

## Article

# Property Rights and the Efficient Exploitation of Copyrighted Works: An Empirical Analysis of Public Domain and Copyrighted Fiction Bestsellers

Paul J. Heald<sup>†</sup>

The primary rationale for copyright protection is grounded in the incentive-to-create theory,<sup>1</sup> which holds that Congress should grant authors the exclusive right to reproduce their works in order to ensure the public a steady supply of new books, music, movies, and the like.<sup>2</sup> Congress has, however, on

---

<sup>†</sup> Allen Post Professor of Law, University of Georgia School of Law. I would like to thank the University of Georgia Law School for financial support for this Article and Tedra Hobson, Christian Perrin, Stephanie Steele, and Sivram Prasad for their work as research assistants. I would also like to thank Fiona MacMillan and the AHRB Copyright Research Network, the faculty and students at Roberta Kwall's DePaul IP Seminar, participants at the Berkeley IP Scholars Conference, the Northwestern University Law School faculty colloquium, Doug Baird, Robert Bartlett, Erica Hashimoto, Justin Hughes, Doug Lichtman, David Mustard, Usha Rodrigues, David Shipley, and Jim Smith for their invaluable comments. A special debt of gratitude is owed to Jennifer LeRademacher, YiMei Cai, and Professor Jaxk Reeves of the University of Georgia Statistics Department, who spent many hours doing the regressions and other statistical work quoted at length in Appendix A. Copyright © 2008 by Paul J. Heald.

1. See *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 450 (1984) ("The purpose of copyright is to create incentives for creative effort."); *Mazer v. Stein*, 347 U.S. 201, 219 (1954) ("The economic philosophy behind the clause empowering Congress to grant patents and copyrights is the conviction that encouragement of individual effort by personal gain is the best way to advance public welfare through the talents of authors and inventors in 'Science and useful Arts.'"); Brett M. Frischmann & Mark A. Lemley, *Spillovers*, 107 COLUM. L. REV. 257, 284 (2007) ("The rights granted by copyright law—specifically, the § 106 rights to reproduce, display, perform, distribute, and make derivative works—provide incentives to create . . ."); Sara Stadler, *Copyright as Trade Regulation*, 155 U. PA. L. REV. 899, 927 (2007) ("[C]opyright law exists *at least* for the purpose of punishing acts that might diminish incentives to create . . .").

2. Copyright law grants authors the exclusive right to reproduce, distribute, perform, display, and prepare derivative works. See 17 U.S.C. § 106

several occasions asserted the power to extend the duration of copyright protection to works that have already been created.<sup>3</sup> Most recently, in 1998, Congress extended an additional twenty years of protection to existing works.<sup>4</sup> Although the wisdom of earlier retroactive extensions in 1790, 1831, 1909, and 1964–1976 was not seriously challenged, objections to the 1998 extension had Congress, the Supreme Court, and economists scrambling to find an alternative to the traditional incentive-to-create rationale.<sup>5</sup>

Why the discomfort? Quite simply, works that have been in existence for years do not require new additional protection to ensure their creation. Their authors, in response to previous legislative stimulus, have already completed their jobs.<sup>6</sup> In the absence of a need to stimulate the creation of a work that already exists, analysts were challenged to provide a reason why works should not fall into the public domain and be made available for any and all to copy for free.<sup>7</sup> Indeed, until 1998, large numbers of works fell into the public domain every year.<sup>8</sup> For the works subject to the present study (bestselling novels published between 1913–1932), the initial term of copyright was twenty-eight years, plus a twenty-eight-year renewal term. In 1964, Congress extended that term nineteen more years for

---

(2000 & Supp. V 2007). Protected works include literary works, musical works, dramatic works, choreographic works, pictorial/graphic/sculptural works, motion pictures, sound recordings, and architectural works. *See id.* § 102(a) (2000).

3. The first copyright statute, enacted in 1790, extended fourteen years of protection to existing works and those yet-to-be-created. *See* Act of May 31, 1790, ch. 15, 1 Stat. 124. In 1831, Congress extended protection to existing and future works to twenty-eight years, plus a fourteen-year renewal period. *See* Act of Feb. 3, 1831, ch. 16, 4 Stat. 436. In 1909, Congress extended the renewal period to twenty-eight years for all works. *See* Act of Mar. 4, 1909, ch. 320, 35 Stat. 1075.

4. *See* 17 U.S.C. § 302(a) (2000).

5. *See* Opposing Copyright Extension, <http://homepages.law.asu.edu/~dkarjala/> (providing a comprehensive anti-term-extension website maintained by Professor Dennis Karjala at Arizona State University).

6. *See* Paul J. Heald & Suzanna Sherry, *Implied Limits on the Legislative Power: The Intellectual Property Clause as an Absolute Constraint on Congress*, 2000 U. ILL. L. REV. 1119, 1169 (using George Gershwin's 1925 *Concerto in F* as an example).

7. *See* *Compco Corp. v. Day-Brite Lighting, Inc.*, 376 U.S. 234, 237 (1964) (explaining why state laws cannot interfere with the public domain because “to forbid copying would interfere with federal policy . . . allowing free access to copy whatever the federal patent and copyright laws leave in the public domain”).

8. *Id.*

those works that had been properly renewed. For example, works published in 1922 fell into the public domain in 1997 (1922 + 28-year initial term + 28-year renewal + 19-year extension = 1997). Those published in 1923 would have fallen into the public domain in 1998 but for the term-extension legislation, which added another twenty years of protection. In theory, works published in 1923 will fall into the public domain in 2018. The effect of the present legislation is to prevent any works from falling into the public domain before that date.<sup>9</sup>

In order to justify this moratorium on the growth of the public domain, legislators, courts, and scholars have offered a rationale that posits the undesirability of letting works fall into the public domain. The theory suggests that works without owners will suffer from underexploitation. A report from the House of Representatives stated that the 1998 extension would “provide copyright owners generally with the incentive to restore older works and further disseminate them to the public.”<sup>10</sup> Similarly, in upholding the extension against a constitutional challenge in *Eldred v. Ashcroft*,<sup>11</sup> the Supreme Court concluded that Congress “rationally credited projections that longer terms would encourage copyright holders to invest in the restoration and public distribution of their works.”<sup>12</sup> Prominent economists William M. Landes and Richard A. Posner reach a similar conclusion: “[A]n absence of copyright protection for intangible works may lead to inefficiencies because . . . of impaired incentives to invest in maintaining and exploiting these works.”<sup>13</sup> Accordingly, they assert that copyright should be indefinitely renewable for works that hold significant value over time.<sup>14</sup> If works that fall into the public domain become less accessible, then retroactive extension could be justified on public interest grounds.

This assertion is, of course, empirically testable. Because works regularly fell into the public domain throughout American history, one can test whether the availability of works from the same general era is dependent on their legal status (still protected by copyright versus fallen into the public domain).

---

9. See 17 U.S.C. § 302; JULIE COHEN ET AL., COPYRIGHT IN A GLOBAL INFORMATION ECONOMY 154–55 (2006).

10. H.R. REP. NO. 105-452, at 4 (1998).

11. 537 U.S. 186 (2003).

12. *Id.* at 207.

13. William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 475 (2003).

14. *Id.* at 518.

Because it is relatively easy to track the availability of books, this study tests the hypothesis that copyright law is presently necessary to prevent underexploitation by examining best-selling fiction published from 1913–1932. The books studied that were published from 1913–1922 fell into the public domain from 1988–1997, seventy-five years after their initial publication dates. However, due to the Copyright Term Extension Act of 1998, the works from 1923–1932 remain protected until at least 2018.<sup>15</sup> A comparison of the two sets of books helps answer the question whether copyright law is necessary to prevent the underexploitation of popular fiction from this era. Part I briefly provides some background and describes the methodology of the study.

Parts II and III show that books in the public domain data set were in print at the same rate as copyrighted books until 2000, and thereafter were in print at a significantly higher rate with significantly more editions per book. Although the average lowest prices for the complete sets of public domain and proprietary fiction are the same, the two smaller subsets of the twenty most popular books from 1913–1922 and 1923–1932 show that the titles in the public domain subset are significantly less expensive than their protected counterparts.<sup>16</sup> Overall, the data suggest that, as to already published books, copyright extension imposes deadweight losses without offsetting efficiency gains. The data also suggest that whether the extension was as deleterious in the context of other types of creative works may depend on the cost of producing and distributing the work: the lower the cost of production, the lower the likelihood of underexploitation.

## I. BACKGROUND AND METHODOLOGY

History and previous studies of the availability of published books help explain the methodology of the study, which focuses on 168 copyright-protected and 166 public domain best-sellers from the same historical era (1913–1932).

