Copyright and Generic Entry in Book Publishing[†]

By IMKE REIMERS*

Taking works off copyright promotes their availability, 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. Evaluating the welfare impact of the 1998 Copyright Term Extension Act, I find that a copyright significantly limits the availability of works, leading to a decrease in consumer surplus, which is significantly larger than any increases in profits to copyright holders. Without changing incentives to create new content, the copyright extension was economically inefficient. (*JEL* L11, L17, L82, O34)

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? Some IP rights extend for such a long period of time—far beyond the life of the original creators—that an extension of rights does not 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 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, a controversial legislation which set in place an abrupt change in copyright protection at the year 1923. Protected works, which were 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 in the book publishing industry.[1] Beyond showing evidence on these dimensions, I quantify the welfare effects of the copyright extension, including whether a work is available at all, effects on variety and prices when the work 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 them. These "orphan" works are unavailable to consumers in the market for new books, obviously creating a loss of surplus.[2] However, institutions such as Google Books and Project Gutenberg have begun to make digital copies of public domain titles available for free online. 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 in late 2005.[3] Eliminating copyright protection for 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 in the book publishing market? In answering the first question, 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 (2008) documents a large effect of the same extension on the availability of books at Amazon, and Li, MacGarvie, and Moser (2015) finds that an increase in copyright for works under the UK 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 1998 Sonny Bono Copyright Extension Act, in addition to a regression discontinuity analysis of availability and prices.[4] I collect information on all in-print and out-of-print editions of the annual top ten bestselling fiction novels from 1910 to 1936. Although my data are based on a rather selected sample, including about 4 percent of all in-print books from that time and only about 0.1 percent of all books that obtained a copyright, these books provide large variation in their demand *today*. They include books that are still successful today as well as several books that have lost their appeal over time and some which have become orphan works

---

[1] See Heald (2008) for an extensive descriptive analysis of availability and prices.

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

[3] The case was eventually 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 and Bosman (2013).

[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, Goldberg, and Gia (2006) estimates a large welfare loss in an antibiotics segment in India due to patent enforcement.

and moved out of print entirely.[5] For this list of 249 titles, I collect monthly edition prices as well as hourly ranking data on Amazon over a 12-month period in 2011 and 2012, which I use to infer the quantity sold.

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, consistent with findings in previous literature. The differences in availability and variety may be driven by cost differences: while the physical costs of production may be similar across copyright regimes, publishers of protected titles pay license fees to the copyright holder to obtain the right to publish. These differences are reflected in the prices of available editions, with editions of protected works being up to 35 percent more expensive than their public domain counterparts. Much of this difference appears to be driven by competition across editions.

The increase in availability and decrease in prices do not necessarily imply an increase in total surplus. 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. Such excess entry is most likely when fixed costs are particularly high and gains from variety are particularly low. It 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, Eizenberg, and Waldfogel 2015). Whether (or how much) the gains from variety—including the free editions through Google Books and Project Gutenberg—outweigh the fixed costs of additional editions in the book publishing industry is an empirical question.

I test for the welfare implications of the copyright extension by estimating a nested logit discrete choice model of demand for book editions. The model allows me 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 *Pride and Prejudice* 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. However, if a title's quality lies between these two types, its copyright status should significantly affect its availability and consumption.

I compare consumer surplus and joint profits (for the copyright holder and publishers) across titles and copyright regimes conditional on this creative quality, keeping in mind that my estimates are based on the most successful works at the time of publication. 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

---

[5] This list of titles is similar to Heald (2008), although Heald includes books listed in the all-time bestsellers list compiled by *Publishers Weekly* in addition to the annual bestsellers. He thus captures more durable titles than I do, without adding more of those titles which may not have stood the test of time.

surplus further. For all other titles, free entry and potential changes in the cost structure increase 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 and publishers' profits. The per-title gains range from about \$1,000 per year to \$50,000 or more annually for the most popular titles, with average gains of just under \$10,000.

I scale these gains to represent the book industry based on an industry-wide distribution of downloads on Project Gutenberg to account for the possibility that my set of titles is disproportionately popular today. Assuming no change in incentives to create new ideas, this implies an annual net decrease in consumer surplus from the book industry of \$2.5 million for each extended year of copyright protection.[6] For a twenty-year extension, accordingly, the annual impact would be twentyfold.

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 (2017) 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), and the total, dynamic impacts of the extension may in fact surpass the static welfare losses estimated in this paper.

I. 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 years 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 to follow the guidelines of the Berne Convention: new works by individual authors were granted copyright protection for 50 years beyond the author's death. For works which had already been published at

---

[6] My scaling method relies on the assumption that Gutenberg downloads are representative of creative qualities, and that their distribution is stable across time. To the extent that this is the case, this approach does not introduce bias.

that time, the maximum term of protection was increased from 56 years to 75 years after publication.

Public debate sparked again in the 1990s over further extensions of the copyright term, eventually leading to the 1998 Sonny Bono Copyright Term Extension Act.[7] Proponents of the extension argued for a uniform copyright term across countries to better protect American works in foreign countries, in addition to incentivizing their owners to preserve and digitize works because they had exclusive rights to them.[8] Opponents argued that an extension from 75 years to 95 years arguably does not impact the incentive to create new works at the margin—the present discounted value will be small even for the most successful works. However, it does impact the availability of so-called "orphan works"—works which cannot be published because hunting down the copyright holder is prohibitively costly—with large impacts on consumer surplus. In addition, extending a copyright would restrict the production of new creative works, which would make use of existing materials.[9] Despite strong public opposition, both houses of the US Congress passed the copyright extension with a voice vote.[10] The Sonny Bono Copyright Extension Term was signed by then-president Bill Clinton on October 27, 1998.

This 1998 copyright extension lengthened copyright protection for cultural goods that 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.[11] Barring further copyright extensions, works will again start moving into the public domain by 2019. More public debate will likely ensue, increasing the need for understanding the implications of such changes.

Importantly, the copyright extension provides exogenous variation in the cost of publishing a book today. If a title is in the public domain, anybody who wants to publish it can do so without having to obtain permission. If a work is protected by copyright, however, the publisher has to obtain a license from the copyright holder in order to publish the work.[12] This license includes a royalty for each individual copy sold on top of an advance to the author. The advance most often ranges between some 10,000 and hundreds of thousands of dollars, and it can amount to

---

[7] 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.

[8] See the Senate Report of the 104th Congress at https://www.congress.gov/congressional-report/104th-congress/senate-report/315.

[9] See, for example, the amicus brief submitted by 17 economists at https://cyber.harvard.edu/openlaw/eldredvashcroft/supct/amici/economists.pdf.

[10] See https://www.congress.gov/bill/105th-congress/house-bill/02589/all-actions and https://www.congress.gov/bill/105th-congress/senate-bill/505/all-actions for overviews of all actions taken from the bills introduction until it became public law.

[11] See http://copyright.cornell.edu/resources/publicdomain.cfm for more detail.

[12] 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.

over a million dollars for the most promising works (Peukert and Reimers 2018). 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 multiproduct monopolist for that title. But because authors (or their heirs) have the option to regain these rights after 28 or 56 years (depending on the author's time of death, see Heald 2018), many of the titles in my dataset have been published by—and are still in print today through—multiple publishers.

II. 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 that can a priori be regarded as similar: the annual ten bestselling fiction titles of the years 1910 to 1936.[13] Although this is neither an exhaustive list of all high-quality titles from the time period nor of the most popular works given today's demand, it consists of titles whose authors had strong incentives to register and later renew a copyright.[14] More important, the list provides a wide variety of demand today. Several titles still face positive demand, while almost 15 percent have moved out of print as orphan works. In comparison, of 10,308 US books published and copyrighted in 1930, only 174 were still in print in 2001.[15] Still, these orphan titles might be available in the used-books market. The underlying dataset consists of 249 titles. Of these, 37 were out of print in 2013, but 34 of those were available as used books at Amazon.

