>45</sup> Found in the third row of her Table 1.

Our results, found in Table 6, have the same sign as found by Reimers and Heald. Although I find more editions of PD titles than CR titles, the difference is only 21% (2.4/2.0) compared to the 590% for Reimers or the 62% for Heald. When I limit the comparison to editions with sales, the difference between PD and CR increases to a somewhat larger 49% (2.9/2.0). Of course, it should not be surprising that PD titles have more editions sold when they are freely available to any publisher.

*Table 6: Editions per Title by Copyright Status*

	total	copyrighted	PD	T-stat
Editions per BiP title	2.2	2.0	2.4	1.63
Editions per title with sales	2.4	2.0	2.9	2.53

## VIII. The economic importance of top vintage titles

It is important, when considering the use of resources in post-creation investment, to understand that although these vintage titles do not have yearly sales that are at the top of the current yearly distributions, these titles can nevertheless be of considerable financial consequence to publishers because of the durability of their sales, thus warranting financial attention. Some of these very old titles continue to sell in large enough quantities year after year so that in a timeframe measured in decades, or even a century, they would often surpass, sometimes by a wide margin, many current best-sellers' lifetime sales (except those destined to become longtime classics themselves). These titles are like the apocryphal tortoise that can outdistance the hare, so to speak.

Over the 13-year period of our sales data, the top 100 yearly bestsellers had 13-year sales that ran from about 1.2 million units to 6.4 million (a single outlier sold 9.2 million), with the 20 top titles each selling more than 3 million units and the top 50 each selling more than 1.9 million. In that same period, the top selling vintage title sold about 2.5 million titles, and the top 5 vintage titles averaged 1.2 million units with a median value of 804 thousand. In addition, the top title from 1895-1950 sold 3.9 million copies and the top old title (1960) was number 2 overall, selling 6.4 million copies.

Back-of-the-envelope extrapolations of these results to a longer timeframe, say 52 years, would lead to sales that are four times as high as found in our 13-year period, ignoring population growth and other changes. Extrapolating values in this manner makes it clear that these top selling vintage titles are very valuable properties indeed. The top vintage title, with expected sales of about 10 million units sold in 52 years, would surpass any bestseller in our 13-year period, and the other leading vintage titles would be among the

very top current sellers. For longer time periods, these leading vintage sellers' sales would be so large that they clearly would justify considerable post-creation investments.<sup>46</sup>

Even the vintage titles that are somewhat below these top sellers have the financial benefit of a very large number of years with sales, improving their lifetime sales well above what a typical title with similar yearly sales might achieve. Therefore, it is not just the top vintage sellers that have sufficient market potential to warrant post-creation investments, but also more moderate sellers that might not otherwise be thought to justify such investment given the yearly sales of these titles.

## **IX. Factors that affect the Cost and Price of Books**

There are many factors that affect the production cost of books that would be expected to influence their prices. The most obvious factor is the physical binding of various editions. Another factor that will influence the cost of production is the number of pages since some titles may run only 100 pages while other titles may run to a thousand or more.

Given the nature of book production, economies of scale are to be expected. The reasons are, first, that publishers are likely to switch from POD to offset, and second, larger print runs decrease the average cost per unit when using offset printing. Appendix S provides evidence of economies of scale in production for hardcover and trade paperbacks, but diseconomies for MM paperbacks. Although the evidence of economies of scale is somewhat mixed, we proxy for larger print runs in the regressions by whether an edition is published by a major publisher.<sup>47</sup> Editions brought to market by major publishers greatly outsell editions from other publishers, with the average and median yearly edition sales of the former (2,974; 133) being much larger than the sales of the latter (235; 7).