---

15. 17 U.S.C. § 302(a) (2000). Given the consistent pattern of retroactive extensions enacted by Congress, it seems likely that the term will be extended once again before 2018.

16. The formal statistical regressions supporting these conclusions are set forth at length in Appendix A.

## A. BACKGROUND

Worries about underexploitation of popular novels have some basis in history. Evidence from the nineteenth century suggests that the absence of copyright protection could lead to the underexploitation of some books, especially when production costs are high. Until 1891, the United States did not grant copyright protection to foreign works, so numerous valuable works by English authors such as Charles Dickens, Lord Alfred Tennyson, Sir Walter Scott, and Robert Browning were effectively in the public domain.<sup>17</sup> According to Professor B. Zorina Khan, costly races by American publishers to be the first to print the newest English fiction resulted in “ruinous competition . . . likely to drive prices down to marginal cost, in which case the high initial fixed investments would not be recovered.”<sup>18</sup> At a time when typesetting was costly and labor-intensive, the absence of a single rights-holder caused serious inefficiencies. In the absence of an assurance of exclusivity, publishers were afraid to make the significant financial investment to typeset and print a book that might arrive second or third on the market. Losing the race to be the first to market the book to the consuming public might well mean a failure to recoup the sunk costs of publication.<sup>19</sup>

One solution to the problem might have been to grant copyright protection to a single publisher. Armed with the exclusive right to reproduce the work, a publisher would not have to worry about being beaten to the market. Interestingly, although Congress did not come to the rescue until 1891, a small group of publishers appeared to solve the problem through the creation of a “synthetic copyright,” whereby they colluded to assign exclusive rights in various English works among themselves.<sup>20</sup> Khan notes that British publishers had similarly colluded earlier over publishing rights to public domain works by Shakespeare and Fielding by the creation of cartels called “printing congers.”<sup>21</sup> Private rules substituted for a formal property right and facilitated the production of books for the reading public. Given the high cost of producing a book in the

---

17. B. Zorina Khan, *Does Copyright Piracy Pay? The Effects of U.S. International Copyright Laws on the Market for Books, 1790–1920*, at 21, 23–24 (Nat’l Bureau of Econ. Research, Working Paper No. 10271, 2004).

18. *Id.* at 21.

19. *Id.* at 21–24.

20. *Id.* at 23–25.

21. *Id.* at 24.

nineteenth century, the absence of copyright (had publishers not creatively colluded) might have led to valuable works being unavailable.

Those currently advocating for the extension of copyright terms for existing works argue that we will see a shortage of popular classics if each work does not have a single owner assigned the sole right to publish.<sup>22</sup> The present study seeks to answer the question whether such a shortage actually materializes when works fall into the public domain.

Previous studies help to narrow the focus of what kind of book shortage is most relevant. Landes and Posner studied the rate of copyright renewal for books (pre-1976 law required renewal to maintain protection) and found that “fewer than 11 percent of the copyrights registered between 1883 and 1964 were renewed at the end of their 28-year term, even though the cost of renewal was small.”<sup>23</sup> This suggests that after only twenty-eight years, relatively few books hold any significant value. If they were valuable, why would their owners voluntarily abandon protection?

Landes and Posner also report that of “10,027 books published in the United States in 1930, only 174, or 1.7%, were still in print in 2001.”<sup>24</sup> Because so few books hold their value over time, they argue persuasively that questions of efficient exploitation, and therefore copyright term extension, only arise when considering the most valuable works produced in any given period.<sup>25</sup> They suggest it would be senseless to provide long terms of protection for the vast majority of works that do not hold their value. Because most copyright owners did not exploit their works, Landes and Posner concluded that extending protection yet further makes little sense in the vast majority of cases. Since the argument for extending protection of insignificant works is very weak, the present study focuses on bestsellers only, those works most likely to hold their value over time. For these works, their argument goes, copyright should be indefinitely renewable in order to ensure their availability.<sup>26</sup>

---

22. See *supra* notes 10–13 and accompanying text.

23. Landes & Posner, *supra* note 13, at 473.

24. *Id.* at 474.

25. *Id.* at 472–73.

26. *Id.* at 471–75.

## B. METHODOLOGY

The primary data set consists of 334 best-selling books collected from three sources. Most titles, approximately fifteen per year, were taken from the end-of-year top-ten bestseller lists compiled by *Publisher's Weekly*, a major print trade publication.<sup>27</sup> This list was supplemented by a second prominent year-by-year listing of popular American books and music<sup>28</sup> and by the all-time bestsellers list compiled by *Publisher's Weekly*.<sup>29</sup> Both captured some popular titles that were not top-ten sellers in any particular year, for example, *A Portrait of the Artist as a Young Man*<sup>30</sup> by James Joyce.

The study examines the years 1913–1932 because of the importance of comparing public domain and copyrighted fiction of approximately the same age. Since 1923 is an important cut-off year (all fiction published before that date is in the public domain),<sup>31</sup> the ten years prior to and immediately after 1923 were selected. The objective was not to discover every popular work of fiction published from 1913–1932, but rather to capture a random sample large enough for statistical purposes. The in-print status and number of available editions for each work were then tracked at five-year intervals<sup>32</sup> beginning with the sixtieth year after publication by consulting Bowker's *Books in Print*.<sup>33</sup> For the 166 works published from 1913–1922, the seventy-fifth year after publication marked their entrance into the public domain. The status of every book in 2006 was tracked through Bowker's *Books in Print* online service.<sup>34</sup> The

27. See ALICE PAYNE HACKETT, 70 YEARS OF BEST SELLERS, 1895–1965, at 111–45 (1967).

28. JULIUS MATTFELD, VARIETY MUSIC CAVALCADE (1962).

29. See HACKETT, *supra* note 27, at 12–30.

30. JAMES JOYCE, A PORTRAIT OF THE ARTIST AS A YOUNG MAN (1916).

31. See 17 U.S.C. § 301 (2000).

32. The decision to measure in-print status at five-year intervals instead of every year was purely pragmatic. There is no online database where information on historic availability can be found; the manual search through Bowker's *Books in Print* is highly labor intensive. Limited resources dictated sampling at five-year intervals.

33. R.R. BOWKER CO., BOOKS IN PRINT (1966–2006). Bowker publishes several volumes every year that list all books currently “in print” (usually meaning in the publisher's inventory) in the United States. The listing is alphabetical by author, title, or publisher and includes the publisher's posted price information. Current in-print status, but not historical information, is available at <http://www.booksinprint.com>.

34. See Booksinprint.com Professional, <http://www.booksinprint.com> (last visited Mar. 10, 2008).

full set of titles can be found in Appendix B.<sup>35</sup> Arguably, comparative sales data would be a better measure of availability than in-print status; however, historical sales data are generally not publicly available.

To meet the objection that even bestsellers do not necessarily hold their value over time, the study also identifies and examines a smaller subset of the forty most currently popular works from 1913–1932 (durable books).<sup>36</sup> Since the works from 1913–1922 are in the public domain, free for any publisher to exploit, the twenty titles with the highest number of editions currently in print were selected.<sup>37</sup> The level of competition among publishers (a minimum of seventeen different publishers for each title) gave some objective evidence of the enduring popularity of those titles. The twenty most durable books from 1923–1932, all still protected by copyright, were chosen somewhat more subjectively. The number of current editions was deemed to be an imperfect indicator of popularity, because a paucity of editions might be indicative of an owner's reluctance to grant a license, as opposed to a lack of consumer demand. So, in addition to considering the number of editions, the expertise of those with a knowledge of American literature was consulted, resulting in the selection of the twenty copyrighted titles.<sup>38</sup> The titles should be familiar to those with even a passing interest in literature, and the content of the list has raised no objections at conferences where the study has been presented.<sup>39</sup> Most importantly, the substitution of different titles

35. Available at <http://local.law.umn.edu/lawreview/issues.html>.

36. Several of these books were identified in a fourth source. See ASA DON DICKINSON, *THE WORLD'S BEST BOOKS: HOMER TO HEMINGWAY* (1953).