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 and text notes. Each title may be available in several ISBNs or *editions*.

I create two datasets based on the underlying set of titles. First, I use the Bowker Books-in-Print directory to obtain characteristics of all available editions for each title. Second, I collect sales data and retailer-specific edition character-istics, 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 edition-level weekly sales data from the Nielsen Bookscan database and with aggregate information on the publishing industry from the Book Industry Study Group BookStats report of 2011 and *Publishers Weekly* periodicals.

---

[13] These can be found on several websites, but Korda (2001) was used as a reliable reference.

[14] Between 1923 and 1936, 12,686 books were copyrighted on average each year; 3,482 of which were renewed after the first 28-year term. See the Internet Archive's Copyright Records at https://archive.org/details/copyrightrecords and Stanford's Copyright Renewal Database at https://exhibits.stanford.edu/copyrightrenewals. Similarly, Landes, and Posner (2003) shows that fewer than 11 percent of the copyrights registered between 1883 and 1964 were renewed after 28 years.

[15] See Landes and Posner (2009). A book is out of print if it is no longer available through the publisher as a new book.

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 were available and in print in 2013, with large variations across titles and copyright regimes. Several (protected) works have moved out of print entirely, while one public domain title (Edith Wharton's *The Age of Innocence*) still has 146 different in-print editions.[16]

I measure the sales of each edition by collecting hourly Amazon sales rankings of a subset of all ISBNs and title-format combinations over the time period of my study (2011 to 2012), using the website novelrank.com. Among researchers facing rankings, there is an active tradition of translating sales ranks into quantities by assuming that sales-rank relationships tend to obey power laws (see Chevalier and Goolsbee 2003; Brynjolfsson, Hu, and Smith 2003; Reimers and Waldfogel 2017). 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 (typically 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 quite unlikely. For each public domain title in this dataset, I collect monthly Project Gutenberg download counts over the same period in addition to the Amazon editions.

Table 1 summarizes the editions' characteristics 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.[17] The table shows that public domain works and protected titles differ from each other on several dimensions. First, public domain editions are significantly more likely to be available as new. Second, the public domain editions on Amazon tend to be published more recently, are less likely published by a major publisher, and are more likely to be in paperback format than editions of comparable protected titles. The editions' characteristics are consistent with those observed in the Bowker dataset—the universe of in-print editions—suggesting that the editions in my dataset are representative.[18]

Finally, a copyright affects prices by changing the cost structure, which can in turn impact the quantity demanded. Public domain titles are indeed cheaper than protected titles. Yet, editions of protected 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. An average Project Gutenberg edition in my dataset was downloaded 118.3 times in April 2011, whereas editions of protected titles were sold 6.7 times on average.

---

[16] Note that Project Gutenberg does not register ISBNs for their editions, so that these editions do not enter the count. In addition, I only count editions that were still in print in 2013 to avoid double-counting of editions, which replace their predecessors.

[17] 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 use the Bowker Books-in-Print directory to infer entry of editions.

[18] I test for the representativeness of my editions more formally in Section IVD.

Table 1—Summary Statistics: Amazon Editions before and after 1923

Variable	Before 1923		After 1923		
	Mean	SD	Mean	SD	t-stat
Available as new	0.741	0.439	0.430	0.496	9.284
Available 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

Note: Summary statistics for 722 editions on Amazon, April 2011.

III. The Effect on Availability and Price

Table 1 suggests that editions of protected titles are different from their public domain counterparts. To examine whether there is a causal effect of a copyright on availability and prices, I take advantage of the fact that the copyright extension in 1998 did not affect the incentives to create a work in the 1920s. Because there was no change in the copyright term between 1909 and 1962, 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 published, plus 28 years if the copyright holder renewed the term.

The 1998 copyright extension thus provides a sharp, exogenous discontinuity in today's copyright-protection status at the year 1923, with no expected discontinuity in the inherent quality and appeal of the works. All titles are fairly "old" (published at least 80 years prior to this study), and all are fiction novels which were fairly successful in the United States upon their original publication. Yet, titles that were originally published before 1923 are in the public domain, whereas more recent titles have been "treated" and are still protected. I use a title's original publication year as the forcing variable in a sharp 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

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

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. 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.[19]

A. *Variety of Editions*

I first examine the impact of a copyright on the number of in-print editions per title and title-format. I control for a work's appeal with three variables. First, I include an indicator variable that equals one if the author has won a Pulitzer Prize. Second, I add 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.[20] Third, I collect British library checkouts between 2010 and 2012 (PLR—Public Lending Rights).[21] This variable captures a title's appeal under two conditions. First, copyright in Britain does not have a discontinuity during the observed time period. This condition is satisfied because British copyright law sets the copyright term at 70 years after the author's death, rather than based on the work's publication date. Second, the British and American markets do not affect each other. While consumers in Britain could find works from before 1923 online on American websites, I show in Section IIIB that this concern is minor.

Table 2 shows the coefficients from this specification. 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 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 1 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 is consistent with the evidence in 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. The remainder of the difference is split between hardcovers (5.4 editions) and e-books (2.9 editions).[22]

The positive coefficients on the number of British library checkouts and on being a Pulitzer award winning author are expected. 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 dataset 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.

---

[19] 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.

[20] 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.

[21] Authors in Britain receive royalties on library checkouts, so that this information is available there.

[22] The effect of the extension on variety is robust. Among the most popular works (the top 25 percent 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.

TABLE 2—REGRESSION DISCONTINUITY: IN-PRINT VERSIONS PER TITLE AND FORMAT

	Title (1)	Hardcover (2)	Paperback (3)	e-book (4)
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
Adjusted R^2	0.764	0.700	0.778	0.402

Notes: Standard errors, clustered by original year of publication, are in parentheses. "PLR demand" denotes the title's checkouts at British libraries, "Pulitzer Prize" is one if the author has won a Pulitzer Prize, and "canon" indicates whether the title (author) is included in Harold Bloom's *Western Canon*.
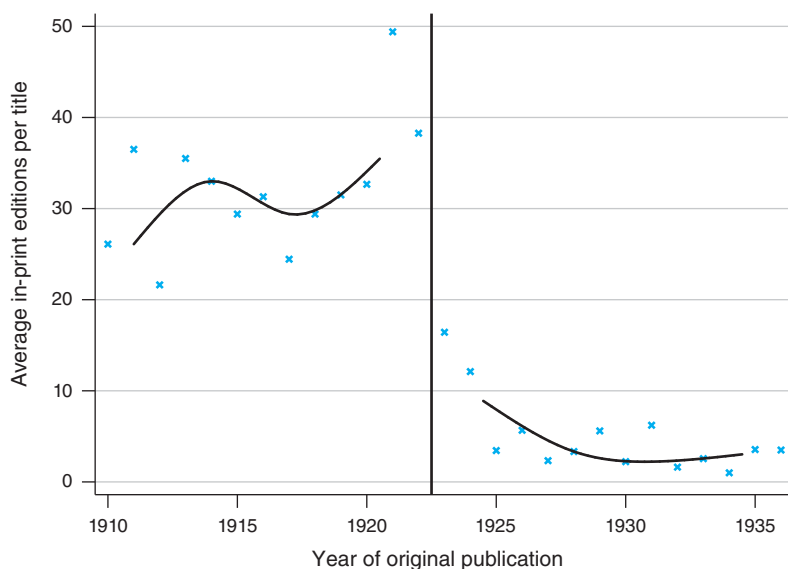