Illustrations in books are another additional expense, both in terms of paying the illustrator and for the additional printing costs, particularly if the illustrations are in color. The relative share of pages taken up by illustrations would be an ideal measure, but we are limited to a dummy variable indicating whether or not the book has illustrations.<sup>48</sup>

---

<sup>46</sup> Even though these classics might sell the most total copies, we would need to know production costs and present values of these numbers to determine how these classic titles compare to traditional bestsellers in terms of profitability

<sup>47</sup> We determine "major" publishers as the top selling fiction publishers in 2004 that are not specialized in self-publishing. These include, in order of sales, Random House/Penguin Group, Simon & Schuster, Hachette Book Group, Harpercollins, Macmillan, Harlequin Books, Houghton Mifflin Co, Kensington Publishing, and WW Norton, each of which had 1% or more of the market in terms of revenue and together represent 72% of the market.

<sup>48</sup> The information on the inclusion of illustrations came from WorldCat.

Other factors, such as paper size and weight, ink quality, dust jackets, cover art, and so forth also affect the cost of book production, but data at this level of detail is in general not available, although it seems likely that imprints, and to a lesser extent publishers, will tend specialize somewhat in the general quality books they produce (see discussion around footnote 24).

The final factor examined is the age of the edition. In 2004, half the editions were more than 3 years old, a quarter of editions were more than 7 years old, and one tenths of the editions were more than 12 years old. Because there was modest inflation over the decade prior to 2004, the later the year of an edition’s publication, the higher the price would be expected to be, ceteris paribus.

*Table 7: Edition Characteristics influencing log price, 2004*

	all	No Imprint FE	hard	paper	MM
Pages (x100)	0.063 (66.23)	0.065 (72.05)	0.043 (28.47)	0.072 (47.06)	0.071 (47.35)
Edition year (x10)	0.0756 (31.22)	0.0715 (33.98)	0.0496 (11.65)	0.0549 (15.73)	0.1017 (21.09)
Majorpub	-0.0379 (-5.15)	-0.04478 (-20.77)	-0.0128 (-0.69)	-0.0828 (-7.16)	0.0208 (2.38)
Illustration	0.0310 (8.45)	0.02206 (6.81)	0.0151 (2.58)	0.0431 (8.08)	-0.0024 (-0.34)
binding fixed effects	x	x			
remove collectables	x	x	x	x	x
Imprint fixed effect	x		x	x	x
N	107,263	107,960	25,376	53,694	28,193
adj. R-sq	0.85	0.72	0.60	0.55	0.58
Dep Var= Log Price; robust t-stat in parenthesis; titles with Price>65 removed.\					

Table 7 shows how these cost factors affect the price of editions, first for the entire market and then separately for each type of binding. In the first column we regress the log of price of an edition on the four cost-factor variables expected to impact the price of books (number of pages, edition year, major publisher dummy, and illustration dummy) and the binding dummies. We also remove collectibles (editions priced over \$65) and control for imprint. All variables have the expected sign and achieve statistical significance. The second column is the same except for the fact that it excludes the imprint fixed effects. The results are fairly similar with or without fixed effects, except for the measure of fit.

The next three columns show the results for each binding type. There are two notable deviations from the overall results, both in the mass market paperback category. First, the major publisher dummy is generally negative whereas it is positive when examining the mm binding category. This would indicate that the larger sales associated with major

publishers does not lower the price of mm paperbacks. This is not a surprise given the lack of evidence for economies of scale for mm paperbacks in contrast to hardcover and paperback editions, as seen in appendix S.<sup>49</sup> Secondly, illustrations do not seem to raise the price for MM paperbacks. The lack of a relationship between price and illustrations might be in part because illustrations are so uncommon in MM paperbacks (4% as opposed to 18% for trade paper and 12% for hardcover).

## **X. The Impact of Copyright on Price**

If we wish to know the impact of copyright on price for our sample of vintage titles, we need to focus on the price controlled by the copyright owner. The copyright owner or their assignee (usually the publisher) determines the wholesale price charged to a retailer and also the suggested retail price, so those would be the prices to examine if they were available.<sup>50</sup> Because we do not have access to the wholesale price charged to a retailer, we will use the retail list price.<sup>51</sup>

As a practical matter, retailers typically charge the list price that is often found on the book cover, with the major exception being discounts for current bestsellers. Author royalties are also normally based directly on the list price, although occasionally they are based on the wholesale price (net cash received) which is itself based on the list price.

