Copyright and Generic Entry in Book Publishing

Imke Reimers*

March 20, 2017

Abstract

Copyright affects the distribution of creative content. Taking works off copyright promotes their availability, which benefits consumers. But it also allows generic entry to dissipate producer surplus. This paper examines the effect of a copyright on the availability and price of books when incentives to create new works are not affected. I evaluate the welfare impact of the 1998 Sonny Bono Copyright Extension Act by estimating differences in consumer and producer surplus across a range of affected works under copyright protection and in the public domain. Using a regression discontinuity analysis, I find that a copyright significantly limits the availability of works. In a demand and entry model, I find that this leads to decreases in consumer surplus which are more than twice as large as any increases in profits to copyright holders and publishers. Without changing incentives to create new ideas, the copyright extension was economically inefficient.

JEL Codes: L17, L43, O3
Keywords: Intellectual property, copyright extension, generic entry, regression discontinuity, discrete choice demand, consumer surplus

*Department of Economics, Northeastern University, i.reimers@northeastern.edu. I thank Thomas Holmes, Amil Petrin and Joel Waldfogel, as well as Jim Dana, Kyoo-il Kim, Gerard McCullough and Petra Moser, for invaluable advice and many helpful comments. I would also like to thank seminar and conference participants at Minnesota, Kansas State, Rensselaer, Northeastern, Dartmouth, Groningen, Bremen, MIT, the NBER, the International Industrial Organization Conference, and the North American Summer Meetings of the Econometric Society for many helpful suggestions.

1 Introduction

Discussions of intellectual property (IP) rights often focus on their role in stimulating new creative activity. This is obviously an important issue, but there is another dimension: to what extent does IP affect the *use* of a creative work? Sometimes, IP rights extend for such a long period of time – far beyond the life of the original creators – that an extension of rights cannot affect incentives for creation at the margin, although it can govern the use of *existing* work. A copyright puts in place a monopolist who controls the varieties in which a creative work might be available and used, as well as the prices at which the work will be offered. It can also create barriers to generic entry where cultural goods from long ago may not even be available in the marketplace due to the transaction costs of tracking down heirs of the long-dead creators.

This paper examines the effects of such IP rights on production and distribution when the incentive to create is not directly affected. It focuses on the 1998 Sonny Bono Copyright Term Extension Act, which set in place an abrupt change in IP protection at the year 1923. Works published after that date remain under copyright protection, whereas works from before that date have been in the public domain for several years or even decades. I use this discontinuity to estimate the effect of a copyright on the availability and prices of creative works. I quantify the welfare effects of the copyright extension in the book publishing industry, including whether or not a title is available at all, effects on variety and prices when the title is available, and the resulting impact on consumer surplus and profits.

The issues of copyright are particularly salient in the book publishing industry as book titles can be available in a wide variety of formats, from large-print or braille versions to audio books and, more recently, e-books. Many copyright-protected titles have owners who cannot easily be identified. They have gone out of print simply because nobody has been able to obtain the right to publish. These "orphan" works are unavailable to consumers in the market for new books, obviously creating a loss of surplus.[1] On the other hand, institutions such as Google Books and the Gutenberg Project have begun to make digital copies of public domain titles available for free online.[2] Google additionally began making snippets of protected works available, which in turn has led to a lawsuit by the Authors Guild and the Association of American Publishers against Google

[1] See the report by the Register of Copyrights at http://www.copyright.gov/orphan/orphan-report.pdf.

[2] Project Gutenberg is a nonprofit organization that offers e-book versions of public domain works for downloads free of charge, comparable in its function to Google Books.

in late 2005.[3] Eliminating the copyright protection of orphan works could make them available again in their entirety.

This paper asks two questions. First, how exactly did the Sonny Bono Copyright Extension affect availability and prices of, as well as demand for works that were affected by this policy change? Second, what are the welfare effects of these changes on the book publishing market? In answering these questions, I add to a growing literature which examines the impact of copyright protection by comparing books published before and after a certain cliff date. For example, Heald (2007) documents a large effect of the same extension on the availability of books at Amazon, and Li et al. (2015) find that an increase in copyright for works under the U.K. Copyright Act of 1814 substantially increases prices. Their results are suggestive of potentially large welfare impacts, but a full welfare assessment requires an analysis of demand in addition to a comparison of consumer and producer surplus under existing and counterfactual regimes. To this end, I estimate the demand for book titles and editions that are affected by the most recent copyright extension in addition to a regression discontinuity analysis of availability and prices.[4]

In order to measure the effects of the copyright extension, I collect information on all in-print and out-of-print editions of a list of 249 book titles from 1910 to 1936, including prices for a subset of these editions on Amazon over a 12-month period in 2011 and 2012, as well as hourly Amazon ranking data, which I use to infer the quantity sold.[5] I utilize the sharp discontinuity in copyright status around the year 1923 in a regression discontinuity design to find that works which have moved into the public domain are much more widely available than those whose copyright has been extended, with a copyright effect of 26.5 additional available editions per title. The differences in availability, variety, and costs are also reflected in the prices of these available editions, with editions of protected works costing up to 35 percent more than their public domain counterparts, although much of this difference is driven by competition across editions, rather than the copyright itself.

[3]The case has eventually been dismissed in November 2014 and again in the second circuit in October 2015 as the amount of text that Google has made available has been decided to constitute fair use, rather than copyright infringement. See, for example, Varian (2006), and Miller, C. C., & J. Bosman. 2013. "Siding With Google, Judge Says Book Search Does Not Infringe Copyright." *New York Times*. November 14. http://www.nytimes.com/2013/11/15/business/media/judge-sides-with-google-on-book-scanning-suit.html.

[4]A lack of data on both demand and supply has traditionally made it difficult to empirically estimate the welfare effects of a copyright. There is, however, some work analyzing the effects of a patent. For instance, Chaudhuri et al. (2006) estimate a large welfare loss in an antibiotics segment in India due to patent enforcement.

[5]I infer the quantity sold by assuming that each improvement in ranking corresponds to the sale of one book. Since the books in the data set are quite old and have lost much of their popularity over the decades, this approach is likely more exact than econometrically motivated models which assume a certain distribution of sales over rankings. See section 2 for more details.

The increase in availability and decrease in prices do not necessarily imply an increase in total surplus. It is possible that an increase in the level of competition, even when combined with lower costs of entry, can lead to an over-provision of a good. While profit maximizing firms will enter a product as long as profit gains are larger than the costs, some of these profit gains may simply be due to business stealing from another version of the book. If firms do not take this externality into account, there may be too many firms. This idea of excess entry has been highlighted in Mankiw and Whinston (1986) and Dixit and Stiglitz (1977), and further supported empirically in several industries (for example, Berry and Waldfogel, 1999; Thomas, 2011; Berry et al., 2015).

I test for the welfare implications of the copyright extension by estimating a nested logit discrete choice model of demand for book editions to determine price elasticities and preference parameters for variety, as well as to obtain estimates of a work's creative quality (the quality of the title, the creative work, itself). For instance, titles like *Gone With the Wind* or *Brave New World* have high creative qualities. They will be produced regardless of their IP status, whereas titles of very low creative quality will not be made available. For titles whose quality lies in between these two types, the copyright status should significantly affect their availability and consumption. Of course, the impact of the extension on consumer surplus depends on this creative quality as well.

Finally, I compare the changes in consumer surplus and profits. I find that a removal of copyright can result in excess entry for two types of works: high-quality titles which still generate a lot of profit for the copyright holder, and low-quality works which are already available in a wide variety, so that additional entry dissipates profits without increasing consumer surplus further. For all other titles, free entry increases welfare. A large majority of the titles – including many orphan works – would benefit from a move into the public domain in the sense that the increase in consumer surplus exceeds the decrease in the copyright holder's profits. Overall, given no change in incentives to create new ideas, this implies a net decrease in surplus from the book industry of $54 million each year.