37. The twenty public domain titles are SHERWOOD ANDERSON, *WINEBURG, OHIO* (1919); EDGAR BURROUGHS, *TARZAN OF THE APES* (1914); WILLA CATHER, *MY ÁNTONIA* (1918); WILLA CATHER, *O PIONEERS!* (1913); WILLA CATHER, *THE SONG OF THE LARK* (1915); F. SCOTT FITZGERALD, *THE BEAUTIFUL AND THE DAMNED* (1922); F. SCOTT FITZGERALD, *THIS SIDE OF PARADISE* (1920); ZANE GREY, *THE LONE STAR RANGER* (1915); JAMES JOYCE, *A PORTRAIT OF THE ARTIST AS A YOUNG MAN* (1916); JAMES JOYCE, *DUBLINERS* (1914); JAMES JOYCE, *ULYSSES* (1922); D.H. LAWRENCE, *SONS AND LOVERS* (1913); SINCLAIR LEWIS, *MAIN STREET* (1920); SINCLAIR LEWIS, *BABBITT* (1922); W. SOMERSET MAUGHAM, *OF HUMAN BONDAGE* (1915); ELEANOR H. PORTER, *POLLYANNA* (1913); RAFAEL SABATINI, *CAPTAIN BLOOD* (1922); RAFAEL SABATINI, *SCARAMOUCHE* (1921); BOOTH TARKINGTON, *THE MAGNIFICENT AMBERSONS* (1918); EDITH WHARTON, *THE AGE OF INNOCENCE* (1920).

38. See *infra* note 39 (listing the twenty copyrighted titles in the study). The individual readers thanked on the first page of this Article were asked to look at the full set of copyright titles and suggest additions to the top twenty.

39. The twenty copyrighted titles are PEARL S. BUCK, *THE GOOD EARTH* (1931); WILLA CATHER, *DEATH COMES FOR THE ARCHBISHOP* (1927); JOHN DOS



from the larger data set of copyrighted titles would not significantly change the statistical comparison, but rather only exacerbate the differences in availability and price between the public domain and copyrighted subsets described in Part III.<sup>40</sup>

The subset of durable books was once again measured by in-print status; however, a shelf-space study in prominent national book store chains was conducted to compare their immediate availability to the purchasing public.

## II. COMPARING AVAILABILITY AND PRICE OF PUBLIC DOMAIN AND COPYRIGHTED BESTSELLERS

One goal of the study is to determine whether extending a property right in an existing work of fiction is necessary to ensure its adequate exploitation. To that end, the set of 166 public domain works, published from 1913–1922, and the set of 168 copyrighted works, published from 1923–1932, were compared over time in terms of their in-print status, number of available editions in 2006, and their current 2006 price. Part II.A explains the availability comparison first in absolute terms (for example, whether a book is in print in 1999), and then in relative terms (for example, whether a book is in print at seventy years after publication). Part II.B analyzes the relative number of editions per book in 2006, as well as price data.<sup>41</sup>

---

PASSOS, MANHATTAN TRANSFER (1925); THEODORE DREISER, AN AMERICAN TRAGEDY (1925); WILLIAM FAULKNER, SANCTUARY (1931); WILLIAM FAULKNER, THE SOUND AND THE FURY (1929); F. SCOTT FITZGERALD, THE GREAT GATSBY (1925); DASHIELL HAMMETT, THE MALTESE FALCON (1930); ERNEST HEMINGWAY, A FAREWELL TO ARMS (1929); ERNEST HEMINGWAY, THE SUN ALSO RISES (1926); ALDOUS HUXLEY, BRAVE NEW WORLD (1932); D.H. LAWRENCE, LADY CHATTERLEY'S LOVER (1930); SINCLAIR LEWIS, ARROWSMITH (1925); SINCLAIR LEWIS, ELMER GANTRY (1927); A.A. MILNE, WINNIE-THE-POOH (1926); CHARLES NORDHOFF, MUTINY ON THE BOUNTY (1932); ERICH MARIA REMARQUE, ALL QUIET ON THE WESTERN FRONT (1929); THORNTON WILDER, THE BRIDGE OF SAN LUIS REY (1927); THOMAS WOLFE, LOOK HOMEWARD, ANGEL (1929); PERCIVAL CHRISTOPHER WREN, BEAU GESTE (1925).

40. One significant difference between the durable public domain books and the durable copyrighted books is the lowest average list price. The average 2006 price listed by [www.booksinprint.com](http://www.booksinprint.com) for the twenty durable copyrighted books was \$8.05. Of the 148 copyrighted bestsellers not chosen for the list of durable books, only ANITA LOOS, GENTLEMEN PREFER BLONDES (1926) (\$6), ELLIOT H. PAUL, IMPROMPTU (1923) (\$5), and EDITH WHARTON, OLD NEW YORK (1924) (\$7) were listed at under \$8.

41. The formal regressions that demonstrate the statistical significance of all the findings in this Section are provided in Appendix A.

## A. PROBABILITY OF BEING IN PRINT

## 1. Probability by Year

This Section demonstrates that from 1988–2001, public domain bestsellers were in print at the same rate as their copyright-protected counterparts. After 2001, the public domain books are available at a significantly higher rate than the copyrighted books.

Figure 1, *infra*, tracks the in-print status of both data sets from 1988–2006.<sup>42</sup> The year-by-year analysis starts in 1988, because that is when the books published from 1913–1922 first began to fall into the public domain. For example, books published in 1913 fell in the public domain in 1988. For this analysis, supplemental data was gathered about the in-print status of bestsellers published between 1907–1912. For example, the 1988 analysis compares the in-print status of books published in 1908 and 1913 (in the public domain in 1988), with books published in 1918, 1923, and 1928 (protected by copyright in 1988). Because in-print status was sampled in five-year intervals, the supplemental data on books published from 1907–1912 was needed to augment the number of available data points so that comparisons for each year from 1988–2006 always measured at least twenty-four books for each of the two data sets (but usually at least forty books for each). Comparative data for 2006 includes all 334 books in the study.

Although the general trend shown in Figure 1 for both sets of works shows increased availability over time, there is no statistically significant difference between the availability of the public domain works and copyrighted works between 1988–2001. Around 2001, however, the public domain works trend turned sharply upwards in terms of in-print status, reaching 98% in 2006, while only 74% of the books in the copyrighted data set are in print in 2006.

---

42. Because works were tracked at five-year intervals, two or three years worth of data were typically available for each data point. For example, data for 1998 public domain availability is taken from works published in years 1908, 1913, and 1918; data for 1998 copyrighted availability is taken from works published in 1923 and 1928.

Figure 1. Percentage in Print by Year

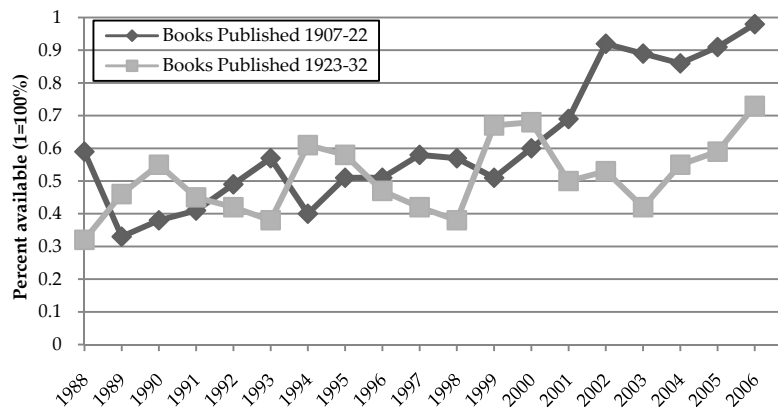


Figure 1 suggests that something significant happened around 2001. Comparative availability varies between 1988 and 2001, with some years (six) showing more public domain books in print, and other years (seven) showing more copyrighted books in print. On the average, the copyrighted books are ten years younger than the public domain books, which may have resulted in marginally increased demand, driving up their availability. On the other hand, publishers of public domain books did not have to pay a licensing fee, which made those titles marginally cheaper to print. In 2001, the availability curves disentangle and the in-print status for public domain works approaches 100% in 2004–2006. One could speculate that that improvements in optical scanning technology and software, lower costs of that technology, and the emergence of new business models allowed publishers of public domain materials to take advantage of the royalty-free status of such works, making it possible to satisfy the small demand for the least popular titles in the data set.<sup>43</sup>

Overall, the probability that works are in print indicate that lack of copyright protection does not lead to underexploitation. In fact, in recent years, more public domain works have been in print than copyrighted works.

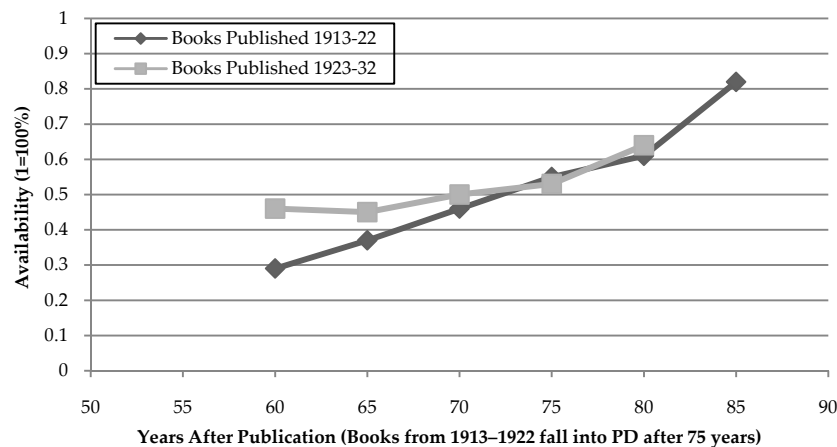
43. This was the case in the music publishing industry, where the appearance of new software in 2001 revolutionized the ability of music publishers to scan and edit sheet music. E-mail from Mark Schweizer, President, St. James Music Press, to author (Dec. 5, 2006, 12:31:01 CST) (on file with author).