Previous researchers have used list price when comparing prices for copyrighted and public domain works, except Reimers (2018), who uses Amazon retail prices.

### **A. What should be the unit of observation?**

Copyright is assigned to a title. Except for titles consisting of but a single edition, it is difficult to conduct an analysis of copyright's influence on the price of titles per se unless you combine the various editions together to form some sort of amalgam title. Amalgam

---

<sup>49</sup> Nevertheless, average sales of mm paperbacks from major publishers are more than four times as great as from minor publishers. Although this difference is less than half of the overall sales difference between major and minor publishes, it is still large and significant.

<sup>50</sup> The list prices do not change over the life of the barcode representing that ISBN, although a price can change for the ISBN if the barcode is changed, although that seems to happen rarely if at all. See the discussion of barcodes at Bowker's FAQs <https://www.myidentifiers.com/faq/barcodes>. I examined any price changes over the 13 years of my data for approximately a million editions and discovered only 241.

<sup>51</sup> Typical trade discounts (off list price) are currently about 55% <https://www.ingramspark.com/blog/why-should-i-discount-my-book> although Greco claims (p. 160) that discounts ranged from 42%-48% in the period prior to 2005.

titles and amalgam title-binding combinations<sup>52</sup> can be created using sales as the “glue” to hold the amalgams together. The alternative to amalgam titles is to treat each edition separately which runs the risk of having a large number of marginal and potentially misleading editions determining the average price. Appendix A describes how I create amalgam titles and amalgam title-binding combinations and provides average prices and copyright premiums.<sup>53</sup> The downside of using amalgam titles is that certain ordinal characteristics of editions, such as the imprint publishing an edition, must be excluded from the amalgam analysis because these ordinal characteristics cannot be combined to find an average.

Many vintage titles sell zero units in 2004 despite being listed as in-print, and some titles contain collectable editions. Eliminating editions with zero sales or selling for more than \$65 leaves 212 (amalgam) titles and 323 title-binding amalgams. Of these 212 titles, 107 are copyrighted and 105 are not.<sup>54</sup>

Many listed title amalgams are not really amalgams, however, because titles with but a single edition are included. Because there are 104 such title singletons, however, almost half the amalgam titles are merely single editions. For this reason, the alternative of using editions as the unit of observation, which we also report, is not as different from using amalgams as might be thought.<sup>55</sup>

## **B. Copyright Price Differentials with Controls.**

In Appendix A we compare the average prices for various types of amalgams and editions. Copyright premiums (in percentage points) range from negative 5% up to almost 40%. In Table 2 we saw that for 2004 vintage editions the copyright premium was 16% (\$20.63/\$17.86). We also saw that there were some important differences between average CR and PD vintage titles in terms of pages, publisher type, and so forth, indicating that it could be important to control for those factors.

---

<sup>52</sup> An example of a title-binding amalgam would be taking multiple hardcover editions of a title and finding a weighted average, which would be the hardcover amalgam for the title.

<sup>53</sup> Briefly, an amalgam title takes an average of each variable weighted by the sales of each edition. Amalgam title-bindings takes an average for each binding type weighted by the sales of each edition of that binding type.

<sup>54</sup> These number of observations are slightly lower than those in Table 3 because we did not remove editions with prices > 65 when we were concerned with sales, but we are concerned here with price so we remove editions where prices are thought to measure unusual characteristics of editions.

<sup>55</sup> Heald simplifies this difficulty by taking the lowest price edition of a title as the price of a title under the assumption that this reflects the best deal available to consumers. Heald’s method suffers from possibly choosing a focal edition that is very different from what consumers are actually purchasing.

Table 8 controls for those factors. We have seen that the number of pages, the type of publisher, the inclusion of illustrations, and the age of the edition influence the price of books in the overall market. We exclude illustrations here since, for these vintage titles, its coefficient always is economically and statistically unimportant and its removal has very little impact on the overall results. For the same reason, we also exclude measures of ownership concentration for titles (Herfindahl) which had been constructed.

