

HOW COPYRIGHT KEEPS WORKS DISAPPEARED

Paul J. Heald*

A random sample of new books for sale on Amazon.com shows more books for sale from the 1880's than the 1980's. Why? This paper presents new data on how copyright stifles the reappearance of works. First, a random sample of more than 2000 new books for sale on Amazon.com is analyzed along with a random sample of almost 2000 songs available on new DVD's. Copyright status correlates highly with absence from the Amazon shelf. Together with publishing business models, copyright law seems to deter distribution and diminish access. Further analysis of eBook markets, used books on Abebooks.com, and the Chicago Public library collection suggests that no alternative marketplace for out-of-print books has yet developed. Data from iTunes and YouTube, however, tell a different story for older hit songs. The much wider availability of old music in digital form may be explained by the differing holdings in two important cases Boosey & Hawkes v. Disney (music) and Random House v. Rosetta Stone (books).

One justification for granting authors a property right in their creations is the assumption that copyright stimulates the production of new works.¹ An alternative justification of growing importance claims that after a work is created, it needs to be protected for a significant period of

*Herbert Smith Fellow & Affiliated Lecturer, Cambridge University; Professor of Law & Raymond Guy James Faculty Scholar, University of Illinois; Professorial Fellow, Bournemouth University (UK), heald@illinois.edu. Thanks to my fabulous research assistants: Anne Lewis, Frank Madden, Laura Meli, Marissa Meli, Carolina Van Moursel, Carlos Ruiz, Jarrett Szczesny, Marc Tan, Dacheng Xie, Xiaoren Xie, Xi Zhao, and for comments provided by Jill Crandall, Greg Mandel, Jason Mazzone, Joe Miller, Arden Rowell, Chris Sprigman, Christian Turner, Tom Ulen, and participants at the 2013 Intellectual Property Scholars Conference, the 2013 Conference on Empirical Legal Studies, and the 2013 conference of the Society for Economic Research in Copyright Issues, especially Clive Bruton. The statistical analysis was performed by PeiBei Shi of the University of Illinois Statistical Consulting Office.

¹ See *Sony Corp. of America v. Universal 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.'").

time to assure its continued availability and distribution.² In the words of one commentator, a work may need “proper husbandry” in order to assure its continued exploitation.³ Influential copyright lobbyists presently circle the globe advocating ever longer terms of copyright protection based on this under-exploitation hypothesis—that bad things happen when a copyright expires, the work loses its owner, and it falls into the public domain.⁴ By analyzing present distribution patterns of books and music, this Article tests the assumption that works will be under-exploited unless they are owned and therefore questions the validity of arguments in favor of copyright term extension.

So far, several studies have tested the assumption that works need owners to be adequately exploited.⁵ Those studies relied on lists of bestselling books and songs from 1913 to 1932 and charted patterns of use and availability both before and after those works fell into the public domain.⁶ The research, summarized in Part I, casts doubt on the wisdom of extending copyright terms in existing works. The new data presented in this article address the same question but from a very different perspective. Rather than starting with a pre-established list of

² See *Eldred v. Ashcroft*, 537 U.S. 186, 207 (2003) (concluding that Congress “rationally credited projections that longer terms would encourage copyright holders to invest in the restoration and public distribution of their works”); *Mills Music, Inc. v. Snyder*, 469 U.S. 153, 187 (1985) (“[The] fundamental objective of the copyright laws requires providing incentives both to the creation of works of art and to their dissemination.”); H.R. REP. NO. 105-452, at 4 (1998) (“[T]he 1998 extension would ‘provide copyright owners generally with the incentive to restore older works and further disseminate them to the public.’”); Miriam Bitton, *Modernizing Copyright Law*, 20 TEX. INTELL. PROP. LJ. 65, 77 (2011) (“If [works enter] the public domain, they [become] obscure and thus no one [will] invest in them due to the problem of free riding. Items which retain enough value for future use should be given indefinite copyrights to maintain their value.”); William M. Landes & Richard A. Posner, *Indefinitely Renewable Copyright*, 70 U. CHI. L. REV. 471, 475 (2003) (“an absence of copyright protection for intangible works may lead to inefficiencies because of impaired incentives to invest in maintaining and exploiting these works.”).

³ See Dennis S. Karjala, *Harry Potter, Tanya Grotter, and the Copyright Derivative Work*, 38 ARIZ. ST. L.J. 17, 37 (2006). It should be noted that Karjala is an opponent of copyright term extension.

⁴ For a summary of extensive international lobbying efforts, see Christopher Buccafusco & Paul J. Heald, *Do Bad Things Happen When Works Enter the Public Domain?: Empirical Tests of Copyright Term Extension*, 28 BERKELEY TECH. L.J. 1, 10-12 (2013).

⁵ See *infra* notes 16-19 and accompanying text.

⁶ See *id.*

older famous works, the present research collects data from a random selection of new editions for sale on *www.amazon.com* (“Amazon”) and music found on new movie DVDs for sale on Amazon.⁷ Research examining what is for sale “on the shelf” reveals a striking finding that directly contradicts the under-exploitation theory of copyright: Copyright correlates significantly with the disappearance of works rather than with their availability. Shortly after works are created and propertized, they tend to disappear from public view only to reappear in significantly increased numbers when they fall into the public domain and lose their owners.⁸ For example, more than twice as many new books originally published in the 1890’s are for sale by Amazon than books from the 1950’s, despite the fact that many fewer books were published in the 1890’s.⁹

Part I briefly summarizes the hypothesis to be tested—that copyright is necessary to assure the adequate exploitation of creative works—and reviews the existing empirical literature. Part II sets forth the methodology of several new studies that examine the mix of public domain and copyrighted books and music presently available. Part III presents the data and compares the disproportionate number of new Amazon books initially published before the public domain cut-off date of 1923 with those initially published after 1923 (“book study”). The study of songs available on new DVD’s sold by Amazon (“song study”) shows less dramatic, but still significant, differences in the availability of music initially published before and after 1923.

After establishing the correlation between copyright status and the diminished availability of books and music-in-movies, Part IV considers consumer demand for out-of-print works and alternative markets that might satisfy that demand. After all, if no one wants the “missing” books on Amazon, or if plentiful substitutes for new hardcover books exist, then diminished

⁷ See *infra* notes 21-49 and accompanying text.

⁸ See *infra* notes 52-56 and accompanying text.

⁹ See *id.*

availability seems less problematic. Surprisingly, eBooks do not provide a significant alternative marketplace for out-of-print books. For example, only 36% of 162 bestselling books from 1923-32 currently had eBook editions in 2014, and none of the eBooks in that data set represented an out-of-print bestseller. In addition, a different data set of randomly chosen books reviewed in the *New York Times Book Review (NYTBR)* from 1930-2010 confirms the conclusion that the eBook market for older copyrighted books is shockingly anemic. Further analysis of the same set of *NYTBR*-reviewed books demonstrates that the market for used books on Amazon and Abebooks.com only partially fills the gap. The story with digital music is much more hopeful. A data set of bestselling tunes from 1913-32 shows a high degree of availability on iTunes, and an analysis of bestselling songs from 1919-26 similarly reveals that copyright does not seem to be an impediment to the preservation of old music on YouTube.com.

The article concludes that present efforts by copyright owners to extend the term of protection for copyright are unsupported by the empirical evidence and contrary to the public interest.

I. THE STORY THUS FAR

Copyright owners are in the business of collecting royalties on existing works, so they advocate extending copyright terms in order to perpetuate revenue streams.¹⁰ Once a work has been published, however, lobbyists lose the ability to make pro-extension arguments based on incentive-to-create rationales because the work already exists.¹¹ Instead, they argue—without empirical support—that bad things will happen to the work when it falls into the public

¹⁰ Lobbying efforts by copyright owners are detailed in Buccafusco & Heald, *supra* note 4 at 10-12.

¹¹ *Id.* at 3-4.

domain.¹² The public interest, so the story goes, requires term extension to prevent a public domain calamity. The history and effectiveness of this argument has been chronicled at length elsewhere,¹³ but one persistent assertion bears repeating: Creative works need owners who will assure their availability and adequate distribution.¹⁴ Although Congress in 1998 relied on this argument in extending the term of protection in the U.S. by 20 years,¹⁵ empirical studies have thus far failed to support this key assertion made by copyright lobbyists.

In fact, Heald (2008) studied bestselling novels from 1913 to 1932 and found that public domain status significantly increased the chance that a book would be in print and increased the number of publishers of it.¹⁶ In the sub-market for audiobooks created from the same set of 1913 to 1932 bestsellers, Buccafusco and Heald (2013) showed that a significantly higher number of the public domain books had audio versions for sale on *www.audible.com*.¹⁷ Although music data is harder to gather, Tim Brooks (2006) showed that non-owners of popular songs from 1890-1965 had converted a significantly higher percentage of them into digital formats than had the songs owners.¹⁸ Finally, Heald (2009) studied a set of popular songs from 1913 to 1932 and

¹² See, for example, *Copyright Term, Film Labeling, and Film Preservation Legislation: Hearing on H.R. 989, H.R. 1248, and H.R. 1734 Before the Subcomm. on Courts and Intellectual Property of the H. Comm. On the Judiciary*, 104th Cong. 217–18 (1995) (statement of Bruce Lehman, Assistant Secretary of Commerce and Commissioner of Patents and Trademarks) (“One reason quality copies of public domain works are not widely available may be because publishers will not publish a work that is in the public domain for fear that they will not be able to recoup their investment or earn enough profit.”). See also *infra* note 36. For a summary of arguments, see Buccafusco & Heald, *supra* note 4 at 13-17.

¹³ See Buccafusco & Heald, *supra* note 4.

¹⁴ See *supra* note 12.

¹⁵ See H.R. REP. NO. 105-452, at 4 (1998) (finding the 1998 extension would “provide copyright owners generally with the incentive to restore older works and further disseminate them to the public”).

¹⁶ See Paul J. Heald, *Property Rights and the Efficient Exploitation of Copyrighted Works: An Empirical Analysis of Copyrighted and Public Domain Fiction Bestsellers*, 92 MINN. L. REV. 1031 (2008) (studying 334 books and finding that after 2001 significantly more of the public domain books were in print and by significantly more publishers).

¹⁷ See Buccafusco & Heald, *supra* note 4 at 22-23 (studying 334 bestsellers from 1913 to 1932 and identifying available professionally recorded audio versions of each book).

¹⁸ See TIM BROOKS, NAT’L RECORDING PRES. BD., LIBRARY OF CONG., SURVEY OF REISSUES OF U.S. RECORDINGS 7–8 & 7 tbl. 4 (2005) (demonstrating that copyright owners had

showed that the public domain songs were no less likely to be in a movie than the copyrighted songs.¹⁹

The dates 1913 to 1932 are important to the studies summarized above because the subset published from 1913 to 1922 fell into the public domain from 1988 to 1998 (they had a 75-year copyright term), while properly renewed works from 1923 to 1932 are still protected by copyright (they have a 95-year term).²⁰ Studying books and music within a decade of the 1923 divide enables researchers to learn what happened to works from 1913 to 1922 after they fell into the public domain and then compare their behavior with copyrighted works from approximately the same era. As useful as such comparisons are, they do not tell policymakers what mix of public domain books and movies is currently “on the shelf.” Published studies so far have only looked at a specific set of older works and tracked them through time. Critically, availability can also be measured by looking at the age and legal status of works presently for sale to the public. If public domain works, for example, are underrepresented in the world’s largest on-line marketplace, Amazon, then copyright owners may have a valid point about under-exploitation.

made only an average of 14% of popular recordings from 1890 to 1964 available on CD’s, while non-owners had made 22% of them available to the public on CD’s).

¹⁹ See Paul J. Heald, *Bestselling Musical Compositions (1913-32) and Their Use in Cinema (1968-2008)*, 6 REV. ECON. RES. ON COPYRIGHT ISSUES 31 (2009) (studying 1294 popular songs from 1913 to 1932 as they appeared in films released from 1968-2008).