This paper relates to literature which examines the effect of generic entry in the pharmaceutical industry, both in terms of prices (Regan, 2008; Berndt and Aitken, 2011) and in terms of entry (Morton, 2000) and profitability (Reiffen and Ward, 2005). The implications of my results reach beyond the setting of the book industry because editions are in effect forms of subsequent innovation. A copyright is narrow enough that an innovation would be covered by a new copyright, but such innovation (in the case of books, new study notes or a foreword) would not happen without the ability to reuse the original work. For example, Williams (2013) finds that IP rights on existing

technology significantly hinder subsequent innovation using human genomes, and Nagaraj (2013) shows that copyright hurts the reuse of information from the Baseball Digest. Moreover, some of these obstacles to follow-on creation can emerge from higher prices (Biasi and Moser, 2016). Similarly, my results indicate that long-lasting IP rights may not be an effective institution for producing cumulative knowledge (Furman and Stern, 2011; Handke, 2011).

2 Copyright in the United States

The British Statute of Anne (1710) set a length of copyright protection of 14 years. The United States originally followed this statute, but gradually extended the copyright term over the next decades and centuries. First, the Copyright Act of 1790 made the 14-year term renewable for an additional 14-year term if the author was still alive, and several extensions followed since then. By 1909, both terms had doubled to 28 years, so that a copyright could last for 56 years. However, the 56-year term only applied to a small fraction of works because registering and especially renewing a copyright often may not have been worth the effort. Copyright law was overhauled again in 1976, when new works by individual authors were granted copyright protection for 50 years after the author's death, which follows the guidelines of the Berne Convention. For works which had already been published at that time, the maximum term of protection was increased from 56 years to 75 years after publication.

Most recently, the 1998 Sonny Bono Copyright Term Extension Act lengthened copyright protection for cultural goods which were still protected at the time by another 20 years – from 75 years (or "life of author plus 50 years" for noncommercial works originally published after January 1, 1978) to 95 years (or "life of author plus 70 years"), creating a cliff at the year 1923. Works created before that date have been in the public domain at least since 1998, whereas works first published in 1923 and later remain under copyright protection at least until 2018, provided that the copyright had been registered and renewed in time.[6] Barring further copyright extensions, works will start moving into the public domain again in 2018, and understanding the implications of such moves is important.

The copyright extension provides exogenous variation in the cost of publishing a book today. If a book title is in the public domain, anybody who wants to publish it can do so without having

[6]See http://copyright.cornell.edu/resources/publicdomain.cfm for more detail. The Sonny Bono Act is also known as the Mickey Mouse Protection Act as it keeps Walt Disney's Mickey Mouse character, which would have moved into the public domain in 2003, under copyright protection until 2023.

to obtain permission. If a work is protected by copyright, on the other hand, the publisher has to obtain a license from the copyright holder in order to publish the work.[7]

When a protected title is published, a contract is set up to determine a royalty for each individual copy sold on top of an advance to the author. The advance most often ranges between some ten thousand and hundreds of thousands of dollars, and it can amount to over a million dollars for the most promising works.[8] Contracts for recent works almost always include exclusive publishing rights for the publisher in the country of publication, essentially transferring the copyright to the publisher, who becomes a multi-product monopolist for that title. While one would suspect this is similar for older protected books (such as those published in the years after 1923), many of the titles in my dataset have been published by – and are still in print today through – multiple publishers.

3 Data

To examine the effect of the copyright extension on the variety and use of products of a particular title, I consider a set of book titles which can a priori be regarded as similar: the annual ten bestselling fiction titles of the years 1910 to 1936.[9] Although this is neither an exhaustive list of high-quality titles from the time period nor of the most popular works given today's demand, it is made up of titles whose authors had strong incentives to register and later renew a copyright. Further, it provides a wide variety in terms of demand today. Several titles still face positive demand, while others have become orphan works – works whose copyright holder cannot be found and which have moved out of print as a result. Still, note that even if these orphan titles are no longer available as new editions, they might be available in the used-books market.

Upon publication, an edition of a book title is assigned an international standard book number (ISBN). An ISBN uniquely identifies a book's title, its publisher, date of publication, its format (for example, hardcover, paperback, and e-book) and suggested retail price, in addition to other characteristics such as the number of pages, the physical dimensions, the font size, and forewords

[7]Without the extension, protected creative works would move off-copyright every year, which would allow me to observe individual titles under both copyright regimes. However, this within-title variation can lead to biased estimates of both supply and demand: Publishers may wait to publish an edition of a title that is close to moving into the public domain to avoid the costs associated with the copyright holder, and consumers may wait to buy the title because they would expect more editions and lower prices in the near future.

[8]Advances have increased significantly in recent years. See Deahl, R. 2014. "The Rise of the Seven-Figure Advance." *Publishers Weekly*, November 21. http://www.publishersweekly.com/pw/by-topic/industry-news/book-deals/article/64848-the-rise-of-the-seven-figure-advance.html. Peukert and Reimers (2017) show that these increases can be attributed to the rise in self-publishing.

[9]These can be found on several websites, but Michael Korda's Book, *Making the List: A Cultural History of the American Bestseller 1900-1999*, was used as a reliable reference.

and text notes. Each title can be available in several ISBNs, or editions. The underlying dataset consists of 249 titles, although not all titles are still in print.[10] Of the 249 titles, 37 are out of print in 2013, but 34 of those are available as used books at Amazon.

I create two datasets for the analysis. First, I use the Bowker Books-in-Print directory to obtain characteristics of all editions that are available for each title. Second, I collect sales data and retailer specific edition characteristics, including the actual sales price and available conditions (new and/or used), from Amazon and Project Gutenberg between September 2011 and August 2012. I supplement these datasets with aggregate information on the size of the publishing industry and average wholesale prices from the Book Industry Study Group BookStats report of 2011 and from Publishers Weekly periodicals.

Information about each title's availability through the Bowker Books-in-Print directory was collected in August 2013 – five years before any title's copyright status would change. The directory includes the issue date, the current in-print status, price, publisher, format and some additional characteristics (whether it is a collector's edition, the language of the text, and so on) of each ISBN that has been published since 1948. On average, 18.2 editions of a title in my list are available and in print in 2013. The number of editions varies substantially across titles and copyright regimes, with several (protected) works having moved out of print entirely, while one public domain title (Edith Wharton's *The Age of Innocence* (1921)) still has 146 different in-print editions.[11]

In order to measure the sales of each edition, I use the website novelrank.com to collect hourly Amazon sales rankings of a subset of all ISBNs and title-format combinations over the time period of my study (2011 to 2012). Among researchers facing rankings, there is an active tradition of translating sales ranks into quantities (see Chevalier and Goolsbee, 2003; Brynjolfsson et al., 2003; Reimers and Waldfogel, 2017) by assuming that sales-rank relationships tend to obey power laws. My set of titles does not require that I make assumptions about sales distributions. Because the quantity demanded of the titles in my dataset is low (only a few units per month), an improvement in the ISBN's ranking over the previous hour can be interpreted as the sale of one single unit because multiple sales in one hour are highly unlikely. Figure 1 shows daily rankings over a ten-month period for the 1961 Signet paperback edition of Arrowsmith by Sinclair Lewis. From the graph, I infer that this edition was sold 12 times through the Amazon platform during this time

[10] Some of the titles made the top ten best seller lists in multiple years. A book is out of print if it is no longer available through the publisher as a new book.

[11] I only count those editions that are still in print in 2013 in order to avoid double counting of editions which replace their predecessors.

period because there are 12 discrete improvements in rankings.

Table 1 summarizes the characteristics of the editions I follow on Amazon in the first month of data collection (April 2011), both for public domain titles (published before 1923) and for protected titles (after 1923). I observe a total of 722 editions on Amazon and 120 editions on Project Gutenberg.[12] The table shows that public domain works and protected titles differ from each other on several dimensions. First, the editions I observe are significantly more likely to be available as new when the title is in the public domain. Second, the public domain editions on Amazon tend to be published more recently, are less likely to be published by a major publisher, and are more likely to be in paperback format than the editions of comparable protected titles.[13] Many public domain editions were published after 2000, and these editions are disproportionately published in paperback format by independent publishers. Whether these editions are welfare increasing is an empirical question.

Finally, a copyright affects the price of editions by changing the cost structure, which can in turn impact the quantity demanded. Public domain titles are indeed less expensive than protected titles, although the difference is barely statistically significant at the ten percent level. On the other hand, editions of protected titles are sold more often on Amazon than editions of public domain titles, although this is could be due to zero-price competition from Project Gutenberg and Google Books. An average Project Gutenberg edition on my data set is downloaded 118.3 times in April 2011.