## 2. Probability by Years After Publication

The full set of data can also be analyzed in terms of availability, measured by years, after publication. A complete comparison is hampered by the fact that some of the books in the copyrighted set are not yet seventy-five years old, and most are not yet eighty years old.

As Figure 2, *infra*, demonstrates, in years 60, 65, and 70 after publication, the newer set of books published from 1923–1932 are marginally more available. However, this difference cannot be due to copyright status, because the books from 1913–1922 did not fall into the public domain until their seventy-fifth year. Therefore, the data points in years 60, 65, and 70 all represent books still protected by copyright. The slightly greater availability for the 1923–1932 books continues in year 75 after publication, at the moment the books from 1913–1922 have fallen into the public domain, possibly because they are on the average ten years newer.<sup>44</sup> Public domain books surpass copyrighted books in year 80. Year 85 data is not available for books in the copyrighted data set yet, but the public domain books trend upward sharply at that point.

**Figure 2. Percent in Print as a Function of Years After Publication**



Given that 98% of the public domain books were in print in 2006, the trends for years 90–95 would likely approach 100%

44. In retrospect, it would have been more revealing to measure status as of year 76. Books falling into the public domain later in year 75 might have only been realistically available to exploit in year 76.

for that set. Whether the availability of copyrighted books will lag or also approach 100% remains to be seen.

Thus, Figure 2 shows that at seventy-five years after publication, copyrighted works are only marginally more available than public domain works. The trend at eighty years after publication seems to indicate that public domain works will be more available than copyrighted works. Regardless of whether the availability of copyrighted works approaches 100%, the current data do not suggest a problem of underexploitation due to the public domain status of a work.

#### B. NUMBER OF EDITIONS IN PRINT IN 2006

The number of editions in print of public domain and copyrighted works in 2006 also varies significantly. Because some publishers list dozens of versions of the same book in the *Books in Print* online database, the study counts only one edition per publisher. There were 1023 editions of the public domain books in print in 2006 and 405 editions of the copyrighted books, an average of 6.3 editions of each public domain book and 3.2 editions of each copyrighted book, respectively. If e-books<sup>45</sup> are subtracted from the number of editions, the average for the public domain set of books drops from 6.3 to 5.2. Overall, the number of editions of public domain works available do not suggest that these works are underexploited.

#### C. COMPARATIVE PRICE DATA

Interestingly, the average lowest list price per book, as gleaned from list price data on *Books in Print* online, was exactly the same (\$20) for both the 125 copyrighted bestsellers still in print in 2006 and the 162 public domain bestsellers in print in 2006. As we shall see in Part III.D, this pricing result does not hold for the smaller subset of the forty most enduringly popular public domain and copyrighted books that have held their value to the present day.

---

45. E-books are available only as a digital download direct to the consumer's computer. The buyer can then print them out or read them on screen. See, e.g., Net Library Home Page, <http://www.netlibrary.com> (last visited Mar. 10, 2008). While there are 180 public domain e-books listed in *Books In Print* online, only fourteen copyrighted editions were available for distribution in digital form. See *Booksinprint.com Professional*, *supra* note 34.

### III. COMPARING AVAILABILITY AND PRICE OF THE MOST ENDURINGLY POPULAR BOOKS FROM 1913–1932

Since the question of efficient exploitation is most salient in the context of the books that hold the most value over time,<sup>46</sup> this Part identifies the twenty most durable works from 1913–1922 and the twenty most durable works from 1923–1932 for closer scrutiny.<sup>47</sup> Part III.A notes that all forty titles are currently in print. Part III.B considers the number of editions in print over time.<sup>48</sup> Part III.C presents the comparative shelf-space data, and Part III.D presents the comparative price data.

#### A. PROBABILITY OF BEING IN PRINT

Not surprisingly, the comparative in-print status of the forty most durable books published from 1913–1932 shows little variation. All of the books in the group of twenty still protected by copyright (1923–1932) have been in print at all intervals measured from 60, 65, 70, 75, and 80 years after their publication dates, for a 100% in-print rate. Of the twenty books now in the public domain (1913–1923), all have been in print at all intervals measured since they fell into the public domain, 75, 80, and 85 years after publication date, for a 100% in-print rate during these periods. However, five books in this group had fallen out of print at various times while they were still covered by copyright.<sup>49</sup> This discrepancy may reflect the relative strength of the twenty books published from 1923–1932, an amazing list that includes *An American Tragedy*, *The Great Gatsby*, *The Sun Also Rises*, *Winnie-the-Pooh*, *A Farewell to Arms*, *All Quiet on the Western Front*, *Lady Chatterley's Lover*, *Brave New World*, and many other classics.<sup>50</sup> The 100% in-print rate for public domain works indicates that these works are not under-exploited.

46. See Landes & Posner, *supra* note 13, at 473–74; see also *supra* notes 23–25 and accompanying text.

47. See *supra* notes 37 and 39 (listing the twenty public domain titles (1913–1922) and the twenty copyrighted titles (1923–1932) in the study).

48. The formal regressions that demonstrate the statistical significance of all the findings in this Section are provided in Appendix A.

49. *Pollyanna* (1913) by Eleanor Porter was out of print in years 60, 65, and 70 after its publication. *O Pioneers!* (1913) by Willa Cather was not in print in year 70 after publication. *The Lone Star Ranger* (1915) by Zane Grey and *Scaramouche* (1921) by Rafael Sabatini were out of print in year 60 after publication. *Captain Blood* (1922) by Rafael Sabatini was out of print in years 60, 65, and 70 after publication.

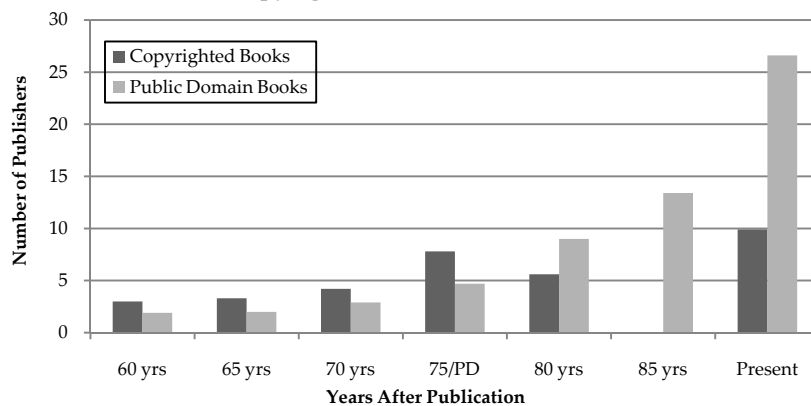
50. See *supra* notes 37 and 39, and app. B, available at <http://local.law.umn.edu/lawreview/issues.html>, for a full listing of both groups.

## B. NUMBER OF EDITIONS IN PRINT OVER TIME

In the absence of historical sales data, one possible measure of availability is variations in the number of editions of each “durable” book available over time. This is an imperfect metric, however, because copyright owners may be unwilling to license their books widely. Since public domain books can be published without permission, multiple editions of those works may tell us little about comparative availability in some venues, for example shelf space occupied in book stores. Data on number of editions do demonstrate, however, the widespread commercial exploitation of the public domain titles. Shelf space data presented in the next Section suggest that the most popular public domain titles are as available in book store venues as those still protected by copyright.

As shown in Figure 3, *infra*, during the time period when both sets of books were protected by copyright, the newer books published from 1923–1932 were somewhat more available in terms of numbers of editions. As with the overall comparison, this trend holds at the seventy-five-year mark when the older books fall into the public domain. Thereafter, the number of editions per public domain book trends sharply upwards, with comparative data for 2006 showing 29.1 editions for each durable book in the public domain and 8.9 editions of each copyrighted durable book. Although the lower number of editions for the copyrighted books does not necessarily signal a lower demand in absolute terms, it seems difficult to conclude from the data that the public domain works are underexploited due to their legal status.