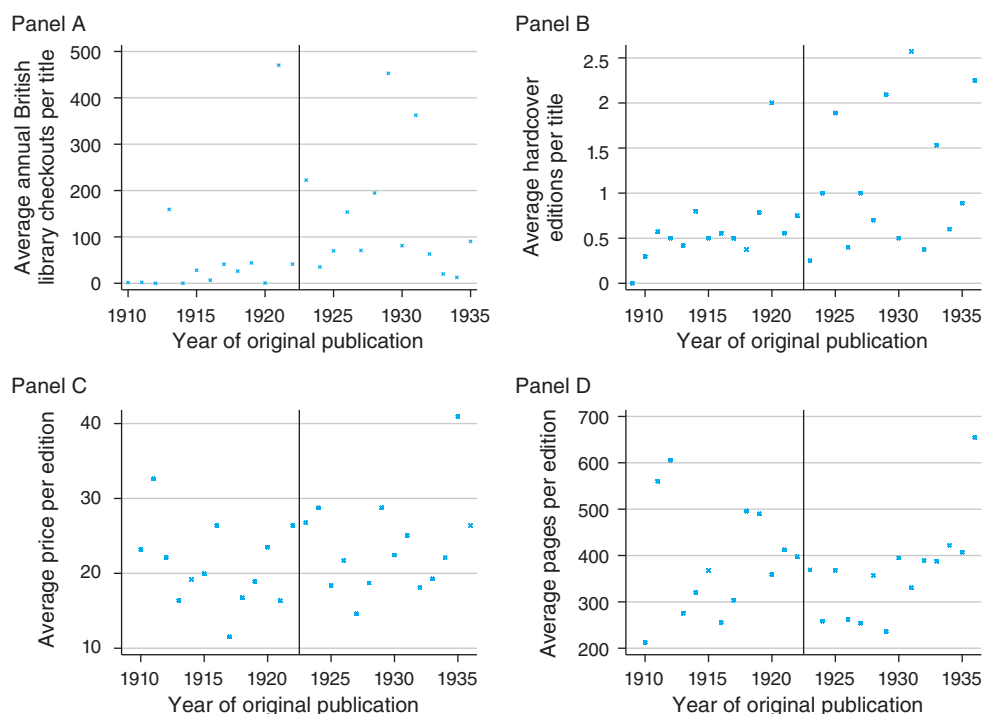FIGURE 1. NUMBER OF CURRENT IN-PRINT ISBNS BY YEAR OF PUBLICATION

Panel A

Panel B

Panel C

Panel D

FIGURE 2. TITLE AND EDITION CHARACTERISTICS BY YEAR OF PUBLICATION

Notes: Panel A shows title-specific demand through British libraries. The other panels show edition characteristics in the first 74 years of the title's life.

B. *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. I test this assumption on the title level using PLR checkouts, and on the edition level by examining editions published within 74 years of the work's original publication—before any of the titles would have moved into the public domain.

Figure 2, panel A shows that while titles that 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 remaining more popular. Graphs for the other explanatory variables in the regression look similar. The remaining panels of Figure 2 examine characteristics of editions which were published when a title was still protected by copyright. Panel B shows that while more recent titles tend to publish more editions, there is no discontinuity at the year 1923. 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 (bottom left) or in the number of pages per published edition (bottom right). This provides further support for the identifying assumption that the bestsellers are similar ex ante.

C. *Prices of Published Editions*

The large discontinuity in variety may be due to two factors: the difficulty of finding the copyright holder and the costs of obtaining the right to publish. Both IP-specific costs 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 more competition. To determine whether the differences across copyright regimes lead to differences in prices, I repeat the analysis from equation (1), using each edition's monthly prices (in logs) on Amazon as the dependent variable.[23] 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.[24] Finally, I include the number of in-print editions of the title and format to examine reasons for possible price differences.

Table 3 displays the results of these regressions. Columns 1 and 4 include editions of all 3 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 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 an edition's price by 12.8 ($= e^{0.121} - 1$) percent, rather than by 27.4 percent, with much of the difference again driven by paperback editions (see column 6). Competition explains part of the price differences, but cost differences across regimes likely also play a role.

IV. 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 that remain under copyright protection are much less likely to be available than comparable books in the public domain. Even among in-print books, the variety

---

[23] The edition's price on Amazon is a more relevant measure than its suggested retail price because edition prices vary over time.

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

Table 3—Regression Discontinuity: Edition Prices on Amazon

	All (1)	Hardcover (2)	Paperback (3)	All (4)	Hardcover (5)	Paperback (6)
Copyright	0.242 (0.0757)	0.133 (0.147)	0.279 (0.118)	0.121 (0.0664)	0.0296 (0.0997)	0.157 (0.0862)
e-book	−2.354 (0.142)			−2.413 (0.158)		
Paperback	−0.656 (0.0407)			−0.530 (0.0333)		
Available as new	−0.131 (0.0623)	−0.216 (0.179)	−0.114 (0.0544)	−0.122 (0.0597)	−0.220 (0.167)	−0.108 (0.0537)
Edition age	−0.00379 (0.00262)	−0.00407 (0.00276)	0.00418 (0.00711)	−0.00352 (0.00247)	−0.00454 (0.00263)	0.00400 (0.00690)
Major publisher	−0.194 (0.0644)	−0.368 (0.244)	−0.218 (0.0758)	−0.196 (0.0683)	−0.403 (0.257)	−0.227 (0.0831)
Editions				−0.0124 (0.00269)	−0.0223 (0.0212)	−0.0107 (0.00383)
Quality:						
PLR demand	✓	✓	✓	✓	✓	✓
Pulitzer Prize	✓	✓	✓	✓	✓	✓
Canon (title)	✓	✓	✓	✓	✓	✓
Canon (author)	✓	✓	✓	✓	✓	✓
Observations	2,659	675	1,966	2,659	675	1,966
Adjusted R^2	0.457	0.121	0.264	0.474	0.155	0.279

Notes: Regressions with natural logs of edition prices on Amazon are shown as the dependent variables. Standard errors, clustered by original year of publication, are in parentheses. "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, a constant, and month fixed effects.

of protected works is limited, and available editions are more expensive. But while these differences are suggestive of significant welfare effects, it is possible that the additional editions are of low enough production quality that they do not increase consumer surplus significantly.

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. 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) to infer product qualities while allowing for varying substitutability across products. I then use these estimates to determine differences in consumer surplus across copyright regimes.

A. *Demand Model*

Define J_{wt} as the set of editions that are 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 w is published in one of four formats, denoted by $k \in \{H, P, E, G\}$, where

H = hardcover, P = paperback, E = e-book, and G denotes the free e-book edition on Project Gutenberg, which is only available if the title is in the public domain. Each consumer 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 the j from the $J + 1$ options that maximizes her indirect utility function given by

$$(2) \qquad u_{ij} = X_j'\beta - \alpha p_j + \phi_w + \xi_j + \zeta_{iw} + (1 - \sigma)\epsilon_{ij}$$

$$= \delta_j + \zeta_{iw} + (1 - \sigma)\epsilon_{ij},$$

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 an indicator that is one if the book's cover is shown on its Amazon page.[25] 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 is $s_j = \dfrac{\exp\{\delta_j/(1 - \sigma)\}}{D_w^\sigma\left(1 + \sum_{w' \in W} D_{w'}^{1-\sigma}\right)}$, where $D_w = \sum_{j \in J_w} \exp\left\{\dfrac{\delta_j}{1 - \sigma}\right\}$. Setting the mean utility from consuming the outside option to zero yields the following regression equation:

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

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

B. *Estimation*

Estimation of the demand model is complicated by three issues, including low demand, endogenous independent variables, and a format (Gutenberg), which is inherently different from the others. I address these issues in detail here.