4 The Effect on Availability and Price

Table 1 suggests that editions of protected titles are different from their public domain counterparts, but a causal effect of a copyright on availability and prices is not yet established. To this end, I take advantage of the fact that the copyright extension in 1998 did not affect the incentives to create a work in the 1920s. In fact, there was no change in the copyright term between 1909 and 1962, so that all works which were published between 1910 and 1936 (the range of years in my dataset) were subject to the same copyright term of 28 years at the time the works were written and published, plus 28 years if the copyright holder renewed the term.

[12] I do not observe the universe of all available editions but rather those which were shown on the first page of the search results when I began collecting this information in April 2011. This approach provides horizontal differentiation within a title-format combination – an important assumption in the demand and welfare estimations – while I can use the Bowker Books-in-Print directory to infer entry of editions.

[13] These characteristics and patterns are consistent with the editions on the Bowker dataset – the universe of in-print editions.

The 1998 copyright extension provides a sharp, exogenous discontinuity in today's copyright protection status at the year 1923, whereas no discontinuity is expected in the inherent quality and appeal of the works. All titles are fairly "old" (published between 80 and 100 years ago), and all are fiction novels which were fairly successful in the United States upon their original publication. Yet, titles which were originally published before 1923 are in the public domain, while more recent titles have been "treated" and are still protected. I use a title's original publication year as the forcing variable in a regression discontinuity design (RDD), where the title is treated – currently protected by copyright – if the publication year is 1923 or later.

Following, among others, Imbens and Lemieux (2008), the equation

$$Y_j = X_j'\beta + \alpha \times 1\{IP_j = 1\} + k(year_j) + \epsilon_j, \tag{1}$$

where X_j includes characteristics of title (edition) j and $k(year_j)$ is a continuous function of the year of original publication, provides information on the immediate effect of a copyright on the outcome variable Y_j through the coefficient α. As dependent variables, I examine the number of in-print editions in each format as obtained from Bowker, and the 2012 price of these editions on Amazon. In order to estimate the effect of a copyright more precisely, I tighten the bound around 1923 to include only those bestsellers which were published between 1915 and 1930 – seven years before and after the cliff year.[14] I control for a work's appeal by including British library checkouts between 2010 and 2012 (PLR – Public Lending Rights).[15] In addition, I include an indicator variable that equals one if the author has won a Pulitzer Prize, and indicator variables that are turned on if the book (or other books by the author) is listed in Harold Bloom's *Western Canon: The Books and School of the Ages* and its appendices.[16]

4.1 Variety of Editions

Table 2 shows the coefficients from equation 1 when the dependent variable is the number of in-print editions of a title. The effect of copyright protection on the number of editions is significantly negative across formats. Protected titles are available in 26 fewer editions in 2013 than similar titles

[14] Omitting the first few years ensures the copyright specific incentives were constant when writing the books, as a change in the copyright term occurred in 1909. The results are robust to including all titles.

[15] Authors in Britain receive royalties on library check outs, so that this information is available there. The British copyright law sets the copyright term at 70 years after the author's death or the publication of the work, so that no discontinuity at the year 1923 is expected.

[16] See http://www.openculture.com/2014/01/harold-bloom-creates-a-massive-list-of-works-in-the-western-canon.html for a list of the titles included in the Canon.

which are in the public domain (column 1). That is, if a currently protected (old) work moved into the public domain today, over 25 new editions would be published (and remain in print) in the next few years. Figure 2 illustrates this discontinuity in available editions at the cliff year 1923.

Table 2 also identifies the effect on a title's availability in each individual format. The largest difference is in the number of paperback editions, where a move into the public domain corresponds to 15 more editions being published. This supports the evidence from table 1, which shows that public domain editions were published more recently, are less likely to be published by major publishers, and more likely to be paperback editions. Paperback editions are cheaper to produce than hardcover editions and require less technical knowledge than the publication of an e-book. The remainder of the difference is split between hardcovers (5.4 editions) and e-books (2.9 editions).[17]

In addition to being available in fewer editions, protected titles are less likely to still be in print at all. Of the 124 protected titles in my data set, 33 are out of print, whereas all 125 of the public domain works are still in print and hence available as new books.

The remaining coefficients are sensible as well. Treating the quantity demanded through British libraries (PLR) as an indicator of quality, its positive effect is expected, as is the positive effect of being a Pulitzer award winning author. The large positive effect of being included in the Western Canon of Literary Works is explained by the fact that only two works in the data set are included in the Canon: Edith Wharton's *The Age of Innocence* (146 editions) and Sinclair Lewis' *Babbitt* (117 editions). The results are virtually unchanged if I remove these two titles from the estimation.

4.2 Identification

The causal effect of the copyright extension on availability is identified if there is no discontinuity in the titles' other observable (and unobservable) characteristics. In theory, very little difference can be expected between the two groups of titles in the dataset except for the fact that works published before 1923 have moved into the public domain. In particular, there is no discontinuity in the titles' "qualities" at the year 1923. Some of this can be seen in figure 3, which shows that while the titles which were written more recently are checked out more often at British libraries, this seems to reflect a gradual progression, rather than an abrupt change, with more recent titles becoming more popular. This change could be due to the fact that people outside of the United

[17]The effect of the extension on variety is robust across the titles' creative qualities as well. Considering only the most popular works (the top 25% in terms of their unit sales in 2012), there is a drop of around 40 editions per title at 1923, whereas there is a drop of close to 15 editions per title for the least popular works. Repeating the analysis with the log of the number of editions as the dependent variable, the copyright extension corresponds to about 80 percent fewer available editions today, compared to those titles which have moved into the public domain.

States can still find works which were originally published before 1923 online on American websites, even if they are protected by copyright in other countries. The graphs for the other explanatory variables in the regression look similar.

To isolate the effect of a change in copyright regimes, I examine the titles' availability and the editions' characteristics before any of the titles would have moved into the public domain by considering only those editions which were published in the first 74 years after the title's original publication. There is no discontinuity in availability, suggested retail prices, or the number of pages. Figure 4 shows that while more recent titles tend to publish more editions within the first 74 years, there is no discontinuity at the year 1923 (top two graphs). Instead, the general increase may be due to gradual improvements in distribution technologies over the years. At the same time, there is no change in suggested retail prices or in the number of pages per published edition (bottom panel). This provides further support for the identifying assumption that the bestsellers are similar ex ante.

4.3 Prices of Published Editions

There are two possible reasons for the large discontinuity in availability and variety: the difficulty of finding the copyright holder, and the costs of obtaining the right to publish, which includes an advance and a royalty. Both of these are IP specific costs which can affect the prices of individual editions. Prices can also be affected by the level of competition across editions within a title itself, with lower prices for editions facing a lot of competition. To determine whether the differences across copyright regimes indeed lead to differences in prices, I repeat the analysis from equation 1, using each edition's price (in logs) as found each month on Amazon as the dependent variable.[18] In addition to the title quality variables from above, I add edition specific variables such as the edition's age, whether it is published by a major publisher and whether it is available as new through Amazon to the specifications.[19] Finally, I include the number of in-print editions of the title and format to examine the mechanism behind possible price differences.

Table 3 displays the results of these regressions. Columns 1 and 4 include editions of all three formats, with indicator variables for paperback and e-book editions, while columns 2 and 5 show the results for hardcover editions, and columns 3 and 6 examine paperback books. Editions of

[18]The edition's price on Amazon is a more relevant measure than its suggested retail price because edition prices tend to vary over time, especially in an increasingly digital world.

[19]While these variables differ across copyright regimes, these differences are indirect consequences of the copyright extension, rather than characteristics which affect the probability of treatment. We see more entry, and this entry is by a certain type of editions.

protected titles are offered at a significantly higher price. Column 1 estimates a price increase due to copyright protection of 27.4 ($= e^{0.242} - 1$) percent. Columns 2 and 3 suggest that the price difference is mostly driven by paperback editions: the price of a protected paperback edition is 32.2 percent larger than that of a public domain paperback version, while the effect on hardcover editions is less pronounced, with a statistically insignificant increase of 14.2 percent.