²⁰ Calculating the copyright term is tedious, and explanation of changes in term length will only be offered when necessary to the analysis of the studies. The first copyright statute, 1790 Act, provided authors with a fourteen-year term of protection that could be renewed for an additional fourteen years. In 1831, Congress extended the initial term of protection to twenty-eight years with a fourteen-year renewal term, and the 1909 Copyright Act extended the renewal term to twenty-eight years. The last major revision of the copyright statute, the 1976 Act, further lengthened the period of copyright protection. For existing works that had not yet entered the public domain, the 1976 Act added forty-seven years of protection to the twenty-eight-year term resulting in a total of seventy-five years of protection. The 1976 Act, which went into effect in 1978, did not retroactively revive copyright protection for works that had already entered the public domain; consequentially, all works published prior to 1923 remain in the public domain. The 1998 Sonny Bono Copyright Term Extension Act (“CTEA”) added an additional twenty years of protection to the copyright term for all existing works. Works created between 1923 and 1978 now receive ninety-five years of protection, while works created since 1978 would be protected for the duration of the lives of their authors plus seventy years, with anonymous works, pseudonymous works, and works made for hire receiving a defined ninety-five-year term of protection.

The two studies discussed below offer a new take on availability by observing books and music presently available to consumers when they shop.

II. METHODOLOGY: SAMPLING THE METAPHORICAL STORE SHELF

Given that in 2014 Amazon currently offers almost 9 million new hardback and 24 million paperback editions for sale in a number of different fiction and non-fiction categories,²¹ the book study used a random sampling technique designed to collect information on representative new fiction books. In order to sample fiction randomly, my research assistant wrote a computer program to generate random 10-digit ISBN numbers that were then submitted as search requests to Amazon using its publicly available application programming interface (API).²² We initially considered submitting requests querying only Amazon's "Literature and Fiction" browse node²³ but saw that it included "Essays and Correspondence" and "History and Criticism" as sub-categories. In an attempt to collect only fiction titles, we submitted to a number of what appeared to be purely fiction sub-categories within "Literature and Fiction," and excluded essays, correspondence, history, and criticism.²⁴ Only data on new books for sale by Amazon (no used books; no books for sale by Amazon "affiliates") were collected.

²¹ See *Books*, AMAZON, http://www.amazon.com/books-used-books-textbooks/b/ref=sa_menu_bo?ie=UTF8&node=283155 (last visited Feb. 3, 2014).

²² See *Application Programming Interface*, WIKIPEDIA, http://en.wikipedia.org/wiki/Application_programming_interface (last visited March 1, 2014) ("Generally speaking, an application programming interface (API) specifies how some software components should interact with each other. In practice in most of the cases an API is a library that usually includes specification for routines, data structures, object classes, and variables.").

²³ Search categories within Amazon are called "browse nodes." For a list of all possible search categories, see *Literature and Fiction*, FINDBROWSENODES, <http://www.findbrowsenodes.com/us/Books/17> (last visited March 1, 2014).

²⁴ The browse nodes chosen were: 10016 – British; 4465 – Comic Literature; 10129 – Contemporary Literature; 2159 – Drama; 16260301 – Foreign Language Fiction; 23 – Romance; 10132 – Literary Books; 10248 – Poetry; 9822 – United States; 542654 – Women's Fiction; 10311 – World Literature; 18 – Mystery & Thrillers; 16190 – Fantasy; 16272 – Science Fiction.

In the group of categories searched, only about one percent of the random ISBN numbers actually corresponded to a new edition of a book for sale by Amazon. Since Amazon allows no more than 2000 requests per hour, it took several weeks of continuous searching to generate a random list of 7000 new editions for sale. Surprisingly, many of the 7000 editions retrieved were not works of fiction. About one-third were works of literary criticism and biography, history, and theology, exactly the sort of works sought be excluded by our choice of browse nodes.²⁵ Another third were works of fiction, and a third were works with foreign language titles in a variety of different categories. The number of foreign language titles was especially notable because that sub-set seemed to be biased toward older works.²⁶

Since 150 of the titles identified were duplicates, the next step was to identify the initial publication date of as many of the 6850 remaining titles²⁷ as possible. Copyright Office records before 1978 are not digitized,²⁸ and using hard copy registration data at the Copyright Office to determine initial publication date was not feasible because records there are indexed chronologically and only by author or title within a particular known year.²⁹ In any case, registration data itself could only be a proxy for date of initial publication because a work can be

²⁵ It may be that Amazon does not do a particularly good job of categorizing its own works or rely on self-reporting by publishers, or it may include some non-fiction in the category “10132 – Literary Books.” *See id.*

²⁶ About one-half of the total works retrieved were accompanied by a date in parentheses as part of the title of the work. All dates were 1922 or earlier, suggesting that Amazon tracks books it believes to be in the public domain. Works with foreign language titles had a disproportionate number of pre-1922 parenthetical dates.

²⁷ About 150 of the 7000 editions retrieved were duplicate titles. We attempted to identify approximately 6850 different titles.

²⁸ *See* U.S. COPYRIGHT OFFICE, <http://www.copyright.gov/records/> (last visited March 1, 2014); U.S. COPYRIGHT OFFICE, CIRCULAR 23 (2012), <http://www.copyright.gov/circs/circ23.pdf>.

²⁹ Pre-1978 Copyright Office records are organized by year, not by author or title, so finding a year of registration with only title and author information requires a painstaking search of every year on file. One professional search service, Thomson, charges \$750 per work for searching through physical copyright registration records in order to determine the initial registration date and renewal of a single work. *See U.S. Copyright Search*, THOMPSON COMPUMARK, <http://trademarks.thomsonreuters.com/searching/title-copyright-entertainment-searches?id=node/230> (last visited March 1, 2014) (One must call the phone number to confirm the price).

initially published before or after registration.³⁰ Instead, my research assistant wrote a program to search U.S. Library of Congress (LOC) records for the earliest edition of each of the Amazon titles held in its collection. The earliest edition in the LOC is a decent proxy for initial publication date as U.S. copyright law provided, and still provides, incentives to deposit a copy of the first published edition with the library.³¹ Deposit is still a routine business practice with major publishers.³²

Nonetheless, not every publisher deposits a book with the LOC, and not every book there is represented by a first edition. A book initially published in 1920, for example, may only be represented in the LOC by a later edition from 1935. For this reason, it is likely that the dates we collect from LOC editions are biased upward. A copy deposited in the LOC may often be a second or third or fourth edition; it should seldom be a copy deposited years before it was published.³³ Some of the upward dating bias may be ameliorated by changes weakening the deposit requirements in the 1976 Copyright Act,³⁴ but even under its predecessor Acts of 1831 and 1909, a failure to make an initial deposit did not result in the forfeiture of copyright, but rather the possibility of sanction if an author ignored an LOC request for a copy.³⁵ Penalties for

³⁰ For example, the registration date on my first novel is 1998, yet it will not be published until 2014.

³¹ See 17 U.S.C. §§ 411-412 (2006) (requiring registration and deposit as a condition of bringing suit, collecting attorney's fees, or collecting statutory damages). See also 2-7 MELVILLE B. NIMMER & DAVID NIMMER, NIMMER ON COPYRIGHT §7.16(B)(6)(a) (2010) (explaining changes in the deposit requirement over time); Committee, *The Library of Congress Advisory Committee on Copyright Registration and Deposit*, 17 COL.-VA J. OF LAW & ARTS 271, 288 (1993).

³² See Email from Mark Schweizer, Owner, St. James Music Press, to Paul J. Heald (July 18, 2013) (on file with author).

³³ But see *supra* note 30.

³⁴ See *supra* note 31.

³⁵ *Id.*; see David Rabinowitz, *Everything You Wanted to Know About Pre-1909 Copyright but Were Too Lazy to Look Up*, 49 J. COPYRIGHT SOC'Y USA 649,655 (2001) (noting that as of 1865 no deposit needed to be made until a request by the LOC with penalties assessed for failure to comply with the request).

failure to deposit were more serious under prior acts,³⁶ which may help to partially correct any dating bias for works initially published in the early 19th Century. There is little doubt, however, that an upward dating bias remains in the sample. This makes the results of the study discussed below even more striking.

Of the randomly selected new fiction editions for sale on Amazon, the software program located 2266 non-duplicate titles³⁷ in the LOC catalog. At least three factors prevented the discovery of all the titles. First, some authors, of course, never deposit a copy of their work.³⁸ Second, the data scraped from Amazon is derived from an edition it is selling, which is not necessarily the same edition as the deposit copy. Therefore, discrepancies between the form of an author's name (for example, the choice to include middle initials) in Amazon records and LOC records are likely. The LOC copy of a first edition of *The Lion, the Witch, and the Wardrobe* might list the author as "Clive Staples Lewis," whereas an edition published decades later and sold by Amazon might list the author as "C.S. Lewis." And even when Amazon is selling the same edition as the one found in the LOC, the Amazon digital record might diverge slightly from what is listed in the title page of the hard cover edition it is selling. Furthermore, LOC records tend to rely on the author's name as listed in the copyright registration document, and publishers may use a variant of that name. For example, the author of *The Hunt for Red October* might be "Tom Clancy" in one place and "Thomas M. Clancy" in another.

Most importantly, more than 2000 of the Amazon titles not found in the LOC were foreign language books. These foreign language works were probably never have deposited or

³⁶ See 35 Stat. 1078 (1909); Case Note, *Copyright—Failure to Deposit Copies Promptly Held Not to Bar Suit For Infringement Prior to Deposit*, 52 HARV. L. REV. 837 (1939).

³⁷ Since 51 were duplicates, 2266 unique titles were dated.

³⁸ For example, the copyright in my second novel, *No Regrets*, published in 2003, has never been registered.

registered.³⁹ Of the titles located in the LOC, only 6% were foreign language works. The data analysis in Part III addresses the 2266 works for which we have publication dates, so the earlier identified age bias within the foreign language sample should have a negligible effect on the findings. Of 2266 located titles, approximately 51% were estimated to be works of fiction (mostly novels, but some drama and poetry) and 43% were estimated to be works of non-fiction (primarily literary history and biography, theology, essays, history, and correspondence).

Collecting a valid random sample of music proved to be more challenging, in part because a song does not have an ISBN number or its equivalent. Taking a random sample of CD's on Amazon, as was done with books, is theoretically possible because Amazon assigns its own number (ASIN) to each CD that it sells, but the results taken from a random sample are difficult to interpret. Amazon data lists the performer of a song and its title, but not the composer, which makes it difficult to determine the original year the music was published (as opposed to the years in which a song was recorded). Songs are not indexed in the copyright office by performer, and a title search for pre-1976 song titles is not possible on the Copyright Office web site. In addition, a large percentage of song titles have multiple entries,⁴⁰ so knowing only the performer and not the composer is little help in determining which copyrighted version Amazon sells. Using the earliest Library of Congress entry for a particular title as a proxy for its publication date, as was done with books, is not possible because on-line access to music score

³⁹ There may be several reasons for this. Foreign authors may have a lower rate of deposit because most foreign jurisdictions do not require deposit. The Berne Convention, which the US only joined in 1989, requires its members to drop all formalities as a prerequisite to the grant of copyright protection, including the deposit requirement. Most countries around the world are longtime members of Berne and did away with deposit requirements long ago. Also, discrepancies in spelling between Amazon editions and LOC editions may proliferate when accent marks and long foreign words may not match perfectly as required by the software.

⁴⁰ See *Three Boys Music Corp. v. Michael Bolton*, 212 F.3d 477 (9th Cir. 2000) (noting that 129 songs registered in the Copyright Office share the title, *Love is a Wonderful Thing*).

holdings is rudimentary. For example, although Irving Berlin wrote over 1500 songs, only 90 are available via an LOC music score search.

Taking a random sample from iTunes is also problematic. One can search for individual song titles on iTunes, but Apple does not assign a unique public identifying alphanumeric for each song it sells, so generating a truly random sample to investigate is difficult. In addition, like Amazon, iTunes does not permit gathering of song data by composer or publication date, only by performer and title. Finally, even were a random sample easily generated, the same problems determining the original publication date for Amazon compositions plagues dating compositions of iTunes songs. Sampling music on YouTube was also considered, but pulling a random sample from YouTube is impossible because its search algorithm is not randomized, but rather based on the queries presented in prior searches.⁴¹

One source of music information, however, does identify the name of the composer, along with the song title. The movie database IMDB.com provides comprehensive information on almost every movie soundtrack and creates the opportunity for taking a random sample of films and then tracking down the initial publication date of the music contained therein. Choosing to sample music in movies has further advantages over attempting to sample music from iTunes or Amazon. Each song in a movie is approved by the director who has determined that it will enhance the value of the film. Since the core debate over term extension revolves around works that hold their value over time,⁴² approval by film directors provides an independent indication of the value of the music chosen. Also, musical compositions as they appear in movies are derivative works. The director must pay a band or orchestra to record the

⁴¹ See Greg Jarboe, *YouTube Algorithm Change: 'Time Watched' Key to Higher Video Search Rankings*, SEARCH ENGINE WATCH, <http://searchenginewatch.com/article/2218696/YouTube-Algorithm-Change-Time-Watched-Key-to-Higher-Video-Search-Rankings> (last visited March 1, 2014) (detailing changes to the YouTube algorithm to account for the amount of time a prior video was watched).