First, many editions experience zero sales on Amazon in a given month, leading to an undefined natural log of the market share (see, for example, Quan and Williams 2018).[26] Because these zeros contain information, I include them in the estimation by adding a small constant (0.0001 units) to each edition's monthly

---

[25] I treat this variable as a proxy for the seller's trustworthiness and professionalism and the consumer's likely perception of the edition's quality. This is particularly important when the book is only available used.
[26] This is the case for almost two-thirds of the observations.

unit sales, noting that the main results are robust to adding any constant between 0.00001 and 1.[27]

Second, 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 its sales within a title. Traditionally, one would instrument for the edition's inside share with the number of products within its nest (title and/or format) and for its price with a cost shifter. I calculate the number of editions per title and format directly from my dataset. It varies across editions and over time because some sellers exit the market. Finding a cost shifter that varies across editions and over time is more difficult. Instead of observing costs directly, I utilize the number of sellers offering the specific edition to proxy for the ease of distributing it.[28] Both instruments can provide insight into the costs of distributing a title as well as the edition's share within the title and format: the correlation between the level of competition within the title-format and the inside share is -0.2344 (its correlation with price is -0.015), and the correlation between the edition's number of sellers and the price is -0.265 (its correlation with the inside share is 0.31).[29]

In addition, the instruments satisfy the exclusion restriction. One might be concerned that the number of sellers offering an edition may be correlated with popularity or consumer price sensitivity, but these concerns are alleviated for two reasons. First, if the number of sellers was negatively correlated with consumer price sensitivity, we would expect prices and the number of sellers to be positively correlated. But they are not—their correlation is -0.265. Second, demand shocks are likely title specific, rather than edition specific, and my estimation includes title fixed effects. The number of sellers, however, is on the edition level. Thus, the instrument is only invalid if the number of sellers for individual editions of a title reacts to demand shocks for those editions only. This is unlikely.

Third, the free Gutenberg editions are inherently different from the editions I observe through Amazon. Importantly, the edition characteristics used to determine δ_j take on different meanings. I treat Gutenberg works as new editions, not from a major publisher, and with a picture on its website, noting that the estimation results are robust to different choices of these variables. Finally, I need to adjust the number of sellers for each edition for all electronic books. To account for the fact that distribution of electronic copies is costless, I set the number of sellers for e-books and Gutenberg editions to 100 or the maximum number of sellers per edition in month t, whichever is larger.[30]

C. *Estimation Results*

Table 4 reports the results of several specifications of equation (3). The first 2 columns show the results from a logit model, where column 2 introduces the number

---

[27] Larger constants would lead to larger distortions of the log term, however. Demand estimation results and changes in consumer surplus from adding different constants are reported in Appendix Section A2.

[28] Amazon lists this information on each edition's respective page, as "*X* New from . . ." and "*Y* Used from . . . ."

[29] A table of the first-stage regressions can be found in Appendix Section A1. All coefficients are as expected.

[30] My results are almost unchanged when instead using different numbers of sellers. These results are available upon request.

TABLE 4—DEMAND ESTIMATION

	Logit: OLS (1)	Logit: IV (2)	Nested logit: Title (3)	Nested logit: Title-format (4)
Price	−0.0978 (0.0121)	−0.382 (0.0473)	−0.138 (0.0523)	−0.311 (0.0426)
Available as new	2.384 (0.289)	4.183 (0.456)	1.715 (0.506)	3.465 (0.467)
Major publisher	0.109 (0.298)	−0.528 (0.405)	−0.134 (0.168)	−0.407 (0.327)
Edition age	−0.0258 (0.00551)	−0.0640 (0.0108)	−0.0237 (0.00889)	−0.0484 (0.00951)
Picture shown	1.448 (0.318)	0.947 (0.463)	0.328 (0.212)	0.750 (0.384)
Hardcover	−8.022 (0.416)	−1.259 (1.124)	−2.088 (0.508)	−3.021 (1.007)
Paperback	−8.680 (0.343)	−4.011 (0.746)	−3.072 (0.369)	−4.772 (0.651)
e-book	−7.987 (0.898)	−6.317 (0.905)	−3.565 (0.671)	−6.981 (1.000)
σ(title)			0.629 (0.126)	0.00637 (0.0472)
σ(title-format)				0.265 (0.106)
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	8,756	8,756	8,756	8,756
Adjusted R^2	0.543	0.385	0.756	0.542
Number of titles	242	242	242	242

Notes: Robust standard errors are clustered at the title level in parentheses. 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.

of sellers as an instrument for the edition's price. Note that the price coefficient becomes significantly negative after introducing the instrument, suggesting that the instrument alleviates the usual attenuation bias. Column 3 represents the preferred specification introduced in Section IVA. 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 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 percent 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 editions within a title-format (σ(title-format)) remains statistically significant, the correlation across formats within a title (σ(title)) is less precisely estimated. Although the coefficients remain similar for all coefficients, I continue with the results from column 3 because many title-format combinations only include one edition, making identification of correlation of tastes within title-formats unreliable.

The title fixed effects provide two insights about each title's creative quality ϕ_w—its inherent appeal in the twenty-first century. First, Figure 3 shows the titles' average creative qualities per year of original publication. There is no significant difference in the size of the fixed effects across copyright regimes, suggesting that publishers create editions with similar production qualities ξ_j across the copyright regimes as well. The argument by Adilov and Waldman (2013) that a longer copyright term triggers more ex post investment in the work finds no evidence in my set of fiction novels.

Second, conditional on copyright status, higher quality titles tend to be published in larger varieties, particularly for public domain works. Figure 4 shows the relationship between a title's creative quality and the number of available editions for both copyright regimes.[31] 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.

D. *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 the increase is unclear. I use the estimated σ, α, and mean utilities for each edition from the demand estimation to quantify the differences in consumer surplus between the two copyright regimes.

I compare the consumer surplus from currently protected titles to that generated from public domain titles with similar creative qualities. More specifically, I rank the public domain and protected titles according to their creative qualities, and I compare the consumer surplus from titles of the same copyright-specific quality ranking. In my application, the difference in consumer surplus per quality-ranking (title) r is

$$(4) \quad \Delta E[CS] = \frac{M}{\alpha}\left[\ln\left(1 + \left[\sum_{j \in J_r^O} \exp\left(\frac{\delta_j}{1-\sigma}\right)\right]^{1-\sigma}\right) - \ln\left(1 + \left[\sum_{j \in J_r^I} \exp\left(\frac{\delta_j}{1-\sigma}\right)\right]^{1-\sigma}\right)\right],$$

where the superscripts describe the choice sets under examination: I represents IP protection, and O denotes the public domain. Note that I directly compare public domain and protected titles, implicitly assuming that the edition managers of the