*Table 8: 2004 Regression Results Explaining Price*

	Amalgam	Binding- Amalgam	Editions	
Copyright	0.165 (2.71)	0.121 (2.81)	0.202 (5.49)	0.164 (4.33)
Pages (x100)	0.065 (3.79)	0.058 (4.29)	0.0846 (6.59)	0.0712 (6.97)
Majorpub	-0.307 (-4.57)	-0.248 (-4.79)	-0.192 (-4.55)	-0.0579 (-0.77)
Edition year (x10)	0.011 (3.14)	0.03 (1.6)	0.0743 (3.84)	0.0292 (1.82)
Binding dum	y	x	x	x
Imprint dum				x
N	201	309	533	533
adj. R-sq	0.49	0.614	0.528	0.874
Log price is the dependent variable, t statistics in parentheses, robust SE, P<65, Sales>0; y: The binding variable is continuous, not a dummy.				

Table 8 provides the results.<sup>56</sup> Copyright appears to have a more homogeneous impact on price than was found looking at average values in Appendix A. When these copyright coefficients (which approximate percentage differences because we use logged prices as the dependent variable) are converted into more precise values, they are, in the order of Table 8, 17.9%, 12.9%, 22.4%, and 17.8%. Although I prefer the amalgam results out of concern that those based on editions might be overly influenced by minor editions, the estimates are all quite close. The evidence indicates that copyright causes a price premium in the vicinity of 15% to 20%.<sup>57</sup>

---

<sup>56</sup> The number of observations in Table 8 is smaller than in Appendix A because some observations do not have values for the additional variables in Table 8. The seemingly different results are not due to the difference in observations.

<sup>57</sup> I also ran this regression for the year 2016, although BookScan does not have data on eBooks or audio books which had significant shares in that year, possibly making these results less reliable. The copyright premiums were higher, particularly for the amalgam estimates. The equivalent 2016 coefficients are 0.35,

All the other variables have the expected sign although they are not always statistically significant. As might be expected, the major publisher dummy is insignificant when the imprint fixed effect is included.

Still, these values appear to be lower than the prior estimates discussed in the literature review. The methodologies used in Heald and Li, MacGarvie, and Moser are so different that it is not clear we are measuring the same thing. Reimers' estimate appears to be only slightly higher than those found here but in fact when used books are removed from her sample, as they should be,<sup>58</sup> and the number of pages is included as a control, the estimated copyright premium using her method and data drops to an insignificant 6.8%, so these estimates are actually higher than what her data indicate.<sup>59</sup>

Current royalties to authors top out at 15% for hardcover sales above 10,000, 10% for paperbacks (over 150,000), and 25% of net receipts for eBooks. Nevertheless, publishers are reported to compete for authors by offering advances that are larger than they expect to recoup (more than half the titles do not recoup their advances), meaning that the effective royalty rates are likely to be considerably higher than these stated rates, perhaps 70% higher.<sup>60</sup> Historical American author royalties in the mid-19<sup>th</sup> century were generally in the vicinity of 10% but reached as high as 50% for famous authors such as James Fenimore Cooper or Washington Irving (Liebowitz, 2016). Because our sample consists of bestsellers mainly by well-known authors, we would expect their royalties to be at the highest end of the distribution. The copyright premiums found for our data, therefore, seem to be within the general vicinity of expected royalty rates. Although we cannot rule out some possible monopoly element in the CR premium, we can say that if there were one, it would not be large.

## **XI. Implications for Welfare**

How do these results fit into the conventional thinking about the welfare effects of copyright? In the most traditional view, copyright was supposed to reduce total surplus from the consumption of copyrighted works, with the reduced consumption surplus being

---

0.27, 0.26, and 0.11. Details are found in Appendix A. In spite of these higher value, these results would not seem to change any qualitative conclusions.