⁴² See Landes & Posner, *supra* note 2 at 475.

work or obtain a license to use an existing recording. Advocates for term extension make a special point of arguing that public domain works will not attract investors interested in making derivative works because they cannot exclude competitors.⁴³ Tracking music in movies permits evaluation of the claim that derivative works will be under-produced.

Two samples of music were collected. First, 134 movies were sampled randomly from www.boxofficemojo.com (BoxOfficeMojo).⁴⁴ Movies from this sample that were not for sale on Amazon were eliminated and replaced randomly with movies that were available in a new DVD version. The music in the 134 movies was identified using the soundtrack search function on www.imdb.com (IMDB)⁴⁵ and 1078 songs were identified. Next, the top 100 highest-grossing movies of all time (adjusted box office figures) were identified from a list on BoxOfficeMojo.⁴⁶ A number of those films either contained no songs or lacked soundtrack information, so a soundtrack search on IMDB generated a shorter list of 836 songs.

Determining the initial publication dates of almost 2000 songs was challenging and required several research assistants to consult several sources, including Google, Wikipedia, a

⁴³ Professor Arthur Miller worries that new works deriving from and based on materials in the public domain will be under-produced. Copyright law gives owners the exclusive right to make or license derivative works like adaptations, sequels, and translations that are based on the original work. Miller argues that derivative works like recordings of musical compositions, adaptations, sequels, and translations will not be made without copyright term extensions. See Symposium, *The Constitutionality of Copyright Term Extension: How Long is Too Long?*, 18 CARDOZO ARTS & ENT. L.J. 651, 693 (2000) (panel comments of Arthur Miller) (“[Miller reasons that] you have to provide incentives for [producers] to produce the derivatives, the motion picture, the TV series, the documentary, whatever it may be—perhaps even a musical! . . . We must incentivize the dissemination industries, the preservation industries, and the derivative work industries.”).

⁴⁴ See *Movies*, BOX OFFICE MOJO, <http://www.boxofficemojo.com/movies/> (last visited March 1, 2014). The website is organized by movie title from A-Z and within each letter group also divided alphabetically. For example, the letter A is sub-divided A-Ac, Ad-Af, Ag-Al, etc. The fifth movie listed in each of the 134 alphabetical sub-divisions was selected. If a movie was eliminated as not available for sale on Amazon, then the sixth movie was chosen and so forth.

⁴⁵ See *Advanced Search*, IMDB, <http://www.imdb.com/search/> (last visited March 1, 2014) (click drop down menu under “title search;” then click “soundtrack” option).

⁴⁶ See *All Time Box Office*, BOX OFFICE MOJO, <http://boxofficemojo.com/alltime/adjusted.htm> (last visited March 1, 2014) (listing top 100 movies in terms of box office gross during the first release of the film adjusted by ticket price inflation).

list of 3700 most popular songs from 1880-1965,⁴⁷ and scanned volumes of the Catalog of Copyright Entries.⁴⁸ Although in some circumstances, images of original sheet music or other authoritative sources could be examined, the publication date used for a song was often the year of its popularity (e.g., when it was a hit on the Billboard charts). Radio chart data or dates when sheet music sales peaked were often used as a proxy for date of publication. Since songs are not technically published when they are played on the radio, but rather when the underlying sheet music is sold, appearance on popularity charts is not an unfailing measure of publication date. However, since songs are published both before and after their sound recordings are popularized,⁴⁹ a systematic bias upward or downward may not be present. Most importantly, popular songs are usually published within several years of the release of the sound recording when the market for sheet music is hottest, another factor reducing bias. To further reduce any dating distortion, the data is presented by decade rather than year-by-year.

III. THE CASE OF THE DISAPPEARING WORKS

The academic literature tells two stories about what happens to works when they fall into the public domain. First, some economists like Landes and Posner suggest that “[a]n absence of copyright protection for intangible works may lead to inefficiencies because . . . of impaired

⁴⁷ Compiled from JULES MATTFIELD, *VARIETY MUSIC CAVALCADE* (1965) (compiling the most popular songs in American history by year).

⁴⁸ Although copyright registration records before 1978 are not available on-line at the copyright office website, The Internet Archive has scanned copies of many volumes available. *See Ebook and Texts Archive*, INTERNET ARCHIVE, <http://archive.org/details/copyrightrecords> (last visited March 1, 2014). Unfortunately, Boolean searching of OCR copies is not possible, so identifying registration records within them is extremely unwieldy, and the quality of the scanning renders them less than completely reliable. The records were therefore not the initial source consulted by my research assistants.

⁴⁹ Unlike books, which are published once copies are sold, a song can be exploited in a recording and technically remained unpublished. This creates the likelihood, not present with books, that a song will be popular in one year, but not technically published until a later year. For example, *White Christmas* by Irving Berlin was a hit in 1941, but the copyright was not registered until five years later.

incentives to invest in maintaining and exploiting these works.”⁵⁰ This is the under-exploitation hypothesis in a nutshell. Why sell a work when others can also exploit it for free and capture market share? Others have argued instead that when works fall into the public domain, they become attractive targets for exploitation because no license fee need be paid to the former owner of the work.⁵¹ They argue exploitation will occur, just as it does in other markets where no one has a monopoly over the object of exploitation (e.g., the markets for string, milk, and pencils). The data collected from Amazon demonstrates the power of the second hypothesis, that books and music become more attractive targets for exploitation after they fall into the public domain.

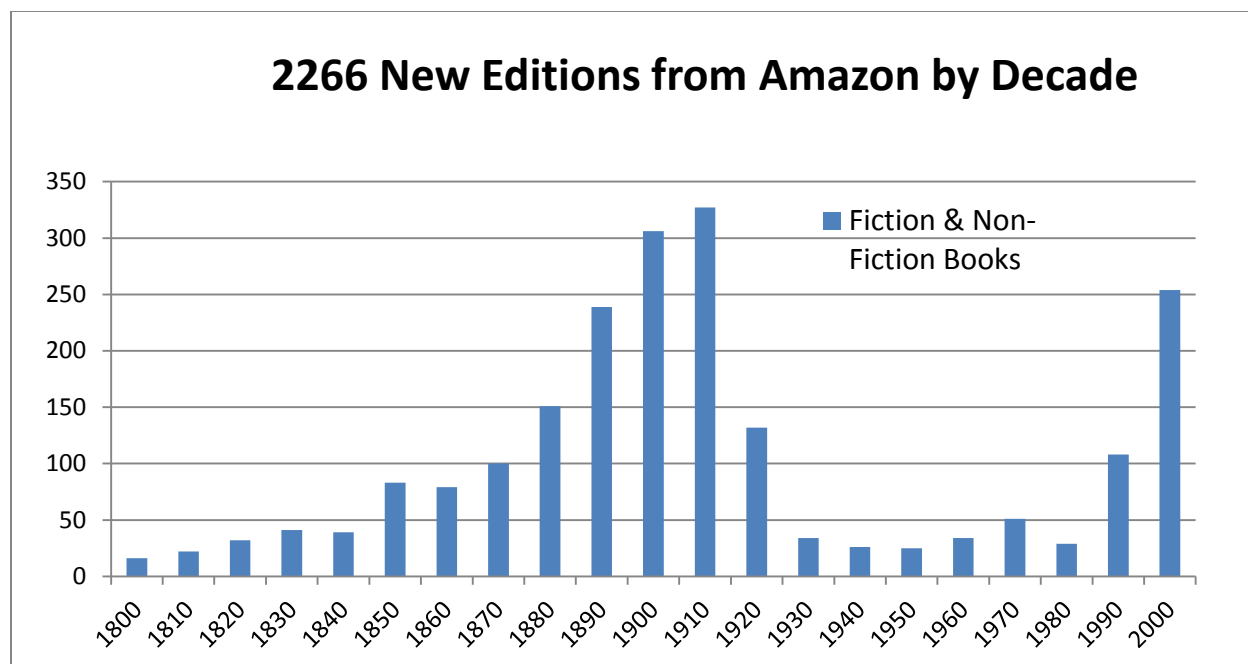
A. The Market for New Books on Amazon

The 2266 random editions of new books available on Amazon during the fall of 2012 are charted in Figure 1 below by the decade of the original publication date for the corresponding title. Both fiction and non-fiction editions are included. Editions of books now in the public domain (those published prior to 1923) constitute 72% of the total (1665/2266), while editions of titles still under copyright constitute 28% (652/2266).

Figure 1

⁵⁰ Landes & Posner, *supra* note 2, at 475.

⁵¹ See Buccafusco & Heald, *supra* note 4 at 18-19.



NOTE: Each edition was identified by query with a randomly chosen ISBN number. Since some book titles have multiple ISBN numbers, approximately 50 duplicate titles were excluded. Editions are ordered by decade based on the year of original publication of the underlying work. For example, a 2005 edition of Tom Sawyer is included in the decade of the 1870's, as its initial publication date was 1876.

The high percentage of public domain editions is probably driven by two factors. First, public domain books typically have more publishers and more editions per title, since there is no copyright owner to restrain exploitation.⁵² Second, Amazon offers as “in stock” multiple new editions of public domain books sold by a growing group of print-on-demand publishers that take advantage of the recent digitization of many old titles.⁵³

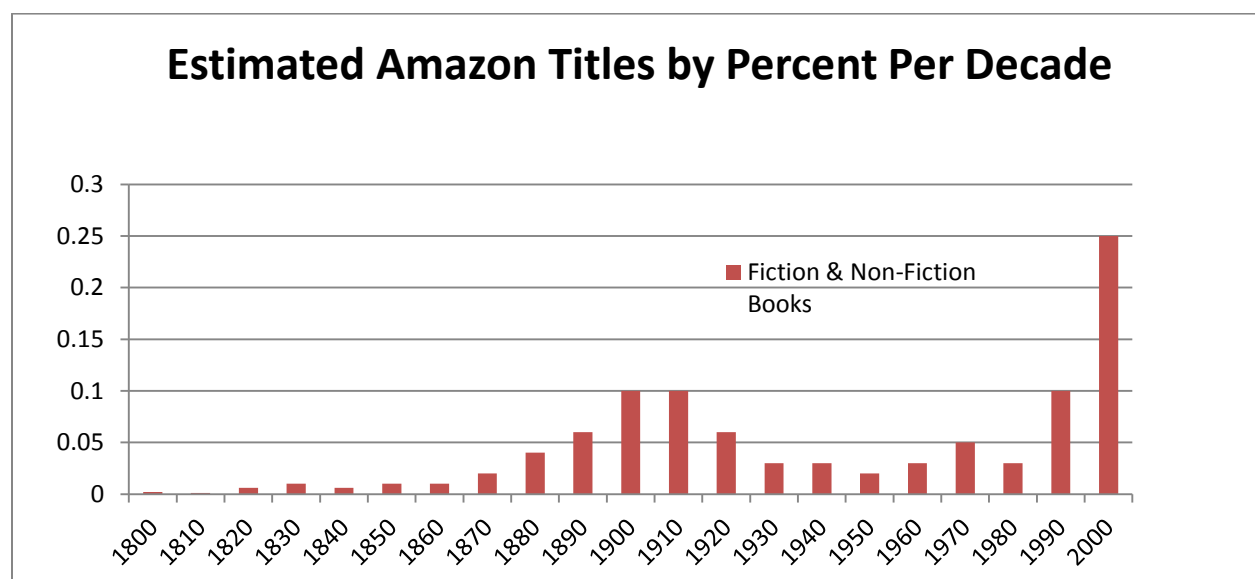
Demonstrating that buyers have the choice of multiple new *editions* of public domain books is not the same as showing that they can choose from more public domain book *titles*. In

⁵² See Heald, *supra* note 16 at 1044-45 (finding bestselling public domain books from 1913 to 1922 have more than twice as many editions as their copyrighted counterparts from 1923 to 1932).