---

[31] I normalize the works' creative qualities to be distributed between zero and ten for illustrative purposes.
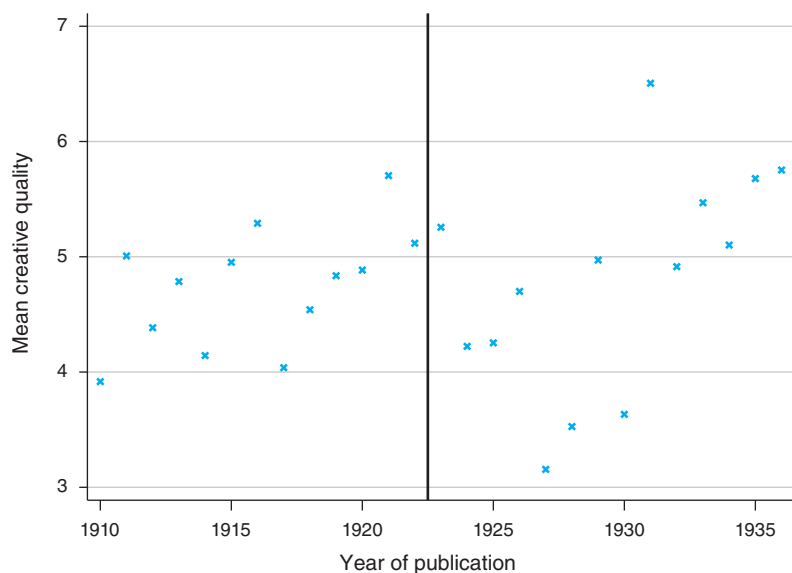
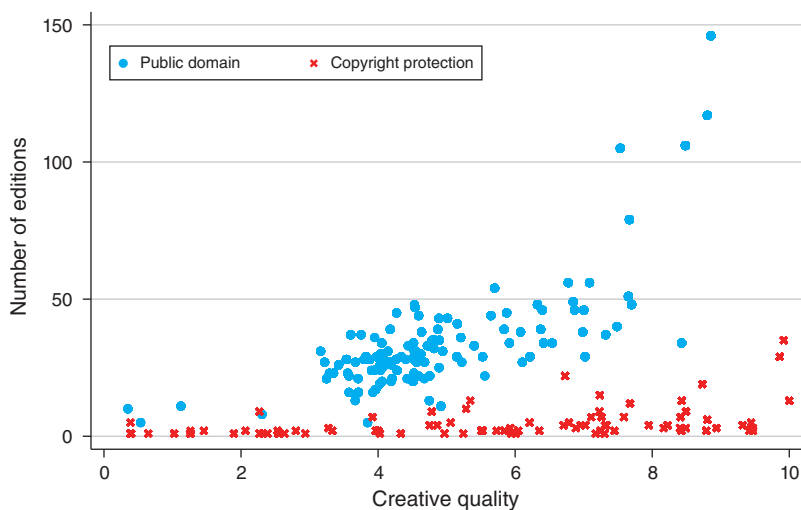FIGURE 3. AVERAGE ESTIMATED CREATIVE QUALITY BY YEAR

FIGURE 4. EDITIONS PER TITLE AS A FUNCTION OF ITS CREATIVE QUALITY

public domain works made optimal entry decisions, and that edition managers of currently protected titles would act the same way.[32]

To accurately report the differences in consumer surplus, I need to address two features of the data. First, I only observe the quantities demanded through two distribution channels: online retail through Amazon and downloads through Project

---

[32] The nature of the data—two datasets with different prices that cannot easily be combined—prevents me from estimating an entry model directly.

Gutenberg. Any consumption through brick-and-mortar bookstores and libraries, or downloads of free editions through Google Books and Amazon, is not observed. Second, I only observe demand for a subset of all in-print editions, but consumer surplus is derived from *all* available editions. To address these issues, I collect sales information from the Nielsen Bookscan database.[33]

To address the first issue, I compare the unit sales I collected through Amazon with total sales as recorded on Nielsen for all observed editions with an ISBN.[34] On average, sales through Amazon account for about 37 percent of the sales reported by Nielsen, although Amazon's share is larger for more obscure books.[35] Based on Amazon's average share, I assume that my dataset captures a share of $\gamma = 0.4$ of the total sales and downloads of my list of editions. That is, I multiply the observed and estimated quantities by 2.5 to cover the entire market.

To obtain Project Gutenberg's market share, I collect Google Trends search volumes for Project Gutenberg and its biggest competitor (Google Books). Throughout the time period of my dataset, Project Gutenberg was searched about one-third as often as Google Books, suggesting a 25 percent market share. This is likely an underestimate, however, as the *Authors Guild v. Google* lawsuit received a lot of attention in 2011, likely leading to more searches for Google Books. To remain conservative, I set Project Gutenberg's market share to 0.4 as well, noting how the results change for smaller market shares when appropriate.

Then, I address the second issue by comparing total sales (as recorded by Nielsen) of the editions I observe with total sales of editions of the same title but not in my dataset. I find that those editions with an ISBN in my dataset sell about eight times as often as the editions I did not collect. Similarly, these editions sell seven times as often as those editions in my dataset without an ISBN. This evidence, along with the fact that the observable edition characteristics of my Amazon editions match those in the Bowker database, suggests that my set of editions is relatively representative. Accordingly, I expand the number of editions for each title-format to the number of in-print editions as in the Bowker database to calculate consumer surplus from each title.

Per-Title Consumer Surplus for Books in My Sample.—While most titles in my dataset generate little consumer surplus regardless of their copyright status, public domain titles generate more consumer surplus than their protected counterparts. On average, a public domain title in my dataset generates $12,278 in consumer surplus (or $16,734 if Project Gutenberg accounted for only 25 percent of all downloaded free editions), compared to just $2,550 for protected titles. The differences are largely due to differences in prices and the popular zero-price format of Project Gutenberg e-books.

The titles with the largest creative qualities generate the largest consumer surplus. For example, Edith Wharton's *The Age of Innocence*—the highest quality

---

[33] Note that I do not use the Nielsen dataset to estimate demand because Nielsen does not report real time prices.

[34] Not all editions list an ISBN on Amazon. For those without an ISBN, Amazon has its own identifier: an ASIN.

[35] For editions in the lowest four quintiles of Amazon's sales, I record 9.8 average unit sales through Amazon, compared to 15 units recorded by Nielsen.

public domain title—generates over $183,000 in consumer surplus per year with 146 different editions at an average price of $4.75, whereas Pearl Buck's *The Good Earth*—the protected title generating the largest consumer surplus—generates just over $59,000 in annual consumer surplus with 35 editions and an average price of $10.55. In a stronger contrast, Margaret Mitchell's *Gone With the Wind*—the highest *quality* protected title, which is available in 13 editions—generates only around $35,000 in annual consumer surplus. However, *The Age of Innocence* is an outlier: the second-most popular public domain title, Sinclair Lewis's *Babbitt* generates a surplus of $73,000.

The differences in per-title consumer surplus vary significantly by creative quality. Many low-quality titles are hardly sold at all and generate very low consumer surplus regardless of their copyright status. Figure 5 compares public domain titles with the protected titles of the same quality rank, plotting the differences in consumer surplus against their quality rankings.[36] While the difference in annual consumer surplus between the copyright regimes is around $50,000 for the most highly ranked titles, this difference does not exceed $5,000 for titles ranked beyond 75 in my sample. Of course, many books which did not enter the annual top ten bestseller lists likely generate even less surplus.

Industry Consumer Surplus.—An assessment of the copyright extension needs to take into account impacts on the entire industry, not just the titles in my sample. Here, I use the consumer surplus estimates from the titles in my dataset to infer how much consumer surplus would increase for *all* books still under copyright. The challenge for extending the calculated surplus changes to the industry level is that my sample is selected, and I do not observe the quality of books outside the top ten, though these are likely to be lower on average. So, to infer an industry-wide impact, I estimate the quality of all impacted books using Project Gutenberg download counts for a random sample of 30,000 public domain works and using two plausible assumptions.[37]

First, I assume the distribution of book qualities is stable over time, and that Gutenberg downloads reflect these qualities. That is, the distribution of downloads of public domain books is the same as what the distribution of downloads would be of books that are still protected because of the Sonny Bono copyright extension. Second, I assume that if two books have the same number of Project Gutenberg downloads, then the increase in consumer surplus from putting them into the public domain is the same as well, conditional on the protected work's in-print status.[38]

In what follows, I use the above assumptions to create a download-weighted, per-title measure of consumer surplus effects. I then make informed guesses about

---

[36] The top-ranked titles—with a consumer surplus difference of almost $150,000—are dropped to display differences for lower ranked titles.