These price differences could be due to cost decreases when the copyright expires, to increased competition within the title, or to a combination of the two. I examine these mechanisms in columns 4 through 6 by including a measure of competition: the number of in-print editions of the respective title-format. The effect of the copyright becomes smaller and less statistically significant in these specifications, with a copyright increasing the price of an edition by 12.8 ($= e^{0.121} - 1$) percent, rather than by 27.4 percent. Interestingly, the effect disappears entirely for hardcover editions, while paperback editions still see a statistically significant impact of a copyright. This suggests that competition explains part of the price differences, but that cost differences across regimes likely also play a role in determining prices.

5 Demand and Consumer Surplus

The regression discontinuity analysis sheds light on some interesting – and possibly unintended – effects of the Sonny Bono Copyright Extension Act. Titles which remain under copyright protection as a result of this extension are much less likely to be available than comparable books in the public domain. Even if the books are in print, the variety of protected works is quite limited, and available editions are more expensive.

While the differences in availability and prices are suggestive of significant welfare effects, it is possible that the additional editions are of low enough production quality that the higher variety and lower prices do not result in larger consumer surplus. At the same time, the demand for individual editions could be higher when the work is protected because each edition faces less competition. The higher profit potential could also increase the incentive to create a high-quality version with nice cover art, a foreword, and study notes.

However, the copyright itself would only have an indirect effect on the number of units sold – through the price, the level of competition, and the edition's production quality – because the titles in my data set were published under the same copyright law. Hence, a regression discontinuity design estimating the demand for a title or edition as a function of its copyright status would not

allow me to identify why one book might sell more than another. Instead, I employ a nested logit model similar to Berry (1994). This approach allows me to infer product qualities while also allowing for varying substitutability among products. I then use these estimates to determine differences in consumer surplus across copyright regimes.

5.1 Demand Model

Consumers choose whether to read a title in my dataset, and then choose which edition of the title to buy. Define J_{wt} as the set of editions available for title (creative work) w in month t, and define the set of all editions across all titles as J_t. An edition j of title (creative work) w is published in one of three formats, denoted by $k \in \{H, P, E\}$, where H = hardcover, P = paperback and E = e-book. For each public domain title, there is one additional format, denoted G, as the book is available through the Gutenberg website. Each consumer then decides in each month whether to buy an edition j of title w, to consume an edition of a different title, or to consume the outside good (not purchasing a book from the choice set). Suppressing the time subscript, a consumer i chooses j from the $J+1$ options that maximizes her indirect utility function given by:

$$\begin{aligned} u_{ij} &= X_j'\beta - \alpha p_j + \phi_w + \xi_j + \zeta_{iw} + (1-\sigma)\epsilon_{ij} \\ &= \delta_j + \zeta_{iw} + (1-\sigma)\epsilon_{ij} \end{aligned} \tag{2}$$

where δ_j is the mean utility from purchasing edition j. The vector X_j includes the edition's age (in months) and format, whether it is available as new, whether it is published by one of the major publishers, and the perceived quality of the book itself.[20] Further, ϕ_w is title w's overall appeal, ξ_j is an unobserved (to the econometrician) quality of edition j, ϵ_{ij} are consumer i's idiosyncratic taste shocks for edition j, and ζ_{iw} represents the idiosyncratic taste shock common to all editions within title w. If ϵ_{ij} and ζ_{iw} follow a type 1 extreme value distribution, then the term $\zeta_{iw}+(1-\sigma)\epsilon_{ij}$ follows a type 1 distribution as well, and the parameter σ measures the correlation of tastes across editions within a title (see Cardell, 1997).

Given the functional form of this model and the idiosyncratic shocks, the market share of each edition follows easily as $s_j = \frac{\exp\{\delta_j/(1-\sigma)\}}{D_w^{\sigma}(1+\sum_{w' \in W} D_{w'}^{1-\sigma})}$, where $D_w = \sum_{j \in J_w} \exp\{\frac{\delta_j}{1-\sigma}\}$. Setting the mean utility from consuming the outside option to zero and inverting the δ_j from observed market shares

[20] In the estimation I approximate this by whether or not a picture is seen on the book's Amazon site.

yields

$$\ln(s_j) - \ln(s_0) = \delta_j + \sigma \ln s_{j|w}, \tag{3}$$

where s_0 is the share of the outside option and $s_{j|w}$ is the share of edition j within title w.

5.2 Estimation and Results

Two of the explanatory variables are affected by the same unobserved factors as the left hand side variable: the edition's price, and the share of an edition's sales within its title. As instruments for the price of an edition, I utilize the number of sellers for a specific edition as a proxy for the ease of distributing the title, and I use the numbers of editions offered on the platform within a title (and format) to instrument for the editions' inside shares. These instruments can give insight into the costs of distributing a title as well as the edition's share within the title and format, but they are not good predictors of an edition's market share across all titles and formats.

Table 4 reports the results of several specifications of equation 3. The first two columns show the results from a logit model, where column 2 introduces an instrument for the price of the edition. Column 3 represents the preferred specification introduced in section 5.1. The correlation of tastes for editions within a title – σ(title) – is large and statistically significant at 0.629, suggesting that the regular logit model is misspecified and that preferences for editions within a title are quite strongly correlated. This specification further suggests a relatively large mean price elasticity of -3.7, and it indicates a strong preference for the free Project Gutenberg editions as the coefficients on the other formats are large and significantly negative. All other coefficients are as expected. Consumers prefer new editions over used ones, and more recent editions over more dated ones. The coefficient on the indicator for a major publisher is negative but insignificant at the 10% level.

Column 4 adds another level to the nesting structure, allowing for preferences to be correlated across editions within a title-format combination as well as across formats within a title. While the correlation of tastes across formats within a title (σ(title)) remains statistically significant, the correlation within title-format combinations (σ(title-format)) is less precisely estimated, presumably because there are several title-format combinations for which I only observe one edition. While the coefficients remain similar for all coefficients, I continue with the results from column 3.

The title fixed effects provide two insights about each title's creative quality – its inherent appeal in the twenty-first century. First, there is no significant difference in the size of these title fixed effects across copyright regimes, suggesting that publishers create editions with similar production

qualities ξ_j across the copyright regimes as well, at least on average. The argument by Adilov and Waldman (2013) that a longer copyright term triggers more ex-post investment in the work finds little to no evidence in my set of fiction novels. Still, a larger variance in creative qualities across protected titles could imply that publishers of "good" protected titles could be willing to create particularly nice editions to extract some monopoly rent, while "bad" protected titles receive less attention.

Second, conditional on the title's copyright status, higher-quality titles tend to be published in much larger varieties, particularly for public domain works. Figure 5 shows the relationship between a title's creative quality ϕ_w (the fixed effects) and the number of available editions, for both copyright regimes.[21] Not surprisingly, the differences in availability are most pronounced at the top of the quality distribution, although public domain works are available in wider varieties throughout that distribution.

5.3 Consumer Surplus

Empirically, a move into the public domain makes titles available in a wider variety and lowers the retail price of each edition. Both of these effects should increase consumer surplus, but the size of this effect is unclear. The estimated σ, α, and mean utilities for each edition, format and title from the demand estimation allow me to calculate these changes in consumer surplus between the two copyright regimes. Here, I compare the consumer surplus from currently protected titles to that generated from public domain titles with similar creative qualities.[22] In my application, the difference in consumer surplus (to the representative consumer) is

$$\triangle E[CS] = \frac{M}{\alpha}\left[\ln\left(1+\sum_{w'\in\mathcal{W}^O}\left[\sum_{j\in J_w^O}\exp\left(\frac{\delta_j}{1-\sigma}\right)\right]^{1-\sigma}\right)-\ln\left(1+\sum_{w'\in\mathcal{W}^I}\left[\sum_{j\in J_w^I}\exp\left(\frac{\delta_j}{1-\sigma}\right)\right]^{1-\sigma}\right)\right], \tag{4}$$

where the superscripts describe the choice sets under examination: I represents IP protection, and O denotes the public domain.

In order to report the changes in consumer surplus, I need to address two features of the data.

[21] In this figure, I normalize the works' creative qualities to be distributed between zero and ten for illustrative purposes.

[22] The nature of the data – two data sets with different prices which cannot easily be combined – prevents me from estimating an entry model directly. In my approach, I implicitly assume that the edition managers of the public domain works made optimal entry decisions, and that edition managers of currently protected titles would act the same way. Differences in consumer surplus are realized as a consequence of the observed differences in entry behavior and prices.