**Figure 3. Number of Editions of Durable Public Domain and Copyrighted Works over Time**



## C. COMPARATIVE SHELF SPACE DATA FOR DURABLE BOOKS

Although the twenty public domain durable books (1913–1922) are widely available in the sense that they can be obtained from numerous different sources, one could argue that availability on bookstore shelves might be of additional economic interest. If publishers prefer to push their copyrighted titles directly to consumers in stores, then a positive informational externality might be generated. If occupying shelf space in front of consumers is the primary means of advertising books over seventy years old, then the fact that all forty durable books are readily available from Amazon.com may not tell the whole story about comparative status. Perhaps copyright generates marginally more information about old titles for consumers.

A list of all forty durable books was faxed to Borders and Barnes & Noble bookstores in Georgia, California, Illinois, and New York during the week before and after Thanksgiving, 2006. Data collected from eight responding stores indicated an availability rate of 80% for the public domain set of twenty books and an 86% availability rate for the set of twenty books still protected by copyright. Each of the eight stores could have at most twenty books in each set, so 100% availability would be indicated by a score of 160/160. The public domain books scored 128/160 (80%), and the copyrighted books scored 138/160 (86%). In total, due to multiple copies of many of the books being offered for sale, 512 copies of the public domain books were available on shelves, and 690 copies of the copyrighted books were available on shelves.

Given the comparative fame of the twenty durable books still protected by copyright (1923–1932), the number of additional copies of those books on shelves today seems surprisingly small, and may actually point to a positive public domain effect. Partial historical sales data on the forty durable books that can be collected from *Publisher's Weekly*<sup>51</sup> suggest that if popularity were the sole variable, then the copyrighted books should occupy significantly more shelf space than the public domain books. For example, as of 1965, when all of the forty durable books were still protected by copyright, only five of the twenty books (1913–1922) that have since fallen into the public domain had sold 1,000,000 copies.<sup>52</sup> As of the same date, eleven of the twen-

51. See HACKETT, *supra* note 27, at 12–55.

52. Those books in order of sales as of 1965 are W. SOMERSET MAUGHAM, *OF HUMAN BONDAGE* (1915), RAFAEL SABATINI, *SCARAMOUCHE* (1921), JAMES



ty books (1923–1932) still protected by copyright today had sold 1,000,000 copies, despite having on the average ten fewer years to accomplish that feat.<sup>53</sup> Even more tellingly, the top five books from the public domain set (1913–1922) had sold a total of only 7,381,709 volumes as of 1965, while the top five sellers from the copyrighted set (1923–1932) had sold 20,289,943 volumes.<sup>54</sup> And as of 1965, the top five books still protected by copyright had fifteen fewer years to sell than those that have since fallen into the public domain.<sup>55</sup> Sales data for books selling fewer than 1,000,000 copies as of 1965 is not publicly available. An update on books that had sold over 2,000,000 volumes by 1975 reemphasizes the comparative popularity of the books published from 1923–1932.<sup>56</sup> Only one of the durable books published from 1913–1922 is on the list (*Of Human Bondage*, with sales of 2,609,236), while seven from 1923–1932 are on the list. Sales of those seven books, as of 1975, totaled 28,732,714.

Although the twenty durable books in the copyrighted (1923–1932) data set are currently somewhat more available on store shelves than their public domain (1913–1922) counterparts, the difference (86% vs. 80% availability; 690 vs. 512 total volumes) seems difficult to attribute solely to their present legal status. Given that the twenty newer books in the copyrighted set have historically been much more popular, we would expect to see comparatively more of them on the shelves. Taking fame into account, the shelf-space data seems consistent with the implication of the data on comparative number of available editions—the public domain books do not seem to be underexploited.

---

JOYCE, *A PORTRAIT OF THE ARTIST AS A YOUNG MAN* (1916), ELEANOR PORTER, *POLLYANNA* (1913), and EDGAR BURROUGHS, *TARZAN OF THE APES* (1914).

53. Those books in order of sales as of 1965 are D.H. LAWRENCE, *LADY CHATTERLY'S LOVER* (1930); ERICH MARIA REMARQUE, *ALL QUIET ON THE WESTERN FRONT* (1929); ALDOUS HUXLEY, *BRAVE NEW WORLD* (1932); CHARLES NORDHOFF & JAMES NORMAN HALL, *MUTINY ON THE BOUNTY* (1932); PEARL S. BUCK, *THE GOOD EARTH* (1931); WILLIAM FAULKNER, *SANCTUARY* (1931); THORNTON WILDER, *THE BRIDGE OF SAN LUIS REY* (1927); SINCLAIR LEWIS, *ELMER GANTRY* (1927); ERNEST HEMINGWAY, *THE SUN ALSO RISES* (1926); DASHIELL HAMMETT, *THE MALTESE FALCON* (1930); and A.A. MILNE, *WINNIE-THE-POOH* (1926).

54. HACKETT, *supra* note 27, at 12–55.

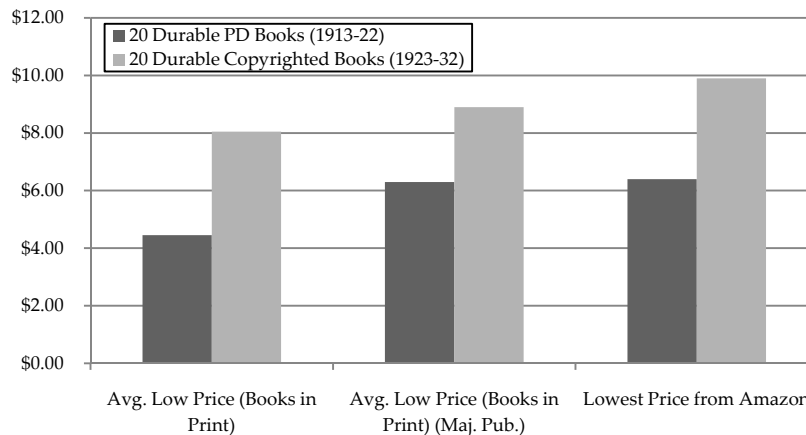
55. The average publication date for the five public domain books was 1916, while the average publication date for the five copyrighted books was 1931.

56. See ALICE PAYNE HACKETT & JAMES HENRY BURKE, *80 YEARS OF BEST SELLERS, 1895–1975* (1977).

## D. COMPARATIVE PRICE DATA FOR DURABLE BOOKS

Figure 4, *infra*, shows that one key difference between the public domain and copyrighted sets of durable books is price. The lowest average price listed by *Books In Print* is \$4.45 for a durable public domain book and \$8.05 for a durable copyrighted book, an 81% higher average price. If only prices of books sold by well-known major publishers are included, the average low price per book rises to \$6.30 and \$8.90 respectively, a 41% higher average price for the copyrighted books. And if one uses the lowest price each book can be obtained new from Amazon.com, then the respective prices rise to \$6.40 and \$9.90, a 55% higher average price for the copyrighted book.<sup>57</sup>

**Figure 4. Price Comparison of Durable PD and Copyrighted Books**



Two factors apart from copyright status might explain this price discrepancy. First, as noted above, the twenty durable copyrighted books (1923–1932) seem to be more popular<sup>58</sup> and may command a somewhat higher price on average. Second, the print quality of the public domain books for sale may be significantly lower. In order to control for these variables, an analysis of the Penguin Classics paperback collection was undertaken. Of the ninety Penguin Classics studied, forty-eight are currently protected by copyright law and forty-two are currently in the public domain.<sup>59</sup> The set of forty-eight copyrighted

57. To avoid polluting the sample with sales of books from secondary markets, only the price of editions available new from the Amazon.com warehouse itself were used. See Amazon.com, <http://www.amazon.com>.

58. See *supra* notes 33–36 and accompanying text.

59. The full list can be found in Appendix C, which is available at <http://>

Penguin Classics does not include any of the very famous books from the set of the twenty most durable copyrighted books studied above. In addition, since the Penguin books are the same quality and have the same source, any price difference found should be attributable to other factors, most likely their legal status (a publisher of a public domain book need not pay royalties to the copyright owner) or relative newness.

The average price per page of the forty-eight copyrighted Penguin Classics was \$0.047, and the average price per page of the public domain Penguin Classics was \$0.03. This represents a difference of slightly more than a penny and a half per page, or \$5.10 for a typical 300-page book. A 300-page copyrighted book in the Penguin Classic series would on average cost \$14.10, or 56% higher than its 300-page Penguin Classic public domain counterpart, which would cost on the average \$9.00. This is almost exactly the difference in the price for the twenty durable public domain and twenty durable copyrighted books found on Amazon.com (55%), suggesting that print quality cannot alone explain the price difference between the twenty durable public domain and copyrighted books. Given the finding in Part II.A.2, that positively correlates year of publication with availability, it may be that both relative newness and legal status explain the price difference.