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

[38] The impact of a move into the public domain likely depends on the work's in-print status. When a title is in print, the impact is on the intensive margin—it depends on the consumers' taste for variety. In contrast, when an orphan work moves back into print, the impact is on the extensive margin—making the book available at all—and may therefore be larger. Consequently, I distinguish between these two types of works even though the differences are not large in my dataset.
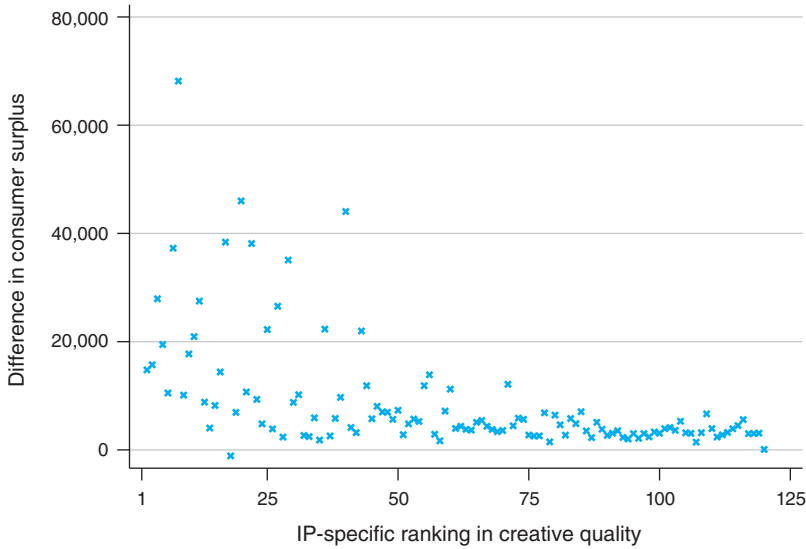
FIGURE 5. IP-RELATED CHANGES IN CONSUMER SURPLUS BY CREATIVE QUALITY
(HIGHEST TO LOWEST)

Note: The figure shows, for each quality rank r, the difference between the consumer surplus from the public domain title of that rank and the consumer surplus from the protected title of the same rank: $CS_r^O - CS_r^I$, where O denotes the public domain and I denotes copyright (IP) protection.

how many books from each year are still in print and how many would move from out-of-print status to in-print status if copyright restrictions were lifted to scale these impacts to the industry level.

First, I take 30,000 Gutenberg books to obtain a baseline distribution of the number of Gutenberg downloads of books in the public domain. By the first assumption, this determines the distribution of downloads (and creative qualities) for all works whose copyright was extended.

Second, my demand estimates yield a mapping from a book's creative quality to the change in surplus from removing its copyright protection. I use this mapping and the observed download counts for the public domain titles in my sample—together with the second assumption—to create a mapping between Gutenberg downloads and the change in consumer surplus, for both in-print and out-of-print protected books.

Third, I assign changes in consumer surplus to all 30,000 public domain books according to this mapping. For works not in the original sample with download counts between those of two titles in the sample, I assign the surplus change of the next lowest download count to remain conservative.[39] Then, taking the average consumer surplus impact across all titles provides a

---

[39] In addition to providing conservative estimates of surplus changes, this simple strategy does not impose any (unverifiable) structure on the title quality distribution. Another option would be to round up to the next largest observed download count. I report how results change under this assumption.

download-weighted, per-title consumer surplus impact, both for in-print and for out-of-print protected works.

Finally, I use educated guesses to determine how many in-print books would move into the public domain each year and how many out-of-print books would move back into print if their copyright was lifted. Some information is available on in-print titles: of the 10,308 titles originally published in 1930, only 174 were still in print in 2001.[40] Consequently, I set the number of in-print titles between 100 and 250 per year of publication. The number of out-of-print titles that would move back into print is more difficult to obtain. Project Gutenberg offered just over 47,000 public domain titles in November 2014. If the first documents were originally published in 1776 (the Declaration of Independence) and the most recent titles are originally from 1922, Gutenberg has around 300 titles per year of publication. But because Gutenberg also carries works from other years, and because not all works are full books, I assume that between 50 and 300 additional titles would move back into print each year. To calculate the industry-wide impact, I multiply the weighted average impacts by these numbers.

Table 5 shows this industry-wide impact given different guesses about the numbers of in-print protected titles (column 1) and out-of-print protected titles that would move back into print (column 2). The table suggests that a one-year decrease in the copyright term would increase consumer surplus by between $1.3 million and $4.0 million annually. If 175 titles are currently in print, and 150 become newly available if in the public domain (my "best guess"), consumer surplus from book publishing increases by about $2.53 million per year.[41]

V. Beyond Consumer Surplus: Profits

Proponents of strict IP laws argue that IP provides an incentive to create new works, and policymakers take these incentives into account. In the case of a copyright, and in particular in the case of the 1998 Copyright Term Extension Act, these incentives are less likely affected. It is possible, however, 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 by examining publisher and copyright holder profits more closely.

A. *Supply Model*

As intuition for the profit calculations, I introduce a motivating model of free entry in two stages. The idea is simple: in the public domain, free entry implies that firms introduce an edition if and only if the expected profit is positive. That is, the marginal edition's operating profits should be close to its fixed costs in expectation. Under copyright protection, the mechanism is more complicated, as the copyright

---

[40] See Landes and Posner (2009). More detailed information for other years is not readily available to my knowledge.

[41] If I instead round up to the next largest observed download count, the calculated annual consumer surplus increase would be $2.76 million per extended year.

Table 5—Consumer Surplus across the Industry: Decreasing
Copyright by One Year

Assumptions		Calculations
Observations (in print)	Observations (made available)	Δ Consumer surplus
100	50	1,280,593
150	100	2,036,443
175	150	2,529,921
200	200	3,023,399
250	300	4,010,356

Notes: Estimates are reported in 2012 dollars per year. Columns 1 and 2 denote assumptions on the number of currently protected in-print titles (column 1) and the number of protected titles that would move into print if the copyright is lifted (column 2). Column 3 reports the calculated industry-wide change in consumer surplus based on these assumptions.

holder maximizes profits from license fees and royalties for all editions. However, even though publishers pay a fee to the copyright holder, the costs of *production* are likely similar across copyright regimes.[42] Taking as given the inferred fixed and marginal costs of production from free entry among public domain works and the observed prices, I calculate the joint profits for copyright holders and publishers of protected editions.[43]

Formally, public domain firms (publishers) compete with each other in a two-stage entry game. In the first stage, they choose whether to enter an edition. In the second stage, they compete in Bertrand fashion. The marginal costs are given by the Lerner index:

$$(5) \qquad \frac{p_j - c_j}{p_j} = -\frac{1}{\epsilon_j},$$

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 estimation.

Given the prices in stage 2, and relaxing the integer constraint under free entry, a public domain edition j's variable profits π_j equal its fixed costs of production.[44] That is,

$$(6) \qquad \pi_j = (p_j - c_j)Ms_j = FC_j,$$

where $s_j = \frac{\exp\{\delta_j/(1 - \sigma)\}}{D_w^\sigma\left(1 + \sum_{w' \in W} D_{w'}^{1-\sigma}\right)}$ depends on edition characteristics and the level of competition, as in the demand model.

---

[42] Recall that the estimated creative qualities are not statistically different across the regimes.

[43] From a welfare perspective, who gets these profits is irrelevant.

[44] I treat the fixed cost as a flow that occurs every year, similar to Eizenberg (2014) and Fan and Yang (2014), because technological innovations are not needed for publication.

B. *Calculation and Profits*

I first recover the format-specific marginal costs from the demand estimation. While I set marginal costs for public domain e-books to zero, marginal costs for hardcover and paperback editions are obtained using equation (5). For editions of protected titles, I subtract royalties of 15 percent of the book's price from these costs because royalties are merely a transfer of wealth.[45] The average marginal costs for protected titles are \$16.63 for hardcover editions and \$9.31 for paperback editions.[46] They are similar for public domain works.