First, I only observe demand for a subset of all in-print editions, but consumer surplus should be derived from all available editions. I therefore expand the number of editions for each title-format to the number of in-print editions as in the Bowker database. I assume the added editions have average retail characteristics for that title-format combination.

Second, I only observe the quantities demanded through two distribution channels: online retail through Amazon, and downloads through Project Gutenberg. Any consumption through brick-and-mortar bookstores and libraries, or downloads of free editions through Google Books or Amazon, are not observed. A Bowker study determined that Amazon accounted for 20% of the publishing market in quarter two of 2011, and for 27% in the second quarter of 2012.[23] This number is likely even larger for old fiction titles that are difficult to find in physical bookstores. I assume that Amazon and Project Gutenberg each capture a share of $\gamma = 0.5$ of the total sales and downloads of my list of editions. That is, I multiply the observed and estimated quantities by two to cover the entire market.[24]

Per-title Consumer Surplus

Most titles in my data set do not generate a lot of consumer surplus regardless of their copyright status. This is not surprising, as many titles only sell a few copies each month. Still, public domain titles generate much more consumer surplus than their protected counterparts. On average, a public domain title in my data set generates close to $9800 in consumer surplus, compared to just over $2000 for protected titles. The differences are due in large part to the differences in prices and to the presence of the popular zero-price format of Project Gutenberg e-books.

The titles with the largest creative qualities generate the largest consumer surplus, and consumers stand to gain the most from such titles moving into the public domain. For example, Edith Wharton's *The Age of Innocence* generates over $148,000 in consumer surplus per year with its 146 different editions at an average price of $4.75, while Margaret Mitchell's *Gone With the Wind* – the highest-quality protected title – generates just close to $27,000 in consumer surplus with its 13 editions and an average price of $11.68.

On the other hand, several high-quality protected titles are available in rather large varieties, generating relatively large consumer surplus as well. For instance, Pearl Buck's *The Good Earth*,

[23] See Publishers Weekly; 11/5/2012, Vol. 259 Issue 45, p6-6. Much of the increase in Amazon's market share can be attributed to Borders going out of business.

[24] If Amazon accounts for more than 50%, the reported results provide an upper bound for the changes in consumer surplus, and vice versa if Amazon accounts for a smaller share.

which is available in 35 different editions, generates almost $50,000 in consumer surplus. Of course, a large number of low-quality titles are hardly sold at all and consequently generate very low consumer surplus whether they are protected by copyright or not.

Another way to measure the effect of the copyright extension is to look at the relative differences in consumer surplus across the copyright regimes. On average, the highest-quality protected titles generate between 70 and 80 percent of the surplus that their public domain counterparts generate, whereas the lowest-quality protected books merely create a fourth of the consumer surplus from low-quality public domain titles, even though these differences are much smaller in absolute terms.

Industry Consumer Surplus

The copyright extension impacts consumer surplus from two types of titles: those which are in print regardless of their copyright status, and those orphan works which are currently out of print but would move back into print if the copyright was removed.[25] Both types of books might affect consumer surplus differently. Thus, to estimate the industry-wide welfare effect of the copyright extension, I collect information on the number of titles which are still protected by copyright today, the number of titles which are currently in print, and of those which would move back into print if they moved into the public domain.

Finding this information requires some detective work. Between 1923 and 1936, an average of 12,686 books were entered into the copyright registry each year, with a slight increase over the years.[26] For the same years, an average of 4,513 copyrights for "books and contributions to periodicals" were renewed after 28 years, and the number of works which are still protected today ranges from 3,208 per year of origin (originally published in 1923) to 5,585 (1936).[27] It is unclear how many of those renewed copyrights pertain to books.

Information on the number of in-print and available books today is more difficult to come by, but some hints are available. For instance, 174 of over 10,000 books originally published in 1930 were still in print in 2001.[28] Based on this, I assume that the number of titles which are still in print is between 100 and 250 per year of original publication.

At the same time, Project Gutenberg offers over 47,000 public domain titles as of November

[25]Obviously, books which do not become available regardless of their copyright status have no impact on consumer surplus.

[26]See the Internet Archive's Copyright Records at https://archive.org/details/copyrightrecords.

[27]See, for example, the University of Pennsylvania Catalog of Copyright Entries at http://onlinebooks.library.upenn.edu/cce/.

[28]See the *American Library Annual and Book Trade Almanac for 1872-1957.*

2014. That is, Gutenberg has approximately 300 titles per year of publication, assuming that the first documents were originally published in 1776 (the Declaration of Independence was Project Gutenberg's first digitized document) and the most recent titles are originally from 1922. Of course, Gutenberg also carries works from before 1776 and after 1923, and most titles are from the early 20^{th} century, which leads me to assume that between 100 and 400 additional titles would move back into print if the copyright term were decreased by one year.

As seen above, titles of different creative qualities are affected differently by a change in copyright status. An assessment of the industry-wide effect of the copyright extension therefore needs to take the distribution of qualities – and of the consumer surplus effects – across all titles into account. I use the set of public domain works in my data set to create a one-to-one mapping of Gutenberg download counts into consumer surplus changes as calculated above. I then collect Project Gutenberg download counts for 30,000 randomly selected works to determine an industry-wide distribution of these consumer surplus effects based on the distribution of downloads.[29] The distribution is created quite simply: For each title not in the original data set, I assign the surplus change of the next lowest observed download count among the original titles. For example, consider two titles in the original data set, one of which has 100 downloads and the other has 200 downloads. If another title has anywhere between 100 and 200 downloads, I assign it the same consumer surplus effect as the 100-download title. For any titles with more than 200 downloads, I assign the 200-download book's surplus change.[30]

Table 5 shows that depending on the number of affected works, decreasing the copyright term by one year would increase consumer surplus by between $1.5 million and $4.3 million annually. If 175 titles are currently in print, and 200 become newly available if in the public domain (my best guess), consumer surplus increases by about $2.7 million per year. Of course, then, a 20-year extension like the 1998 Copyright Term Extension Act, which will have prevented works from 20 years (1923 to 1942) from moving into the public domain by 2018, now leads to a decrease in consumer surplus of more than $54 million each year.[31] Adding the effect on movies and music would likely increase this impact significantly.

[29] See http://gutenberg.readingroo.ms/cache/generated/feeds/ for download counts. The numbers are from April 2014.

[30] This simple measure has two advantages. It allows me to remain conservative in my calculation of the extension's welfare effect, and it does not impose any structure on the title quality distribution which cannot be verified anyway.

[31] A less conservative approximation of the quality distribution, in which titles with between 100 and 200 downloads would be assigned the same surplus change as the 200-download title, would lead to a surplus increase of $66 million due to the 20-year extension.

6 Beyond Consumer Surplus: Entry and Profits

The above differences in consumer surplus are obviously inefficiencies due to the copyright extension: as a result of higher prices and less availability, consumers are worse off if a title is protected than if the same title has moved into the public domain. Of course, proponents of strict IP laws argue that IP provides an incentive to create new works, and policymakers take into account these incentives to create. In the case of a copyright, and in particular in the case of the 1998 Copyright Term Extension Act, these incentives to create are unlikely to be affected. In addition, with impacted books being at least 75 years old, it is unlikely that the original creator of the idea still receives any profits from the extension. As a result, consumers are worse off, no incentives are created, and the person who benefits from the extension did little or nothing to receive such benefits.

Still, policymakers may want to compare the loss to consumers with the gains to the current copyright holder to examine the overall welfare effects of the extension. It is possible that the losses to consumers are outweighed by the gains to the suppliers if fixed costs of additional entry are large compared to the gains from variety. I examine this possibility here by examining publisher and copyright holder profits more closely.

6.1 Supply Model

As intuition for the calculations of the changes in profits, I introduce a model of free entry in two stages. The idea is simple: In the public domain, free entry implies that firms introduce an edition as long as the expected profit is positive, and they do not enter if the expected profit is negative. Then, the marginal edition's operating profit should be close to its fixed costs in expectation. Under copyright protection, the mechanism could be much more complicated, but the costs of production are likely similar. Taking as given the fixed costs of production from public domain works of similar quality, the observed prices, and the inferred marginal costs, I can calculate the profits for protected editions.[32]

Formally, public domain firms compete with each other in a two-stage entry game. In the first stage, they choose whether or not to enter an edition of a title, and they set edition prices in the second stage. In the second stage, firms (publishers) compete in Bertrand fashion.[33] The markups

[32]These profits are likely shared between the copyright holder and the edition managers. From a welfare perspective, who gets these profits is irrelevant, although perhaps publishers would invest their profit to develop more new ideas.