<sup>58</sup> Used books make up 22% of her sample.

<sup>59</sup> Her failure to include number of pages is surprising because in her footnote 45 she states: "the estimated marginal costs are strongly positively correlated with the number of pages in the edition, conditional on the edition's format." Her estimated copyright premium drops to 22.5% if number of pages alone is included.

<sup>60</sup> See this [post](#) from an industry insider who reports the standard royalty rates but also claims "royalties could be roughly 70% higher in a world without advances."

“balanced” or offset by the increase in welfare from copyright’s induced production of new creative works.

The literature contains two possible modifications to the “access” portion of this traditional view. The first is that copyright might benefit society after a work is created by promoting post-creation investment in a title, for which we have found indirect support in the higher sales of copyrighted works. The second is that economies of scale in the physical production of books might allow copyrighted titles to achieve greater surplus relative to public domain works because of lower average book production costs with a single producer. We have found some evidence of economies of scale in the production of hardcover and trade paperbacks, but not for mm paperbacks.

The fact that we have found only a relatively small copyright premium similar in size to the royalty payments typically paid to well-known authors, a large positive impact of copyright on sales, and mixed support for economies of scale in production raises the possibility that the impact of copyright on the consumption of already created works might be positive. If this were the case, however, then the traditional analysis of copyright would be stood on its head. The incentive/access relationship would no longer be a tradeoff since copyright would be socially beneficial on both sides of the ‘balance,’ and that would mean that copyright would clearly be beneficial to society.

But that is not all. The criticisms of retroactive copyright extensions (Akerlof et al., 2003) would be incorrect. Similarly, the supposed benefit to society from allowing already created works to fall into the public domain, which has been taken for granted by so many, would also be incorrect

A welfare analysis of these issues was the main focus of Reimers’ 2019 study. Reimers uses a discreet choice model, following Berry and Waldfogel (1999), to estimate consumer and producer surplus in a world with and without copyright. She concludes that copyright reduces welfare if we ignored its impact on the creation of new works. Her analysis relies on a self-created measure of sales based on changes in Amazon sales rankings, but this sales variable turns out to be inadequate for its task as discussed in Appendix R.

I offer no judgment of her welfare measure methodology except to note that it is accepted within much of the profession. Instead, I treat her method as a black box to see what its results would be if better (BookScan) data were used as the measure of sales. Making this one data adjustment turns out to reverse her conclusions, as demonstrated in Appendix R which also considers other possible issues with Reimers’ analysis.

The first row in Table 9 replicates Reimers’ estimate of the average surplus among her vintage titles for 12 months from September 2011 to August 2012. Total surplus (with the assumption that profit is equivalent to producer’s surplus) is found in the right-most

column. The main result of her analysis is that the average surplus for public domain titles is about 50% larger than the average for copyrighted titles, at approximately \$12,500 per PD title relative to \$8000 per CR title. These are the values leading Reimers to conclude that copyright is generally harmful to social welfare when its impact on the creation of new works is excluded.

*Table 9: Average Results per Title for Vintage Titles using Reimer’s Model*

		Consumer Surplus	Profit	Total Surplus
(1) Reimers Original Results	CR	\$2,550	\$5,460	\$8,010
	PD	\$12,532	\$0	\$12,532
(2) Using BookScan Sales Data	CR	\$6,287	\$22,377	\$28,664
	PD	\$18,461	\$0	\$18,461

Row 2 replaces the sales data used by Reimers (based on counting each improvement in a title’s Amazon sales ranking as a sale) with 2012 data from BookScan.<sup>61</sup> As I explain in Appendix R, her method of counting sales has some deficiencies, particularly when an edition sells dozens of copies per day. This is a particular problem for a handful of top selling vintage titles which, according to BookScan, sell up to 650% more copies than her measure of overall market sales for these titles.<sup>62</sup> Replacing her measured sales with the BookScan sales numbers dramatically alters and reverses her results. Row 2 of Table 9 indicates that replacing her sales values with BookScan values causes CR titles to have a surplus approximately 50% larger than the surplus generated by PD titles. In other words, her results are reversed, making them consistent with our finding that copyright enhances the sales (and presumably the surplus) of titles on the access side of the market.<sup>63</sup>