⁵³ A quick search of Amazon, for example, reveals over 400,000 editions published by Kessinger Press, a prominent print-on-demand service. See KESSINGER PUBLISHING'S RARE REPRINTS, <http://www.kessinger.net/> (last visited March 1, 2014) (describing Kessinger Press). See also *BiblioBazaar*, WIKIPEDIA, <http://en.wikipedia.org/wiki/BiblioBazaar> (last visited March 1, 2014) (describing the business model of print-on-demand publishers Bibliolife and Nabu Press); WEBCITE, <http://www.webcitation.org/5v11RFFZZ> (last visited March 1, 2014).

order to estimate the number of public domain and copyrighted titles, each of the 2266 books was investigated on Amazon (minus 50 duplicates) and the number of editions per title was counted. Not surprisingly, the public domain books averaged 4 times more editions per title than the copyrighted books.⁵⁴ In Figure 2 below, the number of public domain editions is therefore divided by 4 in order to estimate the number of public domain titles available from each decade.

Figure 2



NOTE: Because Amazon only tracks the number of editions it sells (one per ISBN), it does not know the number of titles it offers. Books frequently are published in more than one edition, estimating the number of titles available on Amazon requires dividing the number of editions in the random sample by the average number of editions per title. The estimate is given in terms of the percentage of titles likely available on Amazon from each listed decade with adjustments between decades based on the ratio of editions to titles for books initially published within each decade.

⁵⁴ The 1665 public domain titles had a median of 4 editions per title and a mean of 16. The median number of editions for the copyrighted titles was 1 and the mean was approximately 1.6. Consultant statisticians recommended using the medians of 4:1 as a ratio for two reasons. First, the sample of public domain titles was characterized by massive outliers skewing the mean. A small number of titles with 300, 400, and even 700 editions drove the mean up to 16, while more than 1000 of the 1665 titles had 5 or fewer editions. Second, as noted in the text following Figure 2, the public domain sample of 1665 was likely skewed in favor of works with a higher number of editions. Even using a ratio based on the means, the findings are still significant, although less visually dramatic, when changes in the number of books published per decade is accounted for. See *supra* Figure 3.

We should note that applying a divisor of 4 to the public domain books in the graph above almost certainly under-estimates the number of public domain titles.⁵⁵ Strictly speaking, one should estimate the ratio of public domain editions to titles by using the average number of editions for the entire universe of public domain titles available on Amazon. For practical reasons,⁵⁶ the ratio of 4:1 was calculated from the sample of 2266 works retrieved from the random ISBN queries to Amazon. Those queries, however, were most likely to retrieve books with a higher number of editions and are surely skewed upward. After all, if one feeds a random ISBN to Amazon, one is more likely to retrieve Milton's *Paradise Lost* (with 401 editions and 401 ISBNs) than Lorimer's *A Wife out of Egypt* (1 edition and 1 ISBN). Nonetheless, the distribution of public domain to copyrighted titles is quite dramatic, even after dividing by 4, a likely inflated figure.

In a world without copyright, one would expect a fairly smoothly downward sloping curve from the decade 2000-2010 to the decade of 1800-1810 based on the assumption that works generally become less popular as they age (and therefore are less desirable to market). If age were the only factor, one would expect to see fewer titles available from each successively older decade. Instead, the curve declines sharply and quickly, and then rebounds significantly for books currently in the public domain initially published before 1923. Since age should be a factor that depresses availability, the most plausible conclusion from the data is that the

⁵⁵ Instead of dividing the number of editions in each decade by 4, a separate divisor for each decade was calculated based on the average number of editions per book per decade. So, for example, the number of editions in the 1900's and 1910's was divided by 3, the editions from the 1880's and 1890's by 4, the editions from 1870's by 5, and the editions from the 1840's, 1850's, and 1860's by 6. On average, the reduction for all public domain books is 4. In general, the number of editions increases with the age of the title, perhaps for evolutionary reasons (only the strongest titles survive over time).

⁵⁶ Amazon does not provide a method for searching its database only for public domain titles. This is not surprising, given that Amazon only knows that dates of the editions it sells and not the original publication date of all book titles.

expiration of copyright makes older works reappear. A corollary hypothesis is also supported by the data: Copyright helps make books disappear.

Age seems to be very relevant within both the subset of post-1923 books still under copyright and the subset of pre-1923 books in the public domain. Note, however, the steeper decline in the number of copyrighted books over time: 2000-2010 (254 titles) to the 1990's (109 titles) to the 1980's (29 titles). This is not a gently sloping downward curve! Publishers seem unwilling to sell their books on Amazon for more than a few years after their initial publication. Part IV discusses business models, tax laws, and case law that helps explain why new editions of books disappear; copyright law then deters their reappearance until the copyright expires. On the left side of the graph before 1920, the decline presents a gentler time-sensitive downward sloping curve. The difference in the rate of decline between the public domain subset and the copyrighted subset demonstrates publishers' preferences for marketing books that are less than twenty years old.

The chart, of course, is somewhat misleading because it fails to account for the difference in the number of book titles published each year. Although the number of books published in each year for the last 200 years is not known, fewer books were undoubtedly published in the 1800's, when type was set by hand, as compared with more efficient methods developed during the mechanical typesetting and computer eras.⁵⁷ Of course, the population of the United States also increased over the same time, generating more readers and, as education became more universal in the late 19th and early 20th centuries, a higher percentage of literate consumers appeared. As a proxy for the actual number of books published each year, data were collected from two sources.

⁵⁷ See ROBERT BRINGHURST & WARREN CHAPPELL, *A SHORT HISTORY OF THE PRINTED WORD* (2000).

First, the WorldCat library catalog⁵⁸ of 72,000 libraries around the world was mined to identify the number of titles from each publication year between 1800 and 2010, counting only those titles published in English but not originating in English-speaking countries outside the United States.⁵⁹ Surely, more titles were actually published each year than are held in WorldCat libraries, but as long as the percentage of missing titles does not vary significantly from year to year, then yearly changes in the number of WorldCat books per year should consistently track the changes in books published per year. As one would predict with a direct measure of publication rates, the number of WorldCat titles counted each decade increased steadily until the 1990's when a well-documented decline in the number of printed books published began.⁶⁰

Then, the rate of the increase in the number of WorldCat books over time was compared to the rate of increase in U.S. copyright registrations for books over time. Professor Raymond Ku provided data on the number of yearly copyright registration for books from 1870-2006.⁶¹ Although the number of registrations in a particular year does not necessarily represent the number of books published, the change in the registration rate over time conceivably tracks the change in book publication rates over the same time period. We found, for example that about 6.0 more books were registered in 1990 than in 1910, suggesting a similar increase in the number of books published in 1990. The registration data bolsters confidence in the WorldCat data where the difference in the number of library holdings of books from the 1910's and the 1990's showed a strikingly similar increase of 6.3 times.

⁵⁸ WORLD CAT, <http://www.worldcat.org/> (last visited March 1, 2014); *WorldCat*, WIKIPEDIA, <http://en.wikipedia.org/wiki/WorldCat> (last visited March 1, 2014) (stating that two billion items are searchable in its global consortium of 72,000 cooperating libraries).

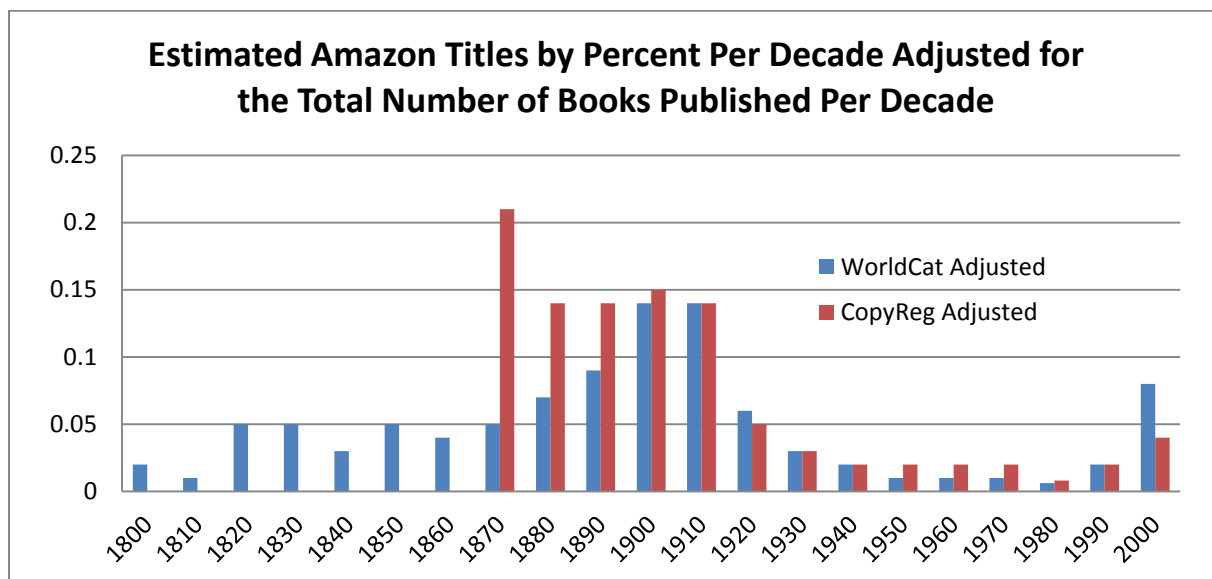
⁵⁹ The search string used in the WorldCat search was "la= "eng" not pl: scotland not pl: ireland not pl: britain not pl: wales not pl: britain not pl: australia not pl: canada and yr: 1800."

⁶⁰ See UNESCO, STATISTICAL YEARBOOK (1990-2010).

⁶¹ See Raymond S. Ku, *Does Copyright Law Promote Creativity? An Empirical Analysis of Copyright's Bounty*, 62 VAND. L. REV. 1669 (2009).

Figure 3 below accounts for the difference in the number of books published each year, normalizing to the decade of the 1990's when the highest number of books was published. The blue columns depict the adjustment based on the WorldCat data and the red columns depict the adjustment based on the copyright registration data. The closeness of the estimates⁶² during the entirety of the twentieth century from these two distinct sources generates some confidence in the accuracy of the attempted measure of the true relationship of copyright status to availability. The negative effect of copyright seen in Figures 1 and 2 becomes even more exaggerated:

Figure 3



NOTE: The publication dates of English language books in WorldCat library holdings and Copyright Office registration data serve as proxies for the number of books published per decade. Patterns of library holdings and changes in copyright registrations are used to estimate changes in the numbers of books

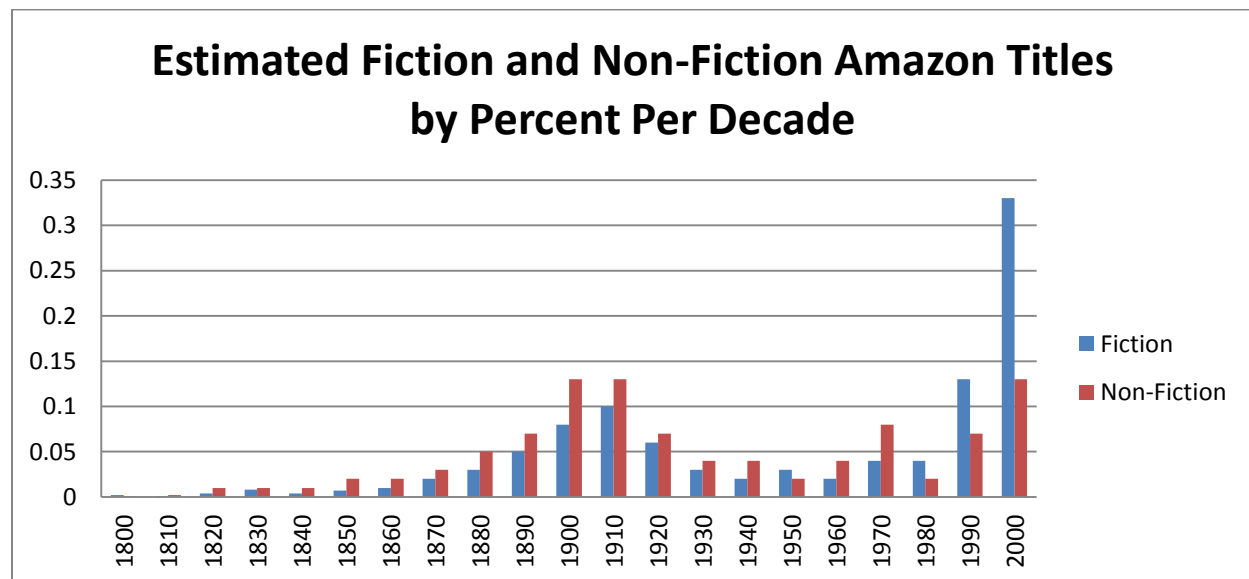
⁶² At both tails, the registration estimates and WorldCat estimates diverge. For the decade 2000-2009, the WorldCat adjustment is higher, perhaps because of tough economic times and the development of digital resources led to fewer purchases of new books. The drop in library holdings for that decade would not necessarily signal a proportional drop in publishing; therefore, the adjustment based on registration data may be more accurate. At the other end of the scale, the registration data results in a higher adjustment for the last decades of the 19th century. It may well be that incentives to register in the era of typesetting by hand were lower than in the era of movable type. An author in the 1870's or 1880's had fewer worries of quick and rampant piracy and therefore a diminished incentive to incur the cost and trouble of registering in Washington, D.C. Comparatively fewer registrations result in a more significant upward adjustment.

published per decade. The percentages in Figure 2 are adjusted to account for estimated trends in book publishing, normalized to the decade when the most books were published, the 1990's.