#### E. OVERALL FINDINGS

When books published between 1913–1932 were written, the copyright term was twenty-eight years, plus a twenty-eight-year renewal term. This fifty-six-year maximum term was enough to stimulate the production of all of the books listed in the data set. In the absence of persistent extensions of copyright duration by Congress, all of the works studied would have fallen into the public domain no later than 1988. The present study makes two important findings relevant to evaluating the economic desirability of these extensions. First, copyright extension was not necessary to maintain the availability of the works studied. From 1988 to 2001, the public domain bestsellers were in print at approximately the same rate as copyrighted bestsellers. And after 2001, they are in print at a significantly higher rate. Second, as for the most enduringly popular titles, the copyrighted books are more expensive. For these works, the public pays a premium price and gets nothing in return in terms of increased availability. The price data suggest

that these copyright term extensions have been costly. Analyzing patterns of exploitation of both public domain and copyrighted books from several different perspectives suggests no offsetting social benefits in the form of increased availability attributable to copyrighted status.

#### IV. APPLICABILITY TO OTHER TYPES OF WORKS

Policymakers should be interested in whether the conclusions that can be drawn from the study of best-selling fiction are typical of most creative works. In particular, does extending the copyright term beyond the length of time necessary to stimulate creation have similar costs as applied to music, drama, paintings, movies, and software, or might underexploitation be a serious problem with some sorts of works?

The cost of reproduction and distribution should be a key factor in identifying public domain works that might be underexploited. As noted above, costly races to be the first to print public domain books in the nineteenth century threatened to reduce the availability of those books to the public.<sup>60</sup> The labor-intensive nature of manual typesetting made choosing to compete in the market for public domain books quite risky.<sup>61</sup> The study suggests that this is not presently the case. The ability to copy books via optical scanning technology, to store them digitally, and to print them with widely available software has lowered the cost of reproduction and distribution to such an extent that exploitation of public domain books has actually increased over that of copyrighted books from the same era. This suggests that other easily reproducible works, such as printed music, recorded music, movies, and software, might behave like books, with little chance of ruinous competition resulting in underexploitation when these works fall into the public domain. Extending protection beyond a term necessary to stimulate creation may be as costly as overprotecting books.

A study commissioned by the National Recording Preservation Board of the Library of Congress provides some support for this hypothesis.<sup>62</sup> The study notes that copyright term extension was motivated in part by the desire “to give owners an incentive to reissue, and thereby preserve, older recordings,”<sup>63</sup> and it then attempts to quantify whether extension has been

---

60. See Khan, *supra* note 17, at 21.

61. See *id.* at 21–24; see also *supra* note 19 and accompanying text.

62. See TIM BROOKS, SURVEY OF REISSUES OF U.S. RECORDINGS (2005).

63. *Id.* at v.

successful in achieving this goal. After an exhaustive canvassing of the availability of over 1500 recordings of historical interest, it concludes that “rights holders appear to have few real-world commercial incentives to reissue many of their most significant recordings.”<sup>64</sup> It finds that extension has not worked to increase the availability of older sound recordings. Only 14% of popular recordings<sup>65</sup> from 1890–1964 have been reissued and made available on CD by their owners.<sup>66</sup> At the same time, due to federal copyright law, virtually no recordings from the same era have fallen into the public domain. Therefore, conducting a direct comparison of the availability of public domain and protected recordings is extremely difficult.<sup>67</sup> Nonetheless, Brooks shows that despite the strictures of copyright law, nonowners have made 22% of the same set of recordings available on CD.<sup>68</sup> Just as with fiction bestsellers, nonowners are responsible for a high rate of availability to the public. Sound recordings, like books, are easy to digitize and distribute in a variety of forms.<sup>69</sup> The striking parallels in vigorous exploitation by nonowners in both the music and book contexts strengthens the hypothesis that lengthy copyright terms are not necessary to ensure the exploitation of easily duplicated works.

The need for extended protection, therefore, may only arise when the cost of making the initial copy available to the public is high, as was the case with books in the nineteenth century. In addition, one would expect barriers caused by high duplication costs might be exacerbated when the cost to free-riders of making subsequent copies is low. For example, if the cost of restoring an old public domain film is sufficiently high, then no one will have adequate motivation to restore it, because the cost of labor expended in the process would not be recouped if the restored work were easily copied and distributed by a sub-

---

64. *Id.* at 14.

65. The author limited the survey to recordings “in which there is a documented interest” as evidenced by their widespread use in source publications. *Id.* at 3.

66. *Id.* at 7.

67. Before 1972, sound recordings were protected by state law without time limitation. In 1976, protection for all recordings was extended to February 15, 2067. 17 U.S.C. § 301(c) (2000).

68. BROOKS, *supra* note 62, at 8. Almost all of these publishers are technically copyright infringers. According to the study, many are based in Europe. Some take advantage of the fact that 16% of owners cannot be easily identified and therefore may be presumed not to know that they have legal title to the works being reproduced. *Id.*

69. See, e.g., E-mail from Mark Schweizer, *supra* note 43.

sequent free-rider. In the context of damaged or fragile older films, the extension of copyright protection may therefore result in greater availability to the public. This may also be true with other sorts of artwork that require costly restoration before they can be made available.<sup>70</sup>

Extension of protection, however, may not always be needed even when restoration is costly and copying is cheap. Consider an expensively restored and easily copied work that incorporates new elements that are independently protectable. For example, if an updated version of a public domain film includes new scenes previously unseen by the public, then a free-rider cannot duplicate the film without violating copyright law, which protects the new contributions.<sup>71</sup> Similarly, “critical” editions of classic books are relatively costly to make available to the public, but if they contain new material not contained in the original edition, it becomes more difficult for free-riders to take advantage of them. For example, a new critical edition of *Sons and Lovers* by D.H. Lawrence that adds scenes cut by the author, a scholarly introduction, explanatory footnotes, and commentary cannot be scanned into digital form and legally distributed without stripping away the new edition’s original contributions.<sup>72</sup> Even if the free-rider were willing to perform the tedious excision, he would be left with an imperfect competitor for the new critical edition. In fact, Sutherland has suggested that one effect of works falling into the public domain is to stimulate new critical editions as a way to package an unprotected underlying work in protectable form.<sup>73</sup>

Therefore, the extension of copyright protection is most likely justified when three conditions are met: (1) the cost of making the initial copy of a work available to the public is high;

---

70. Another example might be the making of expensive derivative works, like movies, from books. One could imagine studios engaging in a costly race to make a movie based on a public domain book, when the market would really only support one film version. The possibility of a race might deter production. On the other hand, the community of movie producers is small and tightly knit, which might encourage a collusive solution similar to the one engineered by publishers of English fiction in the nineteenth century. Indeed, film trade magazines announce plans to make films far in advance of production dates, as if to stake claims to particular story lines. See, e.g., *Calendar of Future Releases*, FILM J. INT’L, Sept. 2007, at 63, 63–66.

71. See 17 U.S.C. § 103(a), (b).

72. See John Sutherland, *The Great Copyright Disaster*, LONDON REV. BOOKS, Jan. 12, 1995, at 3, 3 (describing the strategy of the D.H. Lawrence estate to authorize new critical editions of his works when confronted with their impending public domain status).

73. *Id.*

(2) the cost to free-riders of making subsequent copies is low; and (3) the newly available work does not incorporate independently protectable material. One can find a nice illustration of this hypothesis in the patent context, where Congress determined that some public domain drugs, potentially efficacious in treating diseases with small populations of sufferers, were not being adequately exploited. It responded by rewarding the first pharmaceutical company to satisfy the FDA approval requirements with a seven-year exclusive distribution right to the so-called orphan drug.<sup>74</sup> Congress was presented with a situation where (1) the cost of making the drug available to the public was very high due to stringent FDA testing requirements; (2) it would be relatively easy for generic drug manufacturers to imitate the good after approval; and (3) the party seeking FDA approval would usually have no other proprietary rights in the drug. Parallel protection in the context of copyrighted works, like fragile films needing restoration, may be justifiable to avoid similar problems of underexploitation.

### CONCLUSION

The data presented herein clearly suggest that the public domain status of popular books does not result in underexploitation. Although the public domain books in the data set are on the average ten years older than the copyrighted books, they are in print at a higher rate and have more editions available by more different publishers. If one considers only the subset of the most valuable books, then a significant difference in price can be measured, confirming economists' suspicions about deadweight losses associated with the extension of copyright protection.