Given the marginal costs, I calculate the variable profits for public domain works, which equal their fixed costs of production following equation (6). For hardcover and paperback editions, average annual fixed costs are \$328 and \$268, respectively, and for e-books, they are only \$23.[47]

Then, assuming no differences in costs across copyright regimes, the difference between variable profits of protected editions and the fixed costs of production constitutes the edition-specific joint profits to the copyright holder and publishers. Thus, the title-specific joint profits for protected title w with quality rank r are

$$(7) \qquad \Pi_w^I = \sum_{j \in wH} \left[\pi_j - FC_{rH} \right] + \sum_{j \in wP} \left[\pi_j - FC_{rP} \right] + \sum_{j \in wE} \left[\pi_j - FC_{rE} \right],$$

where Π_w^I denotes the total profit from title w under copyright regime I, $\pi_j = (p_j - c_{rk})Ms_j$ is edition j's variable profit in format $k \in \{H,P,E\}$, and the title- and format-specific costs c_{rk} and FC_{rk} are the average costs for editions of the public domain title with the same quality rank r. In what follows, I report the title-specific joint profits, as calculated in equation (7).

Per-Title Profit.—On average, the in-print protected titles in my dataset generate a profit of \$5,460, compared to zero profits (by construction) in the public domain. These average profits are much smaller than the average differences in consumer surplus of \$9,982. This discrepancy suggests large welfare losses due to the copyright extension.

Most titles generate more total surplus in the public domain than they would if protected. However, some exceptions exist. For example, the protected book with the second largest quality rating (Pearl Buck's *The Good Earth*) generates joint profits of \$139,000 and a consumer surplus of \$59,000 annually, whereas its public domain equivalent (Sinclair Lewis's *Babbitt*) generates 0 profits and a consumer surplus of \$74,000. Similarly, Erich Maria Remarque's *All Quiet on the Western Front* generates \$126,000, which is more than the consumer surplus generated by its public domain equivalent. Several other titles, including Margaret Mitchell's *Gone with the Wind* and Thornton Wilder's *The Bridge of San Luis Rey*, earn

---

[45] Greco, Milliot, and Wharton (2013, 156) lists a standard royalty rate of 10 percent to 15 percent of the price.

[46] The marginal costs reflect costs of production. For example, the estimated marginal costs are strongly positively correlated with the number of pages in the edition, conditional on the edition's format.

[47] If the cost estimates are "too low," then my profit measures can be seen as upper bounds of true profits. As we will see, this will not change my qualitative results.

between $10,000 and $40,000 annually—amounts comparable to the previously calculated difference in consumer surplus from a move into the public domain.

Some superstars—titles like *Gone with the Wind* or *The Good Earth*—might generate more surplus if they remain protected by copyright. Owners of works that stood the test of time (including Disney's *Mickey Mouse*) have been able to lobby for copyright extensions in the past, even though the negative effect of the extension on consumers outweighs the positive effect on producers on average. Without changing incentives for creation, the 1998 Copyright Term Extension was welfare decreasing unless the copyright holders or publishers used the added profits for further innovation.

VI. Conclusion

The Sonny Bono Copyright Term Extension Act was passed in 1998, preventing copyright-protected works from moving into the public domain for at least 20 years. This paper shows that the extension led to fewer editions per title and higher prices, relative to public domain works. As a result, consumer surplus would increase more than producer surplus would decrease for most titles when their copyright protection is lifted.

The copyright extension of 1998 decreased welfare for the publishing industry. This is true especially for the large stock of titles which have become orphans. The Sonny Bono Copyright 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: low- and medium-quality works that have lost much of their appeal since their creation.

A policy that extends the copyright for those titles with the biggest stakes while allowing other titles to move into the public domain would obviously increase total welfare. Since such a policy is difficult to implement, any further increase of the copyright term would likely be welfare decreasing.

APPENDIX A. TECHNICAL APPENDIX

A1. *Demand Estimation: First-Stage Results*

The main analysis includes a discussion of the validity and relevance of the instruments in the demand estimation. Table A1 reports the first-stage coefficients of the instruments. All coefficients have the expected signs and are statistically significant. The coefficients for the other variables are as expected as well.

A2. *Demand Estimation: Adding Different Constants*

Many editions are not sold at all in a given month (almost two-thirds of my observations). In the paper, I add a small constant (0.0001 units) to each edition's monthly unit sales. While the demand estimates vary depending on what constant is added, adding "large" constants would lead to distortions of the demand estimates

TABLE A1—DEMAND ESTIMATION: FIRST-STAGE RESULTS

	DV: Price (1)	DV: Inside share (2)
Sellers per edition	−0.228 (0.013)	0.092 (0.006)
Editions per nest	−0.861 (0.120)	−0.151 (0.060)
Observations	8,756	8,756
Adjusted R^2	0.569	0.390
Number of titles	242	242
F-statistic for both instruments	16.90	26.92

TABLE A2—DEMAND ESTIMATION FOR DIFFERENT CONSTANTS ADDED TO DEMAND

	Add 0.00001 (1)	Add 0.0001 (2)	Add 0.001 (3)	Add 0.01 (4)	Add 0.1 (5)	Add 1 (6)
Price	−0.170 (0.065)	−0.138 (0.052)	−0.105 (0.040)	−0.073 (0.027)	−0.043 (0.016)	−0.021 (0.008)
σ(title)	0.610 (0.136)	0.629 (0.126)	0.654 (0.112)	0.690 (0.095)	0.740 (0.074)	0.789 (0.056)
Available as new	2.121 (0.646)	1.715 (0.506)	1.314 (0.371)	0.922 (0.246)	0.558 (0.139)	0.273 (0.065)
Major publisher	−0.169 (0.209)	−0.134 (0.168)	−0.099 (0.128)	−0.065 (0.089)	−0.035 (0.053)	−0.016 (0.027)
Edition age	−0.030 (0.011)	−0.024 (0.009)	−0.018 (0.007)	−0.012 (0.005)	−0.007 (0.003)	−0.003 (0.001)
Picture shown	0.434 (0.275)	0.328 (0.212)	0.226 (0.153)	0.133 (0.098)	0.054 (0.053)	0.001 (0.025)
Hardcover	−2.336 (0.618)	−2.088 (0.508)	−1.836 (0.402)	−1.578 (0.302)	−1.294 (0.210)	−0.926 (0.135)
e-book	−4.218 (0.826)	−3.565 (0.671)	−2.909 (0.520)	−2.251 (0.378)	−1.601 (0.253)	−0.993 (0.157)
Paperback	−3.566 (0.445)	−3.072 (0.369)	−2.577 (0.295)	−2.079 (0.225)	−1.572 (0.163)	−1.041 (0.114)
Mean elasticity	−4.33	−3.66	−2.99	−2.30	−1.60	−0.92
Observations	8,756	8,756	8,756	8,756	8,756	8,756
$\bar{R}^2$	0.747	0.756	0.768	0.781	0.786	0.766

because demand is low throughout. The demand estimates for different constants (from 0.00001 to 1) added are shown in Table A2.

The differences in the estimated coefficients lead to differences in the estimated consumer surplus effects. Not surprisingly, the impact on consumer surplus becomes larger as I add a larger constant to each edition's monthly unit sales. Per title in my dataset, a move into the public domain adds $9,982 when adding 0.0001 units, as shown in the paper. It would lead to an average increase of $12,973 when instead adding 0.001 units and an increase of $8,112 when adding only 0.00001 units.

TABLE A3—INDUSTRY-WIDE CHANGES IN CONSUMER SURPLUS

Assumptions		Calculated effect		
Observations (in print)	Observations (made available)	ΔCS (add 0.00001)	ΔCS (add 0.0001)	ΔCS (add 0.001)
100	50	997,002	1,292,869	1,588,357
150	100	1,604,237	2,052,414	2,497,514
175	150	2,016,587	2,545,296	3,067,070
200	200	2,428,938	3,038,179	3,636,627
250	300	3,253,640	4,023,943	4,775,740

Table A3 shows annual industry impacts of the extension (see Table 5 in the paper), as I add different constants to unit sales. For my "best guess" numbers of impacted titles, the change in consumer surplus lies between $2 million and $3 million annually.