Consider the comparison of 1980 to 1880 as an illustration. Of the sample, only 25 titles were published in the 1980's whereas an estimated 38 were published in the 1880's. The WorldCat data suggest that almost 7 times as many books were published in the 1980's as in the 1880's. The blue column in Figure 3 above accounts for the seven times difference in the number of books published and provides a further insight into the correlation of copyright to availability. The estimate based on the difference in copyright registrations between 1880 and 1980 suggests any even greater disproportion of almost 18 to 1.

Figures 1, 2, and 3 above include fiction and non-fiction works. Figure 4 below divides the initial publication data into fiction and non-fiction columns, not adjusted for the increasing number of total books published each decade.

Figure 4



NOTE: Amazon only tracks the number of editions it sells (one per ISBN), so it does not know the number of titles it offers. Because books frequently are published in more than one edition, estimating the number of titles available on Amazon requires dividing the number of editions in the random sample by the average number of editions per title. The estimate is given in terms of the percentage of titles

likely available on Amazon from each listed decade. Here, estimates are given for both fiction and non-fiction works, categorized on the basis of guesses made from the title of the each book. Foreign language books and editions with especially ambiguous titles were omitted from the estimate.

The figure excludes 403 titles which could not be easily categorized, including 134 foreign language titles. The general pattern of disappearance and reappearance is approximately the same for both the fiction and non-fiction titles, although the ratio of public domain to copyrighted works in the overall totals varies. Among the fiction works, the public domain/copyright mix is 36% to 64%, while the ratio of public domain to copyrighted works within the non-fiction category is 55% to 45%. Given that the non-fiction category in this study is dominated by literary criticism, literary biography, essays, and theology, it may be that publishers in the 19th Century were more interested in publishing these sorts of works than publishers in the 20th Century.

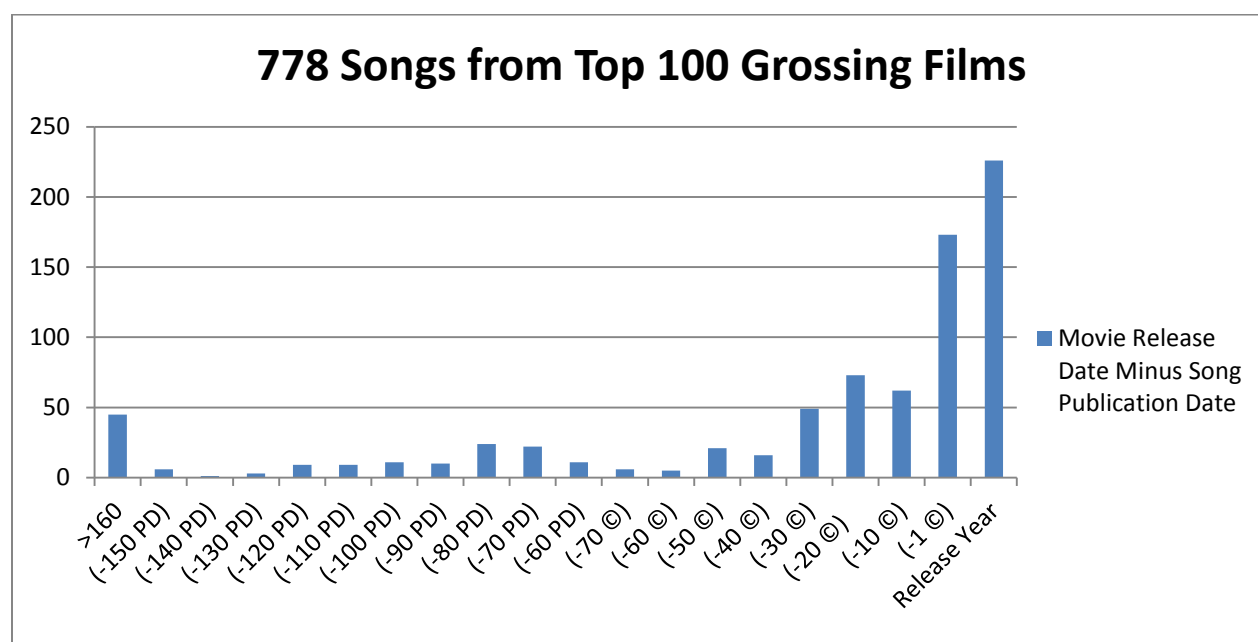
B. The Market for Music on Amazon DVD's

The effect of copyright law on the availability of music as it appears on new DVDs sold by Amazon is also negative, but not quite so dramatic. Figure 5 below displays the distribution of almost 800 songs found in the top 100 highest grossing movies of all time. Rather than organizing the data by the song publication year or movie release date, the chart illustrates the difference between the two. In other words, it measures how far backward movie directors were looking for music. Because the study attempts to measure the effect of legal status on the decision to use a song, it was necessary to compare the date of the movie release with the date of the song's publication to learn how the use of the song correlated with its copyright status at the time the movie was released.

Because of changes in copyright term duration, Figure 5 subdivides songs published 60-80 years before movie release into two categories—songs that were in the public domain at the

time of the movie release and those that were not. In all other categories, copyright status is self-evident (80-plus-year-old songs are always in the public domain while 60-minus-year-old songs are always copyrighted).⁶³ The 60 to 80 year subdivisions are made necessary by changes made in the 1976 Copyright Act (and during some years immediately prior thereto⁶⁴) extending the term of protection from 56 to 75 years for many existing works.⁶⁵ For example, a song that was published sixty years before it appeared in a 1950 film was in the public domain when the director chose to include it. A 60-year-old song appearing in a 1985 movie was not in the public domain at the time of the movie release (nor is it now). Comparing the legal status of songs in the 60 to 80 year prior-to-release categories illustrates in a nutshell the effect of legal status on use:

Figure 5



⁶³ In theory, songs that were 56 to 59 years old at the time of the movie release could also be in the public domain, but the study reveals only a couple of outliers in that category.

⁶⁴ From 1962-1976, Congress on a yearly basis extended the term of copyright for existing works by one year. See *supra* note 20.

⁶⁵ See *supra* note 20 for details on copyright term calculation and historical changes to term length.

NOTE: Songs are grouped by the difference between the publication date of a song and the release date of the movie in which it appears. “Release year” captures all songs that were written for the movie in which they appeared, otherwise the grouping is by 10-year increments, except for >160, which collects all songs that were 160 or more years old at the time of the movie release. Public domain or copyrighted status is indicated for each group, requiring two categories for songs in the 60-80 year range which could be copyrighted or not at the time of release, due to changes in copyright term duration over time.

Although the shape of the curve in Figure 5 roughly tracks the curve for books seen in Figure 2, the reappearance of older songs is much less pronounced. Even so, the upward slope starting with songs in the public domain (60+ years) is statistically significant.⁶⁶ Three times as many 60 to 80 year-old public domain songs (33 titles⁶⁷) were used in movies than 60 to 80 year-old songs still protected by copyright (11 titles). Not surprisingly, the sample is dominated by songs published the same year as the movie’s release date because many songs were written especially for the movie in which they appeared (29%). A high percentage of songs (22%) were 1 to 10 years old at the time of movie release, perhaps reflecting the frequent choice to set a movie plot in the near-present day.

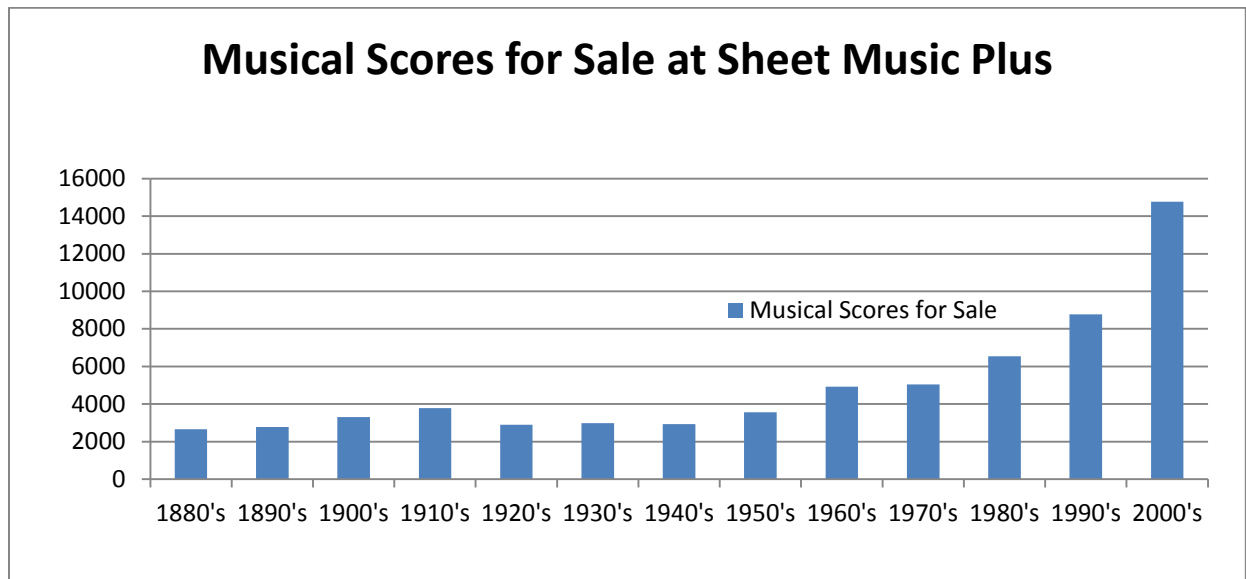
Interestingly, the shape of the chart in Figure 5 also resembles one constructed in preliminary research in an attempt to measure the market for sheet music at the world’s largest sheet music web site, SheetMusicPlus.com (“Sheet Music Plus”). One cannot search by year on the site, but each year from 1880-2010 was entered as a sole search term. The data is very noisy, because searching with a four digit number generates not only scores that list an identical publication year, but also scores that list birth or death dates of composers that correspond to the queried year and also stock numbers that happen to coincide. In addition, a search for 1925, for example, would return a 1925 edition of an 1850 piece. For this reason, the data are not

⁶⁶ See Appendix A for full statistical analysis.

⁶⁷ The data is influenced by the number of public domain songs in the movie *Gone With the Wind*, which contains 14 songs that were published between 70-80 years before the 1939 movie release date.

particularly reliable, but the results may be instructive in light of the more rigorous data collected to construct the movies-in-music chart above in Figure 5.

Figure 6



NOTE: Each year from 1880-2010 was entered as a unique search term using the advanced search function at www.sheetmusicplus.com. Years ending in 00 were omitted because many score descriptions make references to decades outside the context of publication date. The total number of scores identified was 65,000.

As with the music-in-movies chart in Figure 5, one sees the same mild increase in availability once dates prior to 1923 are considered. The drop off from the decade of the 2000's to the 1960's is much less pronounced with sheet music, but that might be explained by a greater present demand for older scores. Any casual listener will have noted the trend on the radio toward "oldie" stations, which keep older musical works in front of consumer. Movie directors favoring a contemporary or futuristic setting for their films might not have the flexibility to look backward as frequently. Although interesting, the noisiness of the data prevents the drawing of any strong conclusions about the sheet music market.