In general, the data show a highly competitive and robust market for the production of public domain books. Markets for other products, such as movies, music, and software, where technology has made the cost of reproduction extremely low, are likely to behave in much the same way. Although market failure is theoretically possible, the burden should be on the party arguing in favor of central control of the production of a good in an apparently competitive market. Taking up Landes and Posner's call to explore analogies between intellectual property and tangible property,<sup>75</sup> one might conclude with a

---

74. 21 U.S.C. §§ 360aa–360ee (2000).

75. See WILLIAM M. LANDES & RICHARD A. POSNER, *THE ECONOMIC STRUCTURE OF INTELLECTUAL PROPERTY LAW* (2003).

simple point. If we trust the market to produce the optimal amount of tangible goods like string, bubble gum, and diet soda without entrusting central control of those products to a single authority, why should we treat intangible public goods like *My Ántonia*, the color yellow, or the word “coffee” any differently? Of course, we need a property right of sufficient duration to ensure that public goods get created in the first place, but extending the property right beyond that point demands affirmative proof that the market is incapable of responding efficiently to ongoing consumer demand for those creative products. Before extending copyright protection to existing works, a legislature should demand proof that without more protection their availability will be diminished.



## APPENDIX A: STATISTICAL ANALYSIS

(1) *Comparative availability of 166 public domain (1913–1922) bestsellers and 168 copyrighted (1923–1932) bestsellers.* If the status of every book in the database is considered at every measured point in time, a formal statistical regression confirms that the public domain books in the data set are not underexploited when compared to the copyrighted books in the data set. In fact, after 2001, the public domain books are exploited at a significantly higher rate.

Table 1. In-Print Frequency

[Frequency Row] [Percent] [Col Percent]	Copyrighted	Public Domain	Total
Out of Print	858 70.39% <b>57.97%</b>	361 29.61% <b>38.48%</b>	1219 50.41%
In Print	622 51.88% <b>42.03%</b>	577 48.12% <b>61.51%</b>	1199 49.59%
Total	1480 61.21%	938 38.79%	2418

The chi-square value for Table 1 is  $\chi^2 = 87.21$ . The p-value is  $P(\chi^2_{1} > 87.21 < 0.0001)$ . The p-value from the chi-square test indicates severe dependency between copyright status and availability of works. More specifically, it shows that works in the public domain were published at a significantly higher rate than those copyrighted. The above result is based on the assumption that all observations are independent from others and was used to determine if there exists an association to warrant further analyses. Since a strong dependency exists between copyright status and works' availability, we proceeded with further analysis. Of course, the results above are exaggerated to some extent because each book appeared, on average, about six times as a "book-event" in the above analysis, and the availability status for a particular book is surely positively correlated over time. However, even under the most severe assumption (that observations for a particular book are completely correlated, so that the sample size is exaggerated by a factor of 6), the  $\chi^2$  value obtained (87.21) would still provide a very strong evidence of a public domain effect.

Using copyright status (PD) alone to model availability (CPUB) might omit other significant factors affecting in-print status. Other variables that could be included in the model are publication year (PUBYR), number of years after publication

(AFPUB), and calendar year of measurement (YR). All four variables (PD, PUBYR, AFPUB, and YR) are possible explanatory variables for CPUB. Since copyright status is the explanatory variable of primary interest, it was the first variable included in the model. Care needs to be taken when choosing additional variables to include in the model to avoid confounding effects since some of these variables are functions of others. For example, copyright status (PD) depends solely on publication year (PUBYR) and age of the work (AFPUB), and the calendar year of the measurement (YR) is the sum of publication year (PUBYR) and age of the work (AFPUB). Since period is another effect of interest and YR was not too highly correlated with PD, it was included in the model. Including either PUBYR or AFPUB in this model (along with CPUB and YR) will cause some confounding, so we did not attempt this.

A preliminary model explaining CPUB by YR shows a somewhat steady increase in books' availability prior to the year 2001, with a large jump in proportion of works in print in 2001. To include this effect in the model, a new dichotomous variable called PY2K (1 if YR > 2000 and 0 if YR ≤ 2000) was created. The regression model specified is:

$$\ln\left(\frac{P_{ij}(\text{in print})}{P_{ij}(\text{out of print})}\right) = \beta_0 + \beta_1 * PD_i + \beta_2 * PY2K_j + \beta_3 * (PD_i * PY2K_j) + \beta_4 * (YR_j - 1966)$$

where  $i = 0, 1$  (0 for work copyright protected and 1 for work in the public domain) and

$$j = 0, 1, 2, \dots, 40 \text{ (} YR_0 = 1966, YR_1 = 1967, \dots, YR_{40} = 2006. \text{ } PY2K_0, \dots, PY2K_{34} = 0 \text{ and } PY2K_{35}, \dots, PY2K_{40} = 1 \text{)}.$$

[ $YR_j - 1966$  was used instead of  $YR_j$  for scaling purposes; it does not affect the outcome.]

Moreover, since each book was repeatedly measured at 60, 65, . . . years after publication, measurements on the same book were correlated. Methods of analysis must take this fact into account. Because the correlation between observations from the same book showed that correlation decreases for measurement between time points further apart, auto-regressive correlation is an appropriate assumption for these data. So, logistic regression with and without repeated measures with autoregressive correlation were performed. The results for these models turned out to be approximately the same. Due to this, the fact

that the response was measured five years apart and it is dichotomous, the repeated measures effect is negligible. The final model did not include repeated measures effects and the results are shown below. The p-values indicate that all variables (PD, PY2K, PD\*PY2K, and YR) are significant in explaining CPUB. PD is significant only from 2001 onwards and the jump between 2000 and 2001 is significant only for books in the public domain. An alternative interpretation of the parameter estimates of this model is

$$\ln\left(\frac{P(\text{in print})}{P(\text{out of print})}\right) = -1.4663 + .0545 * (YR - 1966) + X$$

Where X depends on Y2K and PD, and is given by:

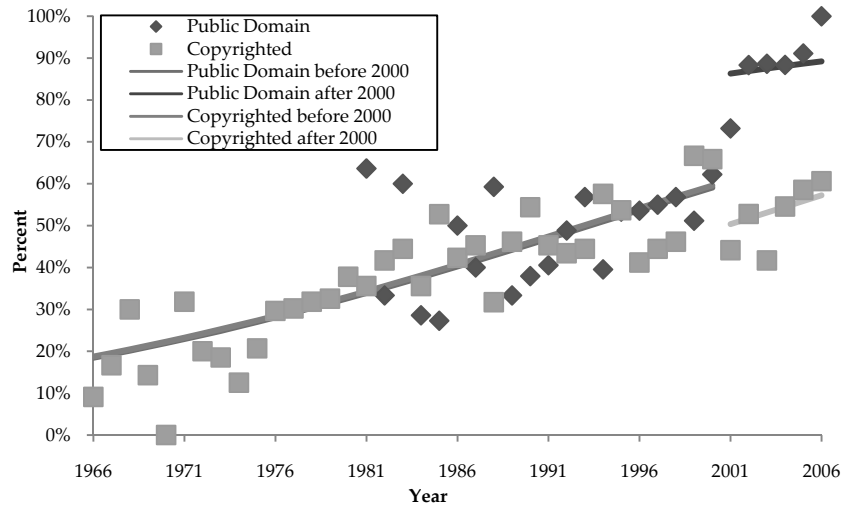
X	PD = 0	PD = 1
PY2K = 0	0.0000	-0.02
PY2K = 1	-0.46	+1.40

The least squares means availability (averaged over all relevant years) by PD\*PY2K are shown in Table 2, *infra*. Note that the average availability pre-2000 is about the same between groups, but after 2000, the public domain books are much more likely to be available.

**Table 2. Least Squares Means by PD\*PY2K**

	Copyrighted	Public Domain
2000 and earlier	0.47	0.47
After 2000	0.43	0.79

Below is a graphic depiction of the model. The dashed line represents the regression line for the public domain books and the solid line represents the copyrighted books. The effect of time (YR-1966) on availability is linear (+.0545/yr) in the logit scale used above, but will be slightly curvi-linear in the actual probability scale, as displayed in Figure 5, *infra*.

**Figure 5. Proportion of Popular Books in Print**

(2) *Number of Editions in Print in 2006.* The number of editions in print of public domain and copyrighted works in 2006 also varies significantly as explained below.