What we can definitively say from this analysis is that Reimers’ methods, ignoring the incentive impact of copyright, cannot be used to claim that PD titles generate greater average surplus than do CR titles. If one has faith in her methodology, however, it appears that a stronger statement can be made, which is that the access portion of the copyright balance is enhanced by copyright, contrary to her published results and to textbook

---

<sup>61</sup> Reimers estimates monthly sales data over 12 months and then multiplies each of her monthly values by 2.5 to bring her yearly sales number into alignment with BookScan values. Although Reimers uses BookScan data to perform this extrapolation, she does not use the BookScan data to adjust her measured Amazon sales for individual titles. This is explained in more detail in Appendix R.

<sup>62</sup> I double checked that the 2011 sales of the top titles were similar to the 2012 data to make sure that the slight mismatch in months covered in our data sets is not responsible for the results.

<sup>63</sup> I also examine Reimers’ measurement of the value of free digital copies of PD titles and conclude that it is overstated and that correcting it can also reverse her conclusions, although only by a little. But that analysis, found in Appendix R, is not needed for my conclusion although it increases the advantage of CR titles beyond what is found in Table 9.

expectations that copyright induces monopoly output restrictions. I am somewhat loath to accept this claim based on someone else's empirical modelling, but I am at least heartened by its general consistency with our earlier result that copyright seems likely to be socially beneficial independent of any incentive impacts.

## **XII. Conclusions**

We have endeavored to use a natural experiment in copyright assignment to examine how copyright has altered the sales, pricing, and availability of vintage fiction titles. Our access to an unusually rich data set has led to findings that call into question the standard understanding of the impacts of copyright.

In particular, we have discovered that copyrighted titles generally sell in much greater quantities than similarly situated public domain titles. The average difference in sales is almost four to one in favor of copyrighted titles, and the sales advantage holds throughout the full distribution of sales. We have tested alternative hypotheses that might have explained this result and found them lacking in support.

The most natural explanation for the greater sales of copyrighted works would seem to be post-creation investment by the publishers of titles, where publishers of copyrighted titles can make marketing investments with the knowledge that they will reap the full return on investment, unlike the publishers of public domain titles who would share the returns with free riding competitors. This is the first evidence of which I am aware, other than case studies, supporting the post-creation investment hypothesis.

We also have two other findings. First, copyrighted titles have higher prices than public domain titles, although the price differential is somewhat smaller than found in most other studies. Also, the price differential is similar to the typical royalty paid by publishers to leading authors. Thus, it is not clear that copyright has any effect on pricing after the higher costs of royalty and investment is taken into account. Second, copyrighted vintage titles are slightly more likely to be sold in the market than are PD titles, contrary to claims based on "in-print" status which tend to find PD titles more available. One reason for this difference appears to be that new business models in publishing have led to serious differences between editions being listed as in-print and actually being sold in the market.

The most intriguing implications of these findings is the possibility that the access/incentive tradeoff may not be a tradeoff at all, at least not with respect to the works of fiction examined here. If copyright provides an incentive for authors to create new works, and if copyright increases the sales of already written works, then copyright would appear to be unambiguously beneficial. It would also mean that retroactive extensions of copyright, which have, in the main, been considered clearly harmful, might in fact also be beneficial. This result makes a strong case for indefinitely renewable copyright. It might also mean that the boundaries of fair use should be narrowed because

the harms from fair use may be larger than previously thought if it interferes with post-creation investment.

Do these results hold for other types of books? Do these results hold for other types of copyrighted works or other types of intellectual property? Clearly, further work on these topics is called for if we are to understand how copyright affects consumption and production of intellectual creations.

AAP StatShot various years, Association of American Publishers, 455 Mass Ave, Suite 700, Washington DC 20001.