The difference in the magnitude of the effect of legal status on books and music in movies is probably explained by the comparative economics of the book and film trades. A book publisher wishing to sell a public domain title need only find the title in the public library and scan it (or find it on Google Books), edit the typeface and graphics with any widely available publishing software program, and send the manuscript off to be printed.⁶⁸ The former copyright owner need not be contacted and no license fee need be paid. Sometimes these tasks can be performed in less than a day,⁶⁹ and the savings over locating a copyright owner, negotiating and paying a licensing fee are substantial.

A movie director likely saves marginally less by choosing a public domain musical composition. A musical composition, standing alone, cannot be employed in a film; it must be played and recorded first. Therefore, a director must hire a singer, band, or orchestra to make a new recording appropriate for inclusion in the film or pay a fee to the copyright owner of an existing sound recording for permission to adapt that recording for the film. A director choosing a recording of the Sex Pistols singing “God Save the Queen” must pay a fee to the owner of the sound recording even though the musical composition is in the public domain.⁷⁰ Although no fee need be paid to the composer, the savings are marginal and are perhaps dwarfed by the cost of making a new recording or obtaining permission to use an existing recording. If the marginal savings of choosing a public domain composition for a film are smaller than the marginal savings of choosing to publish a public domain book, then one would expect to see the more modest increase in the upward curve of older public domain songs depicted in Figures 5 and 6.

⁶⁸ See ANDRA MIKOS, *THE PUBLIC DOMAIN PUBLISHING BIBLE* (2009); ADAM PEARSON, *HOW TO CREATE, FORMAT, PUBLISH, PROMOTE, AND PROFIT FROM THE EBOOK OPPORTUNITY* (2012).

⁶⁹ See *id.*

⁷⁰ The Wikipedia entry for *God Save the Queen* includes an image of sheet music dating from 1745. See *God Save the Queen*, WIKIPEDIA, http://en.wikipedia.org/wiki/God_save_the_queen (last visited March 1, 2014).

The sample of songs from the top 100 grossing movies of all time has particular interest because the songs have been encountered by large numbers of the public. By definition, the list contains no obscure art films that barely reached the silver screen, containing a soundtrack that was heard by virtually no one. Nonetheless, a completely random sample of all films listed on BoxOfficeMojo was also conducted and the data from the songs told an interesting story. The sample of random movies contained many fewer public domain songs than the sample of top-grossing movies. At the time of movie release, only 8% of songs from the randomly selected movies were in the public domain, whereas 25% of the songs appearing in the top-grossing movies were in the public domain at the time of release.

This difference presented a puzzle: Why would the top-grossing films use three times as many public domain compositions as the randomly selected films? Top-grossing films presumably have bigger budgets than randomly selected films, so it seemed unlikely that directors of top grossing films were more price-sensitive and therefore chose to include marginally cheaper public domain compositions. The top-grossing films might have contained more historical plots and settings than the random films, requiring a further reach back into the musical past. A partial answer is found in a bias toward new films in the BoxOfficeMojo database caused by its decision to only list movies with known box office returns. Not surprisingly, movies where box office data are available tend to be newer movies. The box office gross for a 1953 film by a defunct studio may not be available, but almost all newer films report their box office receipts. For this reason, the median release date of the 100 randomly sampled movies from BoxOfficeMojo was 2002. The median release date of the top 100 grossing movies of all-time was 1977, a striking difference.

This difference modified the puzzle: Why would directors of movies with a median release date of 1977 more frequently choose songs that were 60, 70, 80 or more years old at the time of production than did directors of movies with a median release date of 2002? Consistent with the evidence that both legal status and age are relevant to the availability of a work, a testable hypothesis emerged. Because of changes in the duration of copyright, directors of movies released before 1977 did not have to look backward so far to access free public domain material. Since the analysis of both books and music above suggested that the age of a song is also relevant to the movie inclusion decision (51% of songs are published within 10 years of movie release date), one would expect that movie directors who only had to look backward 56 years to access the public domain (e.g., directors of movies from the 1930's to 1960's) would have been more likely to choose a public domain song than the director of a movie, say, in 2010 or 2011, who had to look backward 87 or 88 years respectively to find a public domain song.

This hypothesis was testable by a further examination of the song sample from the top 100 grossing movies (examining the random song sample was fruitless because the earliest movie was released in 1981). The top grossing movies contained equal numbers of films from before and after 1977, a convenient date, given the timing of 1976 term extension. With an equal number of movies from either side of 1977 (and an almost equal number of songs in them⁷¹), the sample provided sufficient data to test whether movie release date, and therefore accessibility to free public domain material, had influenced the availability of older songs in blockbuster movies.

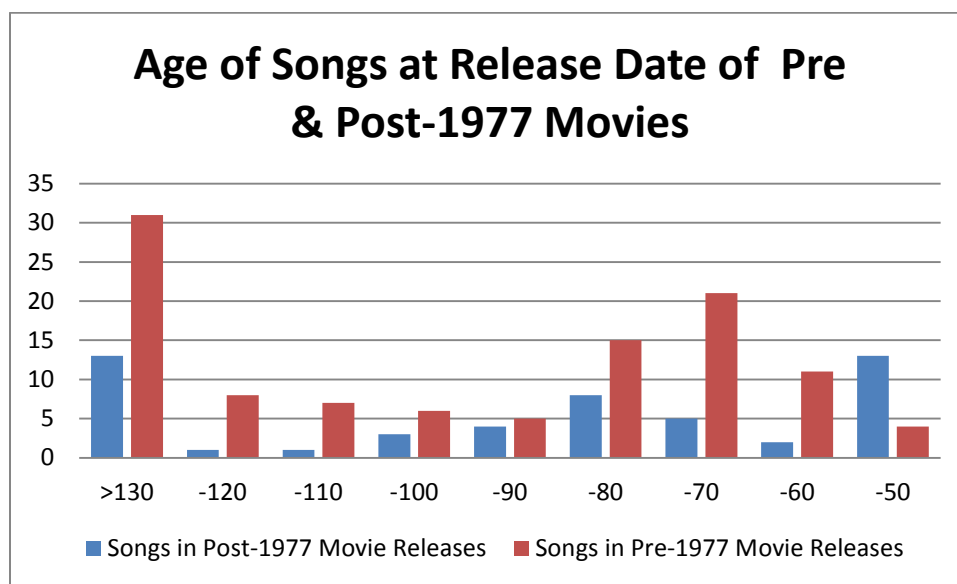
An initial analysis of the distribution of public domain songs in movies on both sides of the 1977 median date suggested a correlation between movie release date and the legal status of songs in the movie. Of the 129 public domain songs in the entire sample, 76% (98/129) were

⁷¹ The data set contains 384 songs from the pre-1977 movies and 341 from the post-1977 movies.

found in movies released prior to 1977. Only 24% (31/129) were found in movies released after 1977.

A more sophisticated analysis, illustrated below in Figure 7, compares the difference in years between the publication dates of the songs and the release dates of the movies in which they appeared in both the pre- and post-1977 sets of movies. The chart begins on the right side with songs, all protected by copyright, that are between 50 to 60 years old and then shows the difference between the two sets as the public domain songs are considered:

Figure 7



NOTE: A paired t-test shows the significance of the difference in 60+ old songs between pre- and post-1977. The p-value is 0.02105. So the rate of use of 60+ year-old songs in the pre-1977 is significantly higher than the post- 1977. Null hypothesis: rates of use are the same. Alternative hypothesis: higher rate of use in pre- 1977 than post- 1977.

	N	Mean	Std
Songs in post- 1977 Movie releases	9	5.556	4.746
Songs in pre- 1977 Movie releases	9	12	8.958

Paired t-test: $t = -2.4161$, $df = 8$, $p\text{-value} = 0.02105$.

One notices immediately that the songs from the pre-1977 movies dominate every age category except 50 to 60 years before the movie release date, when almost all the songs are still protected by copyright.

The analysis supports the earlier suggestion that copyright status has a significant effect on the availability of songs in movies. It also suggests why the sample of 100 randomly selected movies with a 2002 median release date contained so many fewer public domain songs. The directors of the randomly selected movies had to look decades further back to mine the public domain than did the directors of the blockbuster movies that had a media release date of 1977.⁷² If songs get progressively less desirable to place in movies as they age, then copyright seems to work hand-in-hand with Father Time to help make musical works disappear. The analysis, however, does not explain why the pre-1977 movies contain significantly more songs that are 100-130+ years old. Songs of that vintage were in the public domain for all directors of all movies in the sample, so changes in copyright duration should not have a direct effect on the choice to use songs of extreme age.

IV. THE DEMAND FOR OUT-OF-PRINT WORKS AND ALTERNATIVE MARKETS

Changes in copyright term-duration cannot possibly explain every decrease in the availability of new copies of works produced in the 20th Century. Rights holders make decisions that are not driven by copyright law about when to stop marketing new copies of works, and alternative markets may have developed to satisfy the demand for works. In fact, concerns over availability are misplaced if no demand for missing works exists. This section examines the

⁷² Because of legislative changes in copyright length, directors of movies released during the talkie era from 1929 to 1964 only had to look backward 56 years to find a song with a public domain date of initial publication. Directors of movies released from 1964 to 1997 had to look back between 56-75 years, depending on the year of release. Directors of movies in 1998 to 2013 had to look back 75-90 years, depending on the year of release. *See supra* note 20.

demand for out-of-print works and looks at data from digital and used goods markets that could potentially satisfy that demand. Finally, it offers several reasons why in the book publishing market--unlike the music market--copyright still seems to stymie the reappearance of old books.

A. *The Demand for Out-of-Print Works*

If consumer demand for the millions⁷³ of out-of-print works is zero, then the story told by the data is quite banal. If publishers cease production when demand evaporates, then Figure 2 merely depicts the correlation between the age of a work and its value. However, Figure 2 itself suggests that rights holders do not perfectly satisfy demand for their works. Availability spikes for pre-1923 works, suggesting strongly that some pent-up demand exists for older works, at least when cheap and efficient print-on-demand publishers can offer them without having to negotiate for the right to copy. Before 1998, when a new set of out-of-print titles fell into the public domain every year, the more valuable ones were quickly picked up and made available.⁷⁴ This opportunistic business model suggests some rents were being left uncollected, unless firms publishing public domain works consistently operate at a loss. Nonetheless, calculating precisely the demand for books and music is extremely difficult. Publishers consider sales data to be confidential and are unwilling to share it with researchers, making estimates of the size of the unmet demand for out-of-print works speculative. Section C will suggest why copyright owners do not meet demand.

One recent paper takes a random sample of out-of-print books available in Kindle format and compares them to a sample of similar out-of-print books that are unavailable as eBooks.⁷⁵

⁷³ A search for out-of-prints books on www.booksinprint.com on January 23, 2014, returned 3,297,267 editions. This is almost undoubtedly a vast undercount as Books in Print relies on data from publishers willing to provide information on their titles.

⁷⁴ See Heald, *supra* note 16 at 1044-45.

⁷⁵ See Michael D. Smith, Rahul Telang, and Yi Zhang, *Analysis of the Potential Market for Out-of-Print eBooks*, available at <http://ssrn.com/abstract=2141422> (posted August 2012).