**Table 3. Frequency Distribution of Popular Books by Editions in 2006**

Editions Available	Public Domain	Copyrighted	Total
0	8	43	51
1	41	40	81
2	54	34	88
3	33	16	49
4	14	7	21
5–9	52	18	70
10–19	31	9	40
20+	15	1	16
<b>Total</b>	<b>248</b>	<b>168</b>	<b>416</b>

Figure 6. Editions of Popular Books in Print in 2006

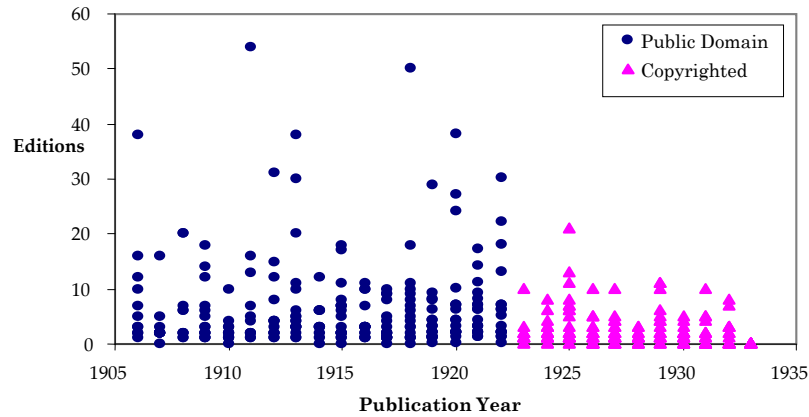


Table 3 and Figure 6 above compare public domain vs. copyrighted books, with both demonstrating generally greater availability for the former group. If one public domain book and one copyrighted book were randomly chosen from this data set, results from a simulation study of 10,000 repetitions indicate that the probability of the public domain work having more editions in print in 2006 is 0.781. This probability is consistent with both Table 3 and Figure 6.

The Poisson regression model specified was

$$\ln(EDT_{ij}) = \beta_0 + \beta_1 * PD_i + \beta_2 * PY2K_j + \beta_3 * (PD_i * PY2K_j + \beta_4 * (YR_j - 1973))$$

[Refer to Part II.A.2 for explanations of the indices.]

The p-values for all parameters are very small indicating that all variables are significant in explaining the number of editions in print for the durable books, although one must be cautious in interpreting these results when interaction terms are present, as they are here.

From the equation below, one can see that as time ( $YR - 1973$ ) goes on, the expected increase (in  $\ln$  scale) is 0.0632 per year, at which point there is a drop ( $-0.1183$ ) in overall editions. The intercept parameter estimate means, in 1973, the durable copyrighted works have  $\exp(-0.2479) = 1.28$  editions in print. Again, since there are no public domain works in our dataset until 1988, it makes no sense to estimate the editions of

durable books in the public domain in 1973. However, it does make sense to explain the parameter estimates in terms of differences between durable books in the public domain and durable books under copyright protection. Before 2000, the number of editions of books in the public domain in print is expected to be  $\exp(0.4972) = 1.69$  times higher than those of copyrighted works. However, from 2001 and on, editions of works in the public domain are expected to be  $\exp(0.5695+0.4972) = 2.90$  times higher than those of copyrighted works. An alternative interpretation of the parameter estimates of this model is

$$\ln(EDT) = 0.2473 + .0682*(YR - 1973) + X$$

where  $X$  (in  $\ln$  scale) is:

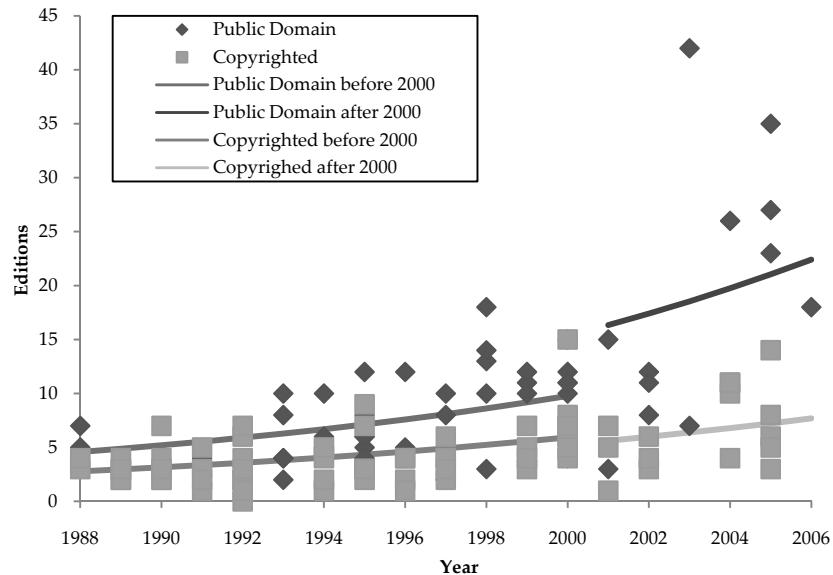
$X$	PD = 0	PD = 1
PY2K = 0	0.0000	+0.4972
PY2K = 1	-0.1183	+0.9484

The least squares means for editions in print by PD\*PY2K are shown in Table 4. These are merely values such as those calculated above when averaged over two representative years, one prior to 2000 and one after.

**Table 4. Least Square Means by PD\*PY2K**

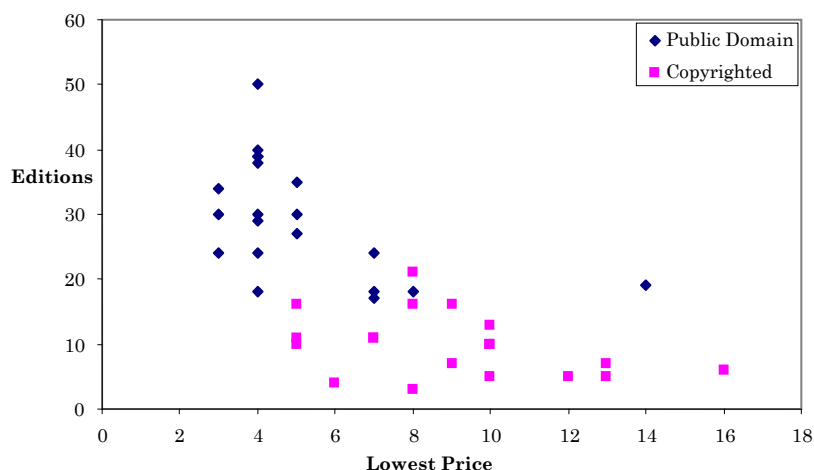
Year	Copyrighted	Public Domain
1986	3.58	4.02
2004	6.78	19.71

Below is a graphic depiction of the model. The curves represent the regression function for public domain books and copyrighted books. There is a small, but significant, difference between these groups even before 2000, but from 2001 onwards, the difference is huge.

**Figure 7. Editions of Durable Books in Print**

Given the dramatic difference in number of editions at the present time, we conducted a further analysis of availability in 2006 and examined its relation to price data, which was available for all forty of the durable books. The significance of the difference in numbers of editions is confirmed by a simulation study of 10,000 repetitions where one public domain book and one copyrighted book were randomly picked from the data set. From these simulation results, the probability of the public domain work having more editions published is 0.9861. This is much higher than the probability of 0.781 obtained previously by simulation for all of the bestsellers in the complete data set. Furthermore, exploring EDT vs. LOW graph in Figure 8, *infra*, also shows a clearer distinction between low prices for works in the public domain versus works under copyright protection. This again agrees with the simulation results that the probability of a randomly chosen public domain book having lower price than a copyrighted book is 0.937.

Figure 8. Editions in Print in 2006 Versus LOW



Analysis of editions of durable books available in 2006 based on copyright status is presented below. The Poisson regression model explaining  $E06$  by PD follows:

$$\ln(E06_i) = 2.2875 + 0.9934 + PD_i$$

As before, copyright status is an important factor, with public domain books having significantly more editions. Based on this model, the average editions available for copyrighted books is estimated to be  $\exp(2.2875) = 9.85$ , while for books in the public domain, it is  $\exp(3.2809) = 26.6$ .

The only explanatory variable considered in the model above is copyright status. If we further include the lowest price as an additional explanatory factor, the model becomes

$$\ln(E06_{ik}) = 2.9086 + 0.7547 * PD_i - 0.0710 * LOW$$

The results from this model indicate significant effects in the expected directions of copyright status and price. The coefficient for LOW is  $-0.0710$ , which means, controlling for other factors, that each dollar increase in price corresponds to a decrease of  $\exp(0.0710) = 0.93$  times editions available. The least squares values estimate the mean number of editions of books in print at the average price (\$7). These are  $\exp(2.4101) = 11.13$  editions for copyrighted books and  $\exp(3.6633) = 23.68$  editions for public domain books. If comparisons were made at a higher



2008]                      *EFFICIENT EXPLOITATION*                      1063

price, such as \$20, the number of expected editions for both groups would decline, but the relative pattern would remain the same (four and ten editions, respectively).