[33]This assumption is consistent with Mankiw and Whinston (1986), who take stage 2 as non-cooperative in the social planning problem of entry.

and marginal costs are given by the Lerner index:

$$\frac{p_j - c_j}{p_j} = -\frac{1}{\epsilon_j}, \tag{5}$$

where $\epsilon_j = \frac{\alpha p_j}{1-\sigma}\left[1 - \sigma s_{j|w} - (1-\sigma)s_j\right]$ is the price elasticity, and α and σ are obtained in the demand model. In stage 1 of the game, firms enter as long as their expected profits are positive. Given the prices in stage 2, an edition's variable profits translate directly into fixed costs of producing edition j in the public domain under free entry.[34] That is,

$$\pi_j = (p_j - c_j)Ms_j, \tag{6}$$

where $s_j = \frac{\exp\{\delta_j/(1-\sigma)\}}{D_w^{\sigma}(1+\sum_{w' \in W} D_{w'}^{1-\sigma})}$ as in the demand model.

These costs of production likely do not depend on the work's copyright status, especially since the estimated creative qualities are not significantly different across the two copyright regimes. They do, however, depend on the edition's format. For example, marginal costs are essentially zero for e-book editions – ignoring the royalties – while they are positive for physical formats. Then, determining the (joint) profits to the copyright holder and publishers is easy: it is the difference between variable profits for editions of in-print protected works, and the estimated fixed costs of production as obtained from their public domain counterparts.

6.2 Calculation and Profits

Given the fixed costs of production, the profits from a protected title depend on the number of editions N_k that the title is available in for each format k, and the operating profit (adding back the royalties) per edition.

While I set marginal costs for public domain e-books to zero, marginal costs for hardcover and paperback editions are obtained using the Lerner markup rule, given each edition's price elasticity. Once the marginal costs for protected titles are found, I subtract likely royalties of 10% of the price of the book.[35] The average marginal costs are then \$17.77 for hardcover editions, and \$9.79 for paperback editions.[36]

[34] I relax the integer constraint here. Moreover, I assume that there is no sunk cost of publishing a new edition because technological innovations are not needed for publication. Instead, I treat the fixed cost as a flow which occurs every year, similar to Eizenberg (2014) and Fan and Yang (2014).

[35] Greco (2013) lists a standard royalty rate of 10% to 15% of the price (page 156).

[36] Note that high marginal costs would imply low fixed costs, so that the size of the marginal costs does not affect the resulting profits much.

Marginal costs are similar for public domain works. These titles provide the fixed costs of production, which are quite low, as expected. For hardcover and paperback editions, average annual fixed costs are \$334 and \$375, respectively, and for e-books, they are calculated as only \$15.[37] With these marginal and fixed costs in mind, title profits are calculated as

$$\pi_w = N_H \times E[\pi_{wH}] + N_P \times E[\pi_{wP}] + N_E \times E[\pi_{wE}], \tag{7}$$

where $\pi_{wk} = (p_{wk} - mc_{wk})Ms_{wk} - FC_{wk}$ with marginal and fixed costs from the public domain, and the expectation is the average per-year profit over all editions within title-format wk.

Per-Title Profit

Much like on the consumer side, most titles generate very little profit whether they are protected by copyright or in the public domain. Yet, there are some titles which generate large profits for the publishers and the copyright holder while protected by copyright. Most notably, Pearl Buck's *The Good Earth* still generates joint profits to the copyright holder and producers of \$105,000 annually. Several other titles, including Margaret Mitchell's *Gone with the Wind* and Thornton Wilder's *The Bridge of San Luis Rey* earn between \$10,000 and \$30,000 annually.[38]

On average, the (in-print) protected titles generate a profit of \$4027, compared to zero profits with free entry in the public domain. These \$4027 are significantly smaller than the average differences in consumer surplus of \$7800. Most titles in my data set would generate more total surplus in the public domain than if they were protected. An industry-wide projection of these profit differences like the one in section 5.3 shows that profits decrease by less than 25% of the increase in consumer surplus when titles move into the public domain, suggesting large welfare losses of the copyright extension.

Still, some creative works – for example, *The Good Earth* – generate more surplus if they remain protected by copyright. The owners of these types of works which stood the test of time (including Disney's *Mickey Mouse*) have been able to lobby for copyright extensions even though the negative effect of the extension on consumers outweighs the positive effect on producers on average. Since an extension of a copyright term beyond 70 years can hardly change incentives for creation, the 1998 Copyright Term Extension was welfare decreasing unless the copyright holder or publishers

[37] If these costs are "too low," then my measures of profit for protected titles can be seen as an upper bound of true profits. As we will see, my estimates of profits are still quite low.

[38] It is likely that *Gone with the Wind* generates much more profit as a movie.

used the added profits for further innovation which would have otherwise been impossible.

7 Conclusion

When the Sonny Bono Copyright Extension Act was passed in 1998, a move into the public domain implied that publication of the work became cheaper. This paper shows that we would have seen an increase in the number of editions per title and format and a decrease in prices. Today, moving a work into the public domain would not only increase the number of *editions* per title-format, but it would also increase the number of *formats* by making the work available for free through Project Gutenberg and Google Books. As a result, consumer surplus would increase much more than producer surplus would decrease when a title moves out of copyright protection.

The copyright extension of 1998 decreased welfare from the publishing industry. This is true especially for the large stock of titles that have become orphans. The Sonny Bono Copright Term Extension, also known as the Mickey Mouse Protection Act, increased profits to creators and owners of works that are still popular today (such as Disney's Mickey Mouse), but it decreased total surplus from the majority of works: the low and medium quality works that have lost much of their appeal since their creation.

A policy which could weed out those titles with the biggest stakes in extending their copyright while allowing other titles to move into the public domain would obviously increase total welfare more. Since such a policy would be difficult to implement, any further increase of the copyright term would likely be welfare decreasing.

References

Adilov, N. and Waldman, M. (2013). Optimal copyright length and ex post investment: a mickey mouse approach. *Economic Inquiry*, 51(2):1101–1122.

Berndt, E. R. and Aitken, M. L. (2011). Brand loyalty, generic entry and price competition in pharmaceuticals in the quarter century after the 1984 waxman-hatch legislation. *International Journal of the Economics of Business*, 18(2):177–201.

Berry, S., Eizenberg, A., and Waldfogel, J. (2015). Optimal product variety in radio markets. Technical report, National Bureau of Economic Research.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Berry, S. T. and Waldfogel, J. (1999). Free entry and social inefficiency in radio broadcasting. *RAND Journal of Economics*, 30(3):397–420.

Biasi, B. and Moser, P. (2016). Effects of copyrights on science: Evidence from the wwii book republication program. *Available at SSRN 2542879*.

Brynjolfsson, E., Hu, Y., and Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, 49(11):1580–1596.

Cardell, N. S. (1997). Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, 13(02):185–213.

Chaudhuri, S., Goldberg, P. K., and Jia, P. (2006). Estimating the effects of global patent protection in pharmaceuticals: a case study of quinolones in india. *The American Economic Review*, 96(5):1477–1514.

Chevalier, J. and Goolsbee, A. (2003). Measuring prices and price competition online: Amazon. com and barnesandnoble. com. *Quantitative marketing and Economics*, 1(2):203–222.

Dixit, A. K. and Stiglitz, J. E. (1977). Monopolistic competition and optimum product diversity. *The American Economic Review*, 67(3):297–308.

Eizenberg, A. (2014). Upstream innovation and product variety in the us home pc market. *The Review of Economic Studies*, 81(3):1003–1045.

Fan, Y. and Yang, C. (2014). Competition, product proliferation and welfare: A study of the us smartphone market. *Available at SSRN 2506423*.

Furman, J. L. and Stern, S. (2011). Climbing atop the shoulders of giants: The impact of institutions on cumulative research. *The American Economic Review*, 101(5):1933–1963.

Greco, A. N. (2013). *The book publishing industry*. Routledge.