The study then analyzes sales data of the out-of-print books available in Kindle versions and calculates the value of the out-of-print book.⁷⁶ Using a Bayesian Propensity Score Matching technique, the authors estimate that making almost 2.7 million out-of-print books available as eBooks would generate \$740 million in revenue and \$860 million in consumer surplus.⁷⁷

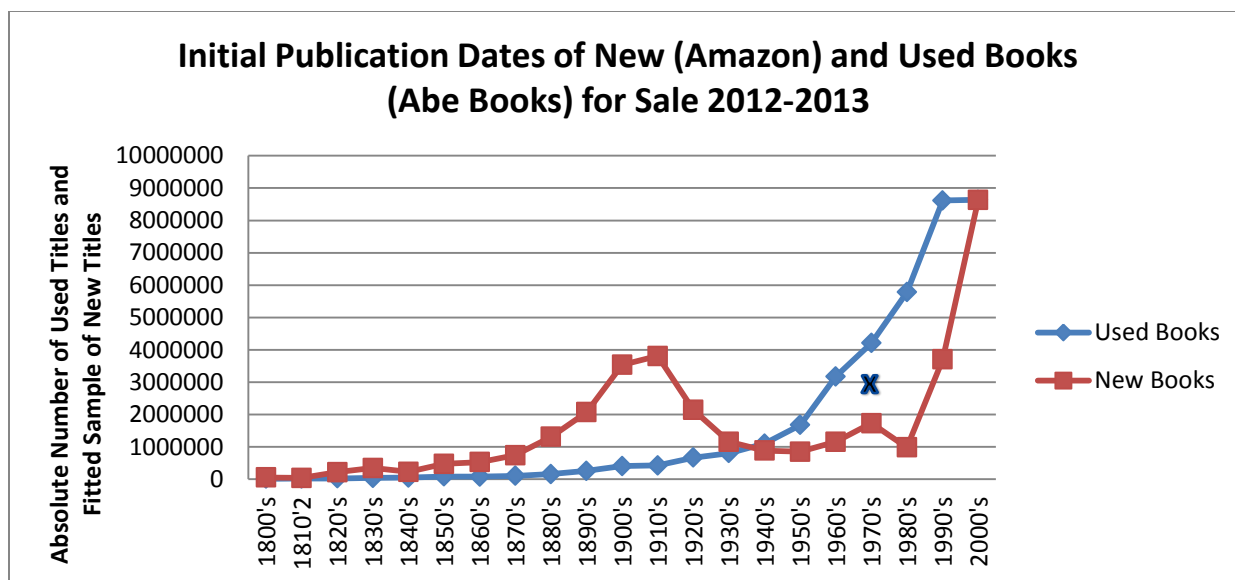
One could also consider the used book market as an indicator of consumer demand for works. Figure 8 below plots the availability of editions by the decade of their publication at the world's largest on-line marketplace for used books: Abebooks.com. According to the Online Booksellers Association, over 13,500 used book dealers sell on Abebooks, which was acquired by Amazon in 2009.⁷⁸ Used booksellers typically buy books in order to sell them at a profit, and their inventories are indicative of their purchasing decisions and their belief that consumer demand exists for the volumes that they offer. Figure 8 below depicts all the volumes of all the editions available from Abebooks.com for the years 1800-2010. Not surprisingly, the number of volumes available from each decade decreases as the books age, but it does not drop nearly so steeply as the number of new books available on Amazon over the same time period. The figure below fits and overlays the downward-sloping curves for new and used books.

Figure 8

⁷⁶ *Id.* at 9-14.

⁷⁷ *Id.* at 1.

⁷⁸ See Chris Volk, *The Pros and Cons of Abebooks.com for Buyers and Sellers*, available at [http://www.ioba.org/newsletter/archive/8\(2\)/toolbox2.php](http://www.ioba.org/newsletter/archive/8(2)/toolbox2.php) (IOBA Standard, the on-line journal of the Independent On-Line Booksellers Association).



NOTE: The used book curve indicates the absolute number of used books for sale in 2013 by their publication years (as offered by Abebooks) including all sellers listing on the web site. The new book curve is the same as in Figure 2 that estimates the percent of new books for sale on Amazon by decade of initial publication. The new book curve is fitted to the graph using a multiplier generated by dividing the number of used books from the 2000's offered by Abebooks by number of new books from 2000's in the Amazon sample.

The downward-sloping curve depicting used books for sale over time is not nearly so steep as the curve depicting new books for sale over the same period. As a proxy for demand, the used book curve shows that demand is sensitive to the age of the book, as one might expect, but not nearly so sensitive as with new books. The gap between the two curves, indicated by x , suggests a demand for works that is being satisfied by the used book sellers that is not being satisfied by publishers. Although the size of that demand cannot be quantified in real terms, it seems quite clear that a continuing demand for older works exists that is not being satisfied by publishers of new editions. Interestingly, at least one study suggests that the market for used books does not suppress the market for new editions of the same books because “used books are

poor substitutes for new books for most of Amazon's customers."⁷⁹ The study finds that cross-price elasticity between the markets is only .088, suggesting strongly that the existence of a used book market does not discourage publishers from reprinting older books.⁸⁰

Unfortunately, evidence of demand for out-of-print music is more difficult to document. No equivalent of the Books-in-Print database⁸¹ exists for music, and there is no mega-marketplace like Abebooks for used sheet music. Sales and licensing data are considered confidential by music publishers, and ASCAP refused to provide the author with airtime data for famous old songs. Moreover, due to amendments to U.S. copyright law, no sound recordings will fall into the public domain until 2067,⁸² which makes comparing sets of copyrighted and uncopyrighted recordings virtually impossible.

B. *Alternative Markets for Out-of-Print Works*

If works that are out-of-print and unavailable in new copies on Amazon can be easily obtained in other forms elsewhere, then the phenomenon of "missing books" on Amazon is hardly cause for alarm.

1. The Market for Out-Of-Print Books in eBook Format

The ease and low cost of digitizing older texts suggests that many out-of-print books might reappear exclusively as eBooks, thereby satisfying latent demand and solving any availability problem. In fact, data on bestselling public domain books from 1913-22 show that eBook publication of old texts is attractive to many publishers. In 2014, 94% of 165 public

⁷⁹ See Anindya Ghose, Michael D. Smith, & Rahul Telang, *Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact*, 17 INFORMATION SYSTEMS RESEARCH 3 (2006).

⁸⁰ *Id.*

⁸¹ www.booksinprint.com (providing publication information on millions of in-print and out-of-print books).

⁸² See 17 U.S.C. § 301(c) (2006) (extending protection to pre-1972 sound recordings to the year 2067).

domain bestsellers from 1913-22 were available in eBook format, up from 48% in 2006.⁸³ An aggressive and competitive market for publishing public domain eBooks is currently thriving.

Data on the eBook availability of copyrighted bestsellers from the same era tells a different story. Of 167 bestsellers from 1923-32 still under copyright, only 27% (45/167) had been made available as eBooks by publishers by 2014. And of those 45 copyrighted eBooks, only one was out-of-print in hard copy format. Even in the eBook market, copyright status seems to stand as a significant impediment to satisfying demand. In fact, the availability gap between public domain and copyrighted bestsellers is even larger in the eBook market than in the in-print market. An earlier study reported that as of 2006, 98% of the 165 bestsellers from 1913-22 were in print, while 78% of the 167 copyrighted bestsellers were in print, a significantly smaller difference than the present eBook gap.⁸⁴

Since the eBook data from 1923-32 bestsellers is based on the market for less than 200 works of fiction published over a 10-year period, a further study was conducted of a sample of 950 fiction and non-fiction books reviewed in the *New York Times Book Review* (NYTBR) from 1930-2009.⁸⁵ Of most interest, perhaps, is the number of out-of-print books from the sample that are now available in eBook form. Of the 292 NYTBR books that are currently out-of-print (31% of the total), only 26 (9%) were available for purchase as an eBook in January 2014.⁸⁶ Figure 9 below charts the availability of all of the NYTBR titles in eBook format.

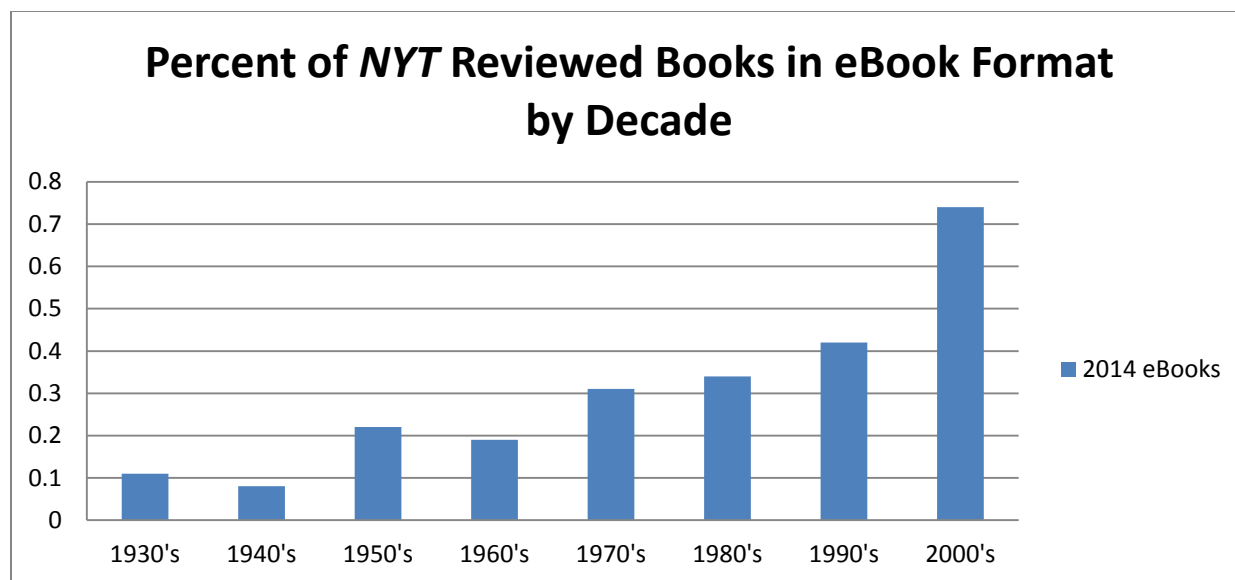
Figure 9

⁸³ This data updates figures reported in Heald, *supra* note 16.

⁸⁴ See Heald, *supra* note 16 at 1040-41.

⁸⁵ The sample consists of the first book reviewed in the *New York Times Book Review* in the first issue of each month. Since the University of Illinois Library does not have an absolutely complete collection, the first book reviewed in the second or third issue of a particular month was sometimes used. In some cases, entire months were missing, which reduced the total sample to 950 (instead of 960).

⁸⁶ Print status and eBook availability data were collected at both at booksinprint.com and amazon.com.



NOTE: The chart uses the date each book was reviewed, not its publication date; however, the *NYTBR* typically reviews books close to their publication dates. The percent recorded is the number books with eBook versions are available on Amazon or Books-In-Print divided by the number of reviewed books sampled from each decade (approximately 120).

Of course, both the list of bestsellers from 1923-32 and the *NYTBR* sample are skewed toward more prominent books, but the higher potential demand for those works makes their absence in the eBook market all the more striking and reemphasizes the effect of copyright on availability. In the absence of copyright, surely one could find a publisher providing eBook versions of popular classics like *The Gulag Archipelago*, *Gentlemen Prefer Blondes*, and *The Magnificent Obsession*.

2. The Market for Used Books

The sample of books reviewed in the *NYTBR* from 1930-2010 generated a list of 292 out-of-print books. This list of out-of-print books created the opportunity to learn whether the used book market might supply a significant number of replacement volumes. All 292 titles were queried on the world's two largest marketplaces for used books, Abebooks and Amazon, and the

number of volumes available from each seller was recorded.⁸⁷ If reasonably priced used editions of out-of-print books are easy to obtain, then any negative effect of copyright on the new book market might be substantially alleviated.

Given that Abebooks and Amazon contain the inventory of at least 13,500 used book dealers, a decent picture of the market was obtained. The out-of-print *NYTBR* titles from the 1930's, 1940's and 1960's had approximately enough used volumes available to satisfy the needs of a single classroom: 31, 32, and 39 volumes on average available per title. Titles from the 1950's averaged 74 available volumes, while titles from the 1970's to 2000's fared rather better, with 136, 190, 148, and 171 volumes available on average. The casual reader or researcher seems to have good access to out-of-print *NYTBR* titles in a used book format, although adequate numbers of older titles do not appear sufficient to satisfy potential classroom needs in many cases.

Optimism about the market for used books, however, should be tempered for three reasons. First, research on price elasticity by Ghose, Smith, and Telang (2006) found that “used books are poor substitutes for new books for most of Amazon customers.”⁸⁸ Although one might doubt this conclusion in some book sub-markets, for example, university textbooks, their paper suggests that the market for out-of-print books might not be wholly satisfied by used book substitutes. In addition, the distribution of used volumes on Abebooks demonstrates that substituting used books for new in the classroom situation is unwieldy, since most sellers only carry one volume. For example, Abebooks offers more than 388 volumes of Willa Cather's *Shadows on the Rock*, but those volumes are held by 366 different sellers. Buying in bulk is therefore clumsy.

⁸⁷ Since most large used book sellers offer their books on both Amazon and Abebooks (owned by Amazon), the largest number of volumes offered by either was used.

⁸⁸ See *supra* note 79 at 3.