Adilov, Nodir, and Michael Waldman. 2013. "Optimal Copyright Length and Ex Post Investment: A Mickey Mouse Approach." *Economic Inquiry* 51 (2): 1101–22.

Berry, Steven T., and Joel Waldfogel. 1999. "Free Entry and Social Inefficiency in Radio Broadcasting." *RAND Journal of Economics* 30 (3): 397–420.

Brief of George A. Akerlof et al. as Amici Curiae in Support of Petitioners at 12, *Eldred v. Ashcroft*, 537 U.S. 186 (2003) (No. 01-618).

Boldrin, M and D. K. Levine *Against Intellectual Monopoly* Cambridge University Press, July 7, 2008.

Clerides, Sofronis K. "Book value: intertemporal pricing and quality discrimination in the US market for books" *International Journal of Industrial Organization* 20 (2002) 1385–1408.

Ellison, Glenn, and Sara Fisher Ellison. Match quality, search, and the Internet market for used books. No. w24197. National Bureau of Economic Research, 2018.

Fisher W.W. (2004), *Promises to Keep: Technology, Law, and the Future of Entertainment*, Palo Alto, California: Stanford University Press.

Greco Albert N., *The Book Publishing Industry*, 2<sup>nd</sup> edition, Lawrence Erlbaum Associates, New Jersey 2005.

Heald, Paul J., Property Rights and the Efficient Exploitation of Copyrighted Works: An Empirical Analysis of Public Domain and Copyrighted Fiction Bestsellers, *Minnesota Law Review*, April 2008 92(4) 1031-1063.

Kitch, Edmund W. "Elementary and Persistent Errors in the Economic Analysis of Intellectual Property" 53 *Vanderbilt Law Review*, November 2000, p. 1727.

Klein, Benjamin, Lerner, Andres V., Murphy Kevin M. "Intellectual Property: Do We Need It? The Economics of Copyright 'Fair Use' in a Networked World," *American Economic Review*, 206-208, May 2002

Landes, William M. and Posner, Richard A. *The Economic Structure of Intellectual Property Law*, Belknap Press, Cambridge, MA 2003.

Landes, W., and R. Posner. "Indefinitely Renewable Copyright." *University of Chicago Law Review*, 70(2), 2003, 471–518.

Li, Xing, Megan MacGarvie, and Petra Moser. "Dead poets' property—how does copyright influence price?" *The RAND Journal of Economics* 49, no. 1 (2018): 181-205.

Liebowitz, Stan J., and Stephen E. Margolis "Seventeen Famous Economists Weigh in on Copyright: The Role of Theory, Empirics, and Network Effects" *Harvard Journal of Law and Technology* 2 Vol. 18, (Spring 2005) pp. 435-457.

Liebowitz, Stan J. "The Myth of Copyright Inefficiency" *Regulation Magazine*, Spring 2009.

Liebowitz, Stan J. "Paradise Lost or Fantasy Island? Voluntary Payments by American Publishers to Authors Not Protected by Copyright." *The Journal of Law and Economics* 59, no. 3 (2016a): 549-567.

Liebowitz, Stan. "The Case for Copyright." *Geo. Mason L. Rev.* 24 (2016b): 907.

Publishers Weekly. "BiblioBazaar: How a Company Produces 272,930 Books A Year" Andrew Albanese, April 15, 2010.

Publishers Weekly. "Self-Published Titles Topped 764,000 in 2009 as Traditional Output Dipped" Jim Milliot, Apr 14, 2010.

Rappaport, Edward Cong. Research Serv., Copyright Term Extension: Estimating the Economic Values (1998), available at <http://countingcalifornia.cdlib.org/crs/pdf/98-144.pdf>).

Reimers, Imke. "Copyright and generic entry in book publishing." *American Economic Journal: Microeconomics* 11, no. 3 (2019): 257-84.

Sorensen, Alan T. "Bestseller Lists and Product Variety" *Journal of Industrial Economics* 2007 55:4 715-738