Handke, C. (2011). Economic effects of copyright. *The emprical evidence so far. Report prepared for the Committee on the Impact of Copyright Policy on Innovation in the Digital Era (USA)*.

Heald, P. J. (2007). Property rights and the efficient exploitation of copyrighted works: an empirical analysis of public domain and copyrighted fiction best sellers. *UGA Legal Studies Research Paper*, (07-003).

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2):615–635.

Li, X., MacGarvie, M., and Moser, P. (2015). Dead poet's property-how does copyright influence price? Technical report, National Bureau of Economic Research.

Mankiw, N. G. and Whinston, M. D. (1986). Free entry and social inefficiency. *The RAND Journal of Economics*, pages 48–58.

Morton, F. M. S. (2000). Barriers to entry, brand advertising, and generic entry in the us pharmaceutical industry. *International Journal of Industrial Organization*, 18(7):1085–1104.

Nagaraj, A. (2013). Does copyright affect creative reuse? evidence from the digitization of baseball digest. *Unpublished Working Paper*.

Peukert, C. and Reimers, I. (2017). Digital disintermediation and the market for ideas. *Unpublished Working Paper*.

Regan, T. L. (2008). Generic entry, price competition, and market segmentation in the prescription drug market. *International Journal of Industrial Organization*, 26(4):930–948.

Reiffen, D. and Ward, M. R. (2005). Generic drug industry dynamics. *Review of Economics and Statistics*, 87(1):37–49.

Reimers, I. and Waldfogel, J. (2017). Throwing the books at them: Amazon's puzzling long run pricing strategy. *Southern Economic Journal*.

Thomas, C. (2011). Too many products: Decentralized decision making in multinational firms. *American Economic Journal: Microeconomics*, 3(1):280–306.

Varian, H. R. (2006). Copyright term extension and orphan works. *Industrial and Corporate change*, 15(6):965–980.

Williams, H. L. (2013). Intellectual property rights and innovation: Evidence from the human genome. *Journal of Political Economy*, 121(1):1–27.

Table 1: Summary statistics - Amazon editions before and after 1923

	Before 1923		After 1923		
Variable	Mean	Std. Dev.	Mean	Std. Dev.	t-stat
Avail. as new	0.741	0.439	0.430	0.496	9.284
Avail. as used	0.826	0.379	0.947	0.224	-5.638
Editions per title (Bowker)	43.732	30.061	7.358	10.062	22.039
Edition age (years)	22.395	33.721	43.423	30.620	-9.010
Major publisher	0.239	0.427	0.515	0.500	-8.267
Hardcover	0.322	0.468	0.502	0.500	-5.212
Paperback	0.627	0.484	0.469	0.500	4.498
E-Book	0.0513	0.221	0.0286	0.167	1.657
Price	14.372	16.321	18.299	41.078	-1.691
Quantity per edition	1.536	4.377	6.727	28.056	-3.434

Summary statistics for 722 editions on Amazon, April 2011.

Table 2: Regression discontinuity - in-print versions per title and format

	(1) Title	(2) Hardcover	(3) Paperback	(4) E-Book
Copyright	-26.48***	-5.441***	-15.17***	-2.874*
	(5.962)	(1.237)	(2.959)	(1.484)
PLR demand	0.00823***	0.00179***	0.00280**	0.00224
	(0.00173)	(0.000204)	(0.00108)	(0.00151)
Pulitzer prize	1.705	0.898	0.0244	0.937
	(2.279)	(0.626)	(1.092)	(0.669)
Canon (title)	66.02***	9.558***	27.11***	5.161
	(10.40)	(3.005)	(5.904)	(7.385)
Canon (author)	10.46*	1.767	5.065*	1.189
	(5.410)	(1.514)	(2.599)	(0.968)
year	-0.561	-0.447*	-0.478	0.260
	(1.156)	(0.236)	(0.572)	(0.305)
$year^2$	-0.142***	-0.0154	-0.0610**	-0.0269**
	(0.0472)	(0.0119)	(0.0229)	(0.0121)
$year^3$	0.00497	0.00461	0.00453	-0.00350
	(0.0160)	(0.00351)	(0.00771)	(0.00419)
Constant	33.055***	7.374***	19.02***	2.276*
	(2.820)	(0.595)	(1.419)	(1.126)
Observations	148	148	148	148
adj. R-squared	0.764	0.700	0.778	0.402

Standard errors are clustered by original year of publication. *** p<0.01, ** p<0.05, * p<0.1. "PLR demand" denotes the title's checkouts at British libraries, "Pulitzer prize" is 1 if the author has won a Pulitzer prize, and "Canon" indicates whether the title (author) is included in Harold Bloom's *Western Canon*.

Table 3: Regression discontinuity - edition prices on Amazon

	All	Hardcover	Paperback	All	Hardcover	Paperback
	(1)	(2)	(3)	(4)	(5)	(6)
Copyright	0.242***	0.133	0.279**	0.121*	0.0296	0.157*
	(0.0757)	(0.147)	(0.118)	(0.0664)	(0.0997)	(0.0862)
E-book	-2.354***			-2.413***		
	(0.142)			(0.158)		
Paperback	-0.656***			-0.530***		
	(0.0407)			(0.0333)		
Avail. as new	-0.131*	-0.216	-0.114*	-0.122*	-0.220	-0.108*
	(0.0623)	(0.179)	(0.0544)	(0.0597)	(0.167)	(0.0537)
Edition age	-0.00379	-0.00407	0.00418	-0.00352	-0.00454	0.00400
	(0.00262)	(0.00276)	(0.00711)	(0.00247)	(0.00263)	(0.00690)
Major publisher	-0.194***	-0.368	-0.218**	-0.196**	-0.403	-0.227**
	(0.0644)	(0.244)	(0.0758)	(0.0683)	(0.257)	(0.0831)
Editions				-0.0124***	-0.0223	-0.0107**
				(0.00269)	(0.0212)	(0.00383)
Constant	3.309***	3.421***	3.431***	3.632***	2.600***	2.813***
	(0.104)	(0.101)	(0.159)	(0.271)	(0.106)	(0.115)
Quality:						
PLR demand	✓	✓	✓	✓	✓	✓
Pulitzer prize	✓	✓	✓	✓	✓	✓
Canon (title)	✓	✓	✓	✓	✓	✓
Canon (author)	✓	✓	✓	✓	✓	✓
Observations	2659	675	1966	2659	675	1966
adj. R-squared	0.457	0.121	0.264	0.474	0.155	0.279

Regressions with natural logs of edition prices on Amazon as the dependent variables. Standard errors are clustered by original year of publication. *** p<0.01, ** p<0.05, * p<0.1. "PLR demand" denotes the title's checkouts at British libraries, "Canon" indicates whether the title (author) is included in Harold Bloom's *Western Canon*, "edition age" is the time (in years) since the edition was published, "Amazon picture" denotes whether the edition's Amazon page shows a picture of the cover, and "edition" is the number of distinct in-print editions that the title-format is available in. I further include a cubic function of the title's original year of publication and month fixed effects.

Table 4: Demand estimation

	(1) Logit: OLS	(2) Logit: IV	(3) Nested logit: Title	(4) Nested logit: Title - Format
Price	-0.0978***	-0.382***	-0.138***	-0.311***
	(0.0121)	(0.0473)	(0.0523)	(0.0426)
Available as new	2.384***	4.183***	1.715***	3.465***
	(0.289)	(0.456)	(0.506)	(0.467)
Major publisher	0.109	-0.528	-0.134	-0.407
	(0.298)	(0.405)	(0.168)	(0.327)
Edition age	-0.0258***	-0.0640***	-0.0237***	-0.0484***
	(0.00551)	(0.0108)	(0.00889)	(0.00951)
Picture shown	1.448***	0.947**	0.328	0.750*
	(0.318)	(0.463)	(0.212)	(0.384)
Hardcover	-8.022***	-1.259	-2.088***	-3.021***
	(0.416)	(1.124)	(0.508)	(1.007)
Paperback	-8.680***	-4.011***	-3.072***	-4.772***
	(0.343)	(0.746)	(0.369)	(0.651)
E-book	-7.987***	-6.317***	-3.565***	-6.981***
	(0.898)	(0.905)	(0.671)	(1.000)
σ(title)			0.629***	0.265***
			(0.126)	(0.106)
σ(title-format)				0.00637
				(0.0472)
Title fixed effects	✓	✓	✓	✓
Month fixed effects	✓	✓	✓	✓
Sellers of edition		IV	IV	IV
Editions of title			IV	IV
Editions of format				IV
Mean Elasticity	-1.129	-4.415	-3.664	-4.880
Observations	8756	8756	8756	8756
Adj. R-squared	0.543	0.385	0.756	0.542
Number of titles	242	242	242	242

Robust standard errors clustered at the title level in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Columns 1 and 2 show regular logit, column 3 shows a nested logit estimation with nests at the title level, and column 4 shows nested logit results with nests at the title-format and the title level.