Finally, and most importantly, the sample of *NYTBR* books is skewed toward prominent and culturally important publications, books that are more likely to be collected by used book sellers. One measure of the prominence of the titles in the sample can be seen in the holdings of the Chicago Public Library system. Let's consider only the oldest 470 titles reviewed between 1930-1970. During that period, more than 2.14 million books were registered with the copyright office, a rough proxy for the number of books published from 1930-1970. The Chicago Public Library system shows approximately 160,000 books (7%) in its collection from the same 40-year period. Yet, of the 950 *NYTBR* books published from 1930-1970, more than 58% are contained in the library holdings. The *NYTBR* books are clearly more prominent than the average book published at the same time. More research needs to be done on the availability of more typical and obscure books from the mid-twentieth century before concluding that the used book market fully satisfies the demand for older books.

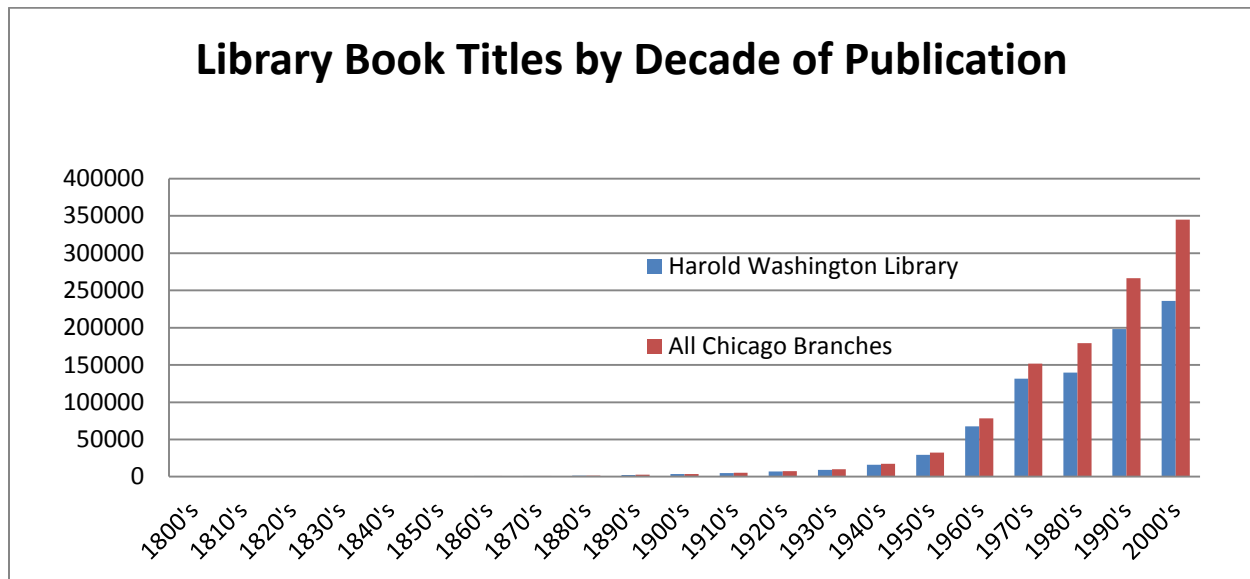
3. Public Libraries

Googlebooks, operating without the constraint of copyright, has made millions of books published before 1923 available for free download. Given the ready access to digital versions of pre-1923 books in the United States, it would be difficult to conclude that digital public domain books are suffering from an availability crisis, although these .pdf versions are often not as readable or as well-formatted as an eBook or an in-print edition.

Public libraries cannot make copyrighted works freely available in digital form, but they do maintain an important reservoir of works. Figure 10 below suggests that one major public library, the Chicago Library System, makes some books available that are not offered in new copies on Amazon. Although the number of volumes from each decade declines steadily over

time, the drop is not nearly so precipitous as in Figure 2, suggested that books disappear from the Amazon bookshelf before they disappear from a large metropolitan library.

Figure 10



NOTE: Total Chicago Public Library system holdings and total holdings at the largest downtown branch were searched by decade of book publication date (as reported by the library). The library search engine reports a 1995 edition of an 1895 book as published in 1995, which skews the curve toward newer books.

Despite the availability of many titles, the holdings still constitute a relatively small percentage of the books published in each decade. For example, the 79 branches of the Chicago Library system offer only 33,000 titles of the approximately 500,000 published in 1950's.⁸⁹

Of most interest may be the number of out-of-print books that are preserved for the public in libraries. The Chicago library system does an imperfect job of collecting even the more prominent titles from the twentieth century. Of the 292 out-of-print books in the *NYTBR* sample, approximately 55% may be obtained from one of Chicago's libraries. The availability rate of

⁸⁹ More books, however, originally published 1950's may be available. The library web site search engine reports publication date of the edition it holds, not the original publication date of the title, so a book from 1952 that was reprinted in 1962 is charted above in the decade of the 1960's. Of course, some the books currently listed in the decade of the 1950's would have been 1940's reprints, and so on. This cascading effect undoubtedly skews the overall curve upward somewhat.

those books varies significantly by decade: 1930's (35%), 1940's (29%), 1950's (57%), 1960's (30%), 1970's (62%), 1980's (78%), 1990's (82%), 2000's (79%).

4. Digital Music and YouTube

Data from Brooks (2005) suggested that many older musical titles were missing in the markets for CD's and digital music downloads.⁹⁰ He found that only 14 percent of famous historical recordings from 1890-1964 had been digitized by their owners (22 percent had been digitized by others, almost certainly without authorization).⁹¹ This suggested that looking for old tunes from that era on iTunes or on CD's might often be futile; however, testing whether Brooks' findings hold true in 2014 is difficult. The Amazon advanced search function for music does not allow searching by the author of a musical composition, only by title and artist. In addition, searching by title does not permit the use of quotation marks to tie words together. So, searching for the 1926 hit "Yankee Rose" returns 27 results that contain the words "yankee" and "rose" somewhere on the page, with apparently no results containing "yankee rose" conjoined.

The search function on iTunes, the world's most important marketplace for digital music, is even more research- and consumer-unfriendly. The current version of iTunes completely lacks an advanced search function, so it's impossible to limit a search to title, artist, or composer. For example, a search for the 1926 hit "Yankee Rose" returns 40 hits, topped by several versions by David Lee Roth of what looks to be a heavy metal tune from the '80's.⁹² One of the hits does identify an album entitled "Retro-Specht (1925-31)," which likely contains the 1926 song by Holden and Frankel, but without access to composer information, one can only make an educated guess. Searching becomes impossible for songs with more common titles, as with other hits from 1926 entitled "Horses," "Babyface," "Lucky Day," "Mississippi," or "Valentine." Finally,

⁹⁰ See *supra* note 18 and accompanying text.

⁹¹ *Id.*

⁹² See [http://en.wikipedia.org/wiki/Yankee_Rose_\(song\)](http://en.wikipedia.org/wiki/Yankee_Rose_(song)).

although the iTunes API allows somewhat greater precision in searching, the absence of an ISBN-like system for identifying music makes generating a truly random sample of what is on the iTunes “shelf” extremely difficult.

Nonetheless, one can get an idea of the extent of iTunes holdings by searching only the titles of older songs that have uniquely worded titles. A sample search of unique sounding titles in the iTunes database suggests that music publishers, unlike book publishers, have done a comprehensive job of digitizing their back catalogs since the 2005 Books report. Using a database of popular songs from 1913-32 from in a previous study,⁹³ five songs were selected alphabetically from the end of the list for each year which had long and probably unique titles (for example, “Where Did Robinson Crusoe Go with Friday on Saturday Night?” (1916) and “When Yankee Doodle Learns to Parlez-Vous Francais” (1917)). Of those 100 obscure old hits, 85 were available on iTunes as song downloads, most of them with recordings from multiple artists.⁹⁴ As opposed to books, the digital revolution seems to be ameliorating the problem of missing older songs (at least for those that were among top-selling 60 or 70 in a particular year). The sampled songs are listed in Appendix B.

More evidence of increased availability can be found on web sites like YouTube, where individuals with copies of musical recordings can upload them for free without rendering the website liable.⁹⁵ Under either the Digital Millenium Copyright Act safe harbor provisions⁹⁶ or

⁹³ See Heald, *supra* note 16.

⁹⁴ See Appendix B for a list of the songs and their ITunes availability.

⁹⁵ See *Viacom Int’l Inc. v. YouTube, Inc.*, No. 07 Civ. 2103, 2013 WL 1689071 (S.D.N.Y. Apr. 18, 2013) (granting Google summary judgment in lawsuit brought by Viacom suing YouTube for hosting infringing uploads). Although the Viacom litigation is still on appeal, much academic commentary has concluded that YouTube will prevail in cases where it lacks actual knowledge that uploaded material is infringing. Edward Lee, *Decoding the DMCA Safe Harbors*, 32 COL.-VLA 233 (2012); Jordan Sundell, *Tempting the Sword of Damocles: Reimagining the Copyright/Dmca Framework in a UGC World*, 12 MINN. J.L. SCI. & TECH. 335, 337 (2011); Jennifer M. Urban & Laura Quilter, *Efficient Process or “Chilling*

analogous common law rules,⁹⁷ YouTube appears to be neither directly⁹⁸ nor secondarily⁹⁹ liable for infringement until it receives notice from a complaining copyright owner.¹⁰⁰ This creates a potential market for older musical works that remain available to consumers if the copyright owner is willing to monetize the upload or otherwise tolerates the infringement.

In order to measure whether YouTube functions as an alternative market for old songs and to measure the possible effect of copyright law on availability within that market, a list of 385 hit songs from 1919-1926 was identified from prior research and each song was queried on

Effects”? *Takedown Notices Under Section 512 of the Digital Millennium Copyright Act*, 22 SANTA CLARA COMPUTER & HIGH TECH. L.J. 621 (2006); Andrey Spektor, *The Viacom Lawsuit: Time to Turn Youtube Off?*, 91 J. PAT. & TRADEMARK OFF. SOC'Y 286, 290-91 (2009).

⁹⁶ See 17 U.S.C. § 512(c)(1) (limiting liability to injunctive and equitable relief unless the service provider has actual or constructive knowledge, derives financial benefit, or does not remove infringing material). Most cases applying the DMCA have found Internet service providers in positions analogous to YouTube to qualify for the DMCA safe harbor. See *Perfect 10, Inc. v. CCBill, LLC*, 488 F.3d 1102, 1114 (9th Cir. 2007); *UMG Recordings, Inc. v. Veoh Networks, Inc. (UMG I)*, 620 F. Supp. 2d 1081, 1088 (C.D. Cal. 2008); *Io Group, Inc. v. Veoh Networks, Inc.*, 586 F. Supp. 2d 1132, 1148 (N.D. Cal. 2008); *Corbis Corp. v. Amazon.com, Inc.*, 351 F. Supp. 2d 1090, 1110-11 (W.D. Wash. 2004). See also *Viacom Int'l. Inc. v. YouTube, Inc.*, No. 07 Civ. 2103, 2013 WL 1689071 (S.D.N.Y. Apr. 18, 2013) (“But the governing principle must remain clear: knowledge of the prevalence of infringing activity, and welcoming it, does not itself forfeit the safe harbor. To forfeit that, the provider must influence or participate in the infringement.”).

⁹⁷ For a discussion of the common law approach to liability for on-line platforms, see Alfred Yen, *Third-Party Copyright Liability After Grokster*, 91 MINN. L. REV. 184 (2006); see also Brett White, *Viacom v. Youtube: a Proving Ground for DMCA Safe Harbors Against Secondary Liability*, 24 ST. JOHN'S J. LEGAL COMMENT. 811, 814-21 (discussing common law safe harbors as applied to YouTube).

⁹⁸ See *Cartoon Network, LP v. CSC Holdings, Inc.*, 536 F.3d 121 (2d Cir. 2008) (finding cable company not liable for making system available for customers to copy programs remotely on its servers); *Religious Technology Center v. Netcom On-Line Communication Services, Inc.*, 907 F. Supp 1361 (1995) (finding on-line platform that provided open storage for uploaded material was not directly liable for infringement unless it committed a voluntary act beyond merely making space available); Andrey Spektor, *The Viacom Lawsuit: Time To Turn Youtube Off?*, 91 J. PAT. & TRADEMARK OFF. SOC'Y 286, 290-91 (2009) (no direct infringement by YouTube).

⁹⁹ Since most cases hold that on-line platforms like YouTube qualify for DMCA safe harbor provisions, the application of other secondary liability doctrines like contributory liability and vicarious liability remains underdeveloped. Commentators have made persuasive arguments that YouTube lacks the requisite mental state and control