Finally, the impact on estimated profits varies as well, with average per-title profits lying between $4,000 and $5,500 among the titles in my dataset. Overall, the qualitative results in my paper do not depend on what constant I add to unit demand.

REFERENCES

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

Berndt, Ernst R., and Murray L. Aitken. 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, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *RAND Journal of Economics* 25 (2): 242–62.

Berry, Steven, Alon Eizenberg, and Joel Waldfogel. 2015. "Optimal Product Variety in Radio Markets." NBER Working Paper 21621.

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

Biasi, Barbara, and Petra Moser. 2016. "Effects of Copyrights on Science—Evidence from the WWII Book Republication Program." https://www.gc.cuny.edu/CUNY_GC/media/CUNY-Graduate-Center/PDF/Programs/Economics/Course%20Schedules/PT.pdf.

Brynjolfsson, Erik, Yu Hu, and Michael D. Smith. 2003. "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers." *Management Science* 49 (11): 1580–96.

Cardell, N. Scott. 1997. "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity." *Econometric Theory* 13 (2): 185–213.

Chaudhuri, Shubham, Pinelopi K. Goldberg, and Panle Gia. 2006. "Estimating the Effects of Global Patent Protection in Pharmaceuticals: A Case Study of Quinolones in India." *American Economic Review* 96 (5): 1477–1514.

Chevalier, Judith, and Austan Goolsbee. 2003. "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com." *Quantitative Marketing and Economics* 1 (2): 203–22.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67 (3): 297–308.

Eizenberg, Alon. 2014. "Upstream Innovation and Product Variety in the US Home PC Market." *Review of Economic Studies* 81 (3): 1003–45.

Fan, Ying, and Chenyu Yang. 2014. "Competition, Product Proliferation and Welfare: A Study of the US Smartphone Market." https://ideas.repec.org/p/net/wpaper/1414.html.

Furman, Jeffrey L., and Scott Stern. 2011. "Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research." *American Economic Review* 101 (5): 1933–63.

Greco, Albert N., Jim Milliot, and Robert M. Wharton. 2013. *The Book Publishing Industry*. 3rd ed. Abingdon-on-Thames: Routledge.

Handke, Christian. 2011. *Economic Effects of Copyright: The Empirical Evidence So Far*. Rotterdam: National Academies.

Heald, Paul J. 2008. "Property Rights and the Efficient Exploitation of Copyrighted Works: An Empirical Analysis of Public Domain and Copyrighted Fiction Bestsellers." *Minnesota Law Review* 92: 1031–63.

Heald, Paul J. 2018. "Copyright Reversion to Authors (and the Rosetta Effect): An Empirical Study of Reappearing Books." http://www.serci.org/congress_documents/2018/heald.pdf.

Imbens, Guido W., and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2): 615–35.

Korda, Michael. 2001. *Making the List: A Cultural History of the American Bestseller 1900–1999: As Seen through the Annual Bestseller Lists of Publishers Weekly*. New York: Barnes and Noble Books.

Landes, William M., and Richard A. Posner. 2003. "Indefinitely Renewable Copyright." *University of Chicago Law Review* 70 (2): 471–518.

Landes, William M., and Richard A. Posner. 2009. *The Economic Structure of Intellectual Property Law.* Cambridge, MA: Harvard University Press.

Li, Xing, Megan MacGarvie, and Petra Moser. 2015. "Dead Poet's Property—How Does Copyright Influence Price?" NBER Working Paper 21522.

Mankiw, N. Gregory, and Michael D. Whinston. 1986. "Free Entry and Social Inefficiency." *RAND Journal of Economics* 17 (1): 48–58.

Miller, Claire Cain, and Julie Bosman. 2013. "Siding with Google, Judge Says Book Search Does Not Infringe Copyright." *New York Times*, November 14. https://www.nytimes.com/2013/11/15/business/media/judge-sides-with-google-on-book-scanning-suit.html.

Morton, Fiona M. Scott. 2000. "Barriers to Entry, Brand Advertising, and Generic Entry in the US Pharmaceutical Industry." *International Journal of Industrial Organization* 18 (7): 1085–1104.

Nagaraj, Abhishek. 2017. "Does Copyright Affect Reuse? Evidence from Google Books and Wikipedia." *Management Science* 64 (7): 3076–90.

Peukert, Christian, and Imke Reimers. 2018. "Digital Disintermediation and Efficiency in the Market for Ideas." https://www.aeaweb.org/conference/2019/preliminary/paper/2BE6NEs3.

Quan, Thomas W., and Kevin R. Williams. 2018. "Product Variety, Across-Market Demand Heterogeneity, and the Value of Online Retail." *RAND Journal of Economics* 49 (4): 877–913.

Regan, Tracy L. 2008. "Generic Entry, Price Competition, and Market Segmentation in the Prescription Drug Market." *International Journal of Industrial Organization* 26 (4): 930–48.

Reiffen, David, and Michael R. Ward. 2005. "Generic Drug Industry Dynamics." *Review of Economics and Statistics* 87 (1): 37–49.

Reimers, Imke, and Joel Waldfogel. 2017. "Throwing the Books at Them: Amazon's Puzzling Long Run Pricing Strategy." *Southern Economic Journal* 83 (4): 869–85.

Reimers, Imke. 2019. "Copyright and Generic Entry in Book Publishing: Dataset." *American Economic Journal: Microeconomics*. https://doi.org/10.1257/mic.20170100.

Thomas, Catherine. 2011. "Too Many Products: Decentralized Decision Making in Multinational Firms." *American Economic Journal: Microeconomics* 3 (1): 280–306.

Varian, Hal R. 2006. "Copyright Term Extension and Orphan Works." *Industrial and Corporate Change* 15 (6): 965–80.

Williams, Heidi L. 2013. "Intellectual Property Rights and Innovation: Evidence from the Human Genome." *Journal of Political Economy* 121 (1): 1–27.

This article has been cited by:

1. Jeremy Watson, Megan MacGarvie, John McKeon. 2022. It Was 50 Years Ago Today: Recording Copyright Term and the Supply of Music. *Management Science* **36**. . [Crossref]

2. Barbara Biasi, Petra Moser. 2021. Effects of Copyrights on Science: Evidence from the WWII Book Republication Program. *American Economic Journal: Microeconomics* **13**:4, 218-260. [Abstract] [View PDF article] [PDF with links]

3. Imke Reimers, Joel Waldfogel. 2021. Digitization and Pre-Purchase Information: The Causal and Welfare Impacts of Reviews and Crowd Ratings. *American Economic Review* **111**:6, 1944-1971. [Abstract] [View PDF article] [PDF with links]

4. Kristelia García, James Hicks, Justin McCrary. 2020. Copyright and Economic Viability: Evidence from the Music Industry. *Journal of Empirical Legal Studies* **17**:4, 696-721. [Crossref]

5. Martin Watzinger, Thomas A. Fackler, Markus Nagler, Monika Schnitzer. 2020. How Antitrust Enforcement Can Spur Innovation: Bell Labs and the 1956 Consent Decree. *American Economic Journal: Economic Policy* **12**:4, 328-359. [Abstract] [View PDF article] [PDF with links]

6. Kristelia Garcia, James Hicks, Justin McCrary. 2020. Copyright and Economic Viability: Evidence from the Music Industry. *SSRN Electronic Journal* . [Crossref]

7. Abhishek Nagaraj, Imke Reimers. 2019. Digitization and the Demand for Physical Works: Evidence from the Google Books Project. *SSRN Electronic Journal* . [Crossref]