Table 5: Consumer surplus across the industry: decreasing copyright by 1 year

Assumptions		Effect
N(in print)	N(made available)	$\triangle$ Consumer Surplus
100	100	1,499,112
150	150	2,248,668
175	200	2,720,631
200	300	3,386,965
250	400	4,330,892

Estimates reported in 2012 $ per year.

Figure 1: Sales Ranks, Arrowsmith (Mass Market Paperback)

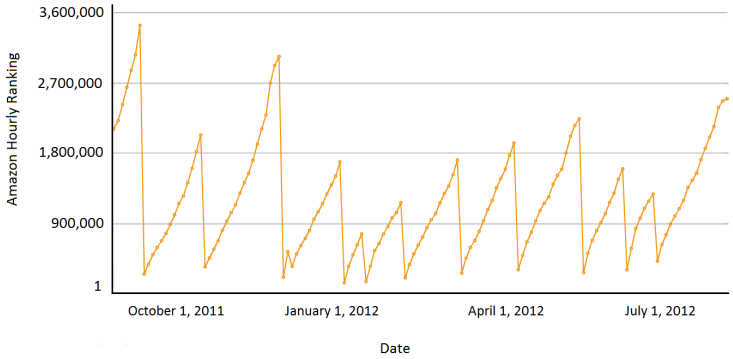

Figure 2: Number of current in-print ISBNs by year of publication

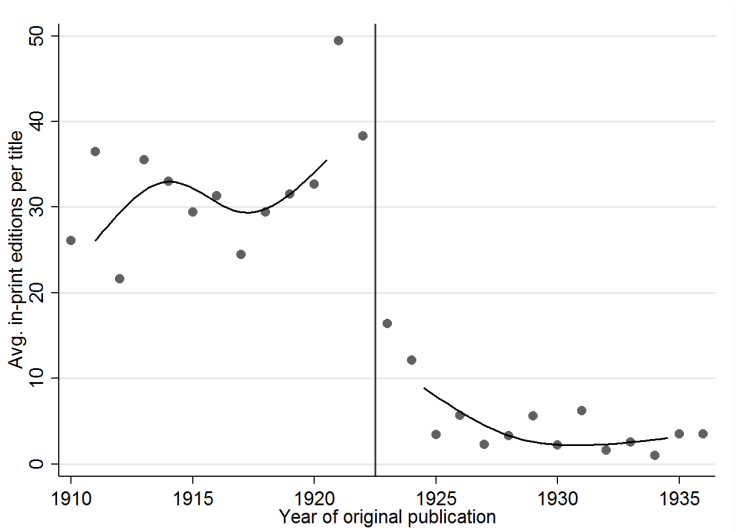

Figure 3: Average title PLR library check-outs in 2011, by year of publication

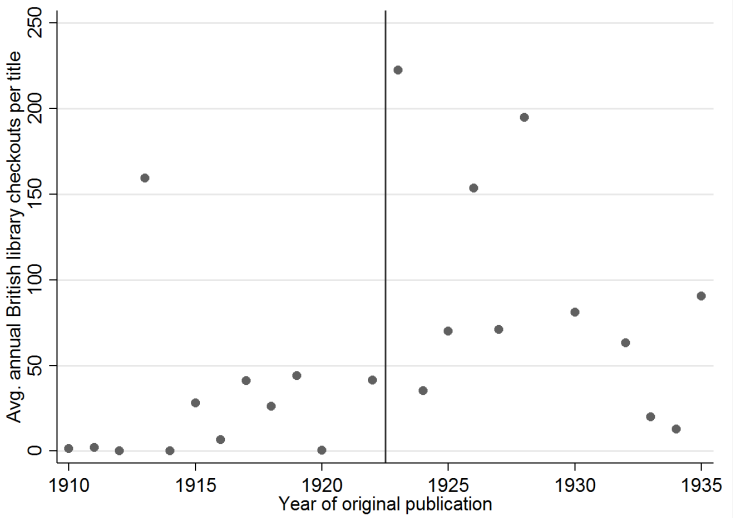

Figure 4: Title and edition characteristics first 74 years, by year of publication

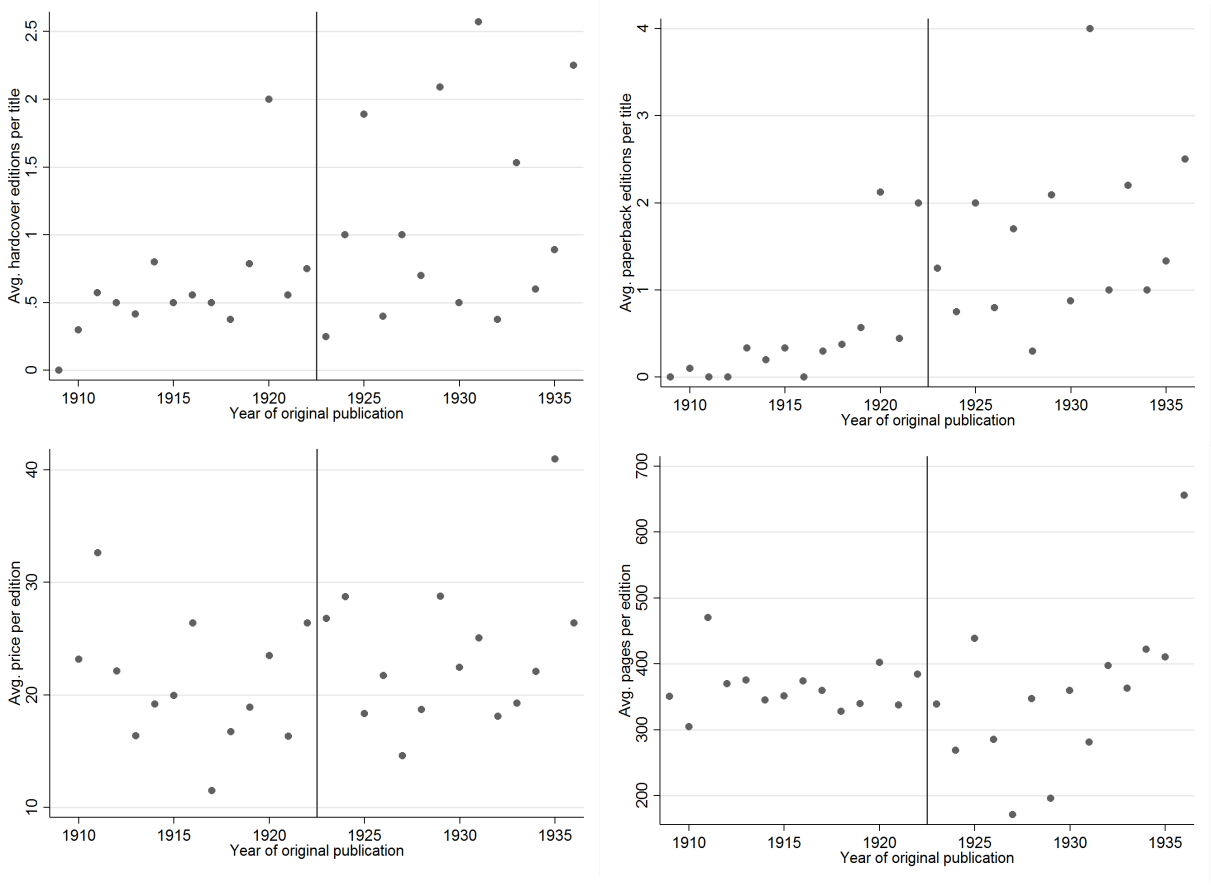

Figure 5: Editions per title as a function of its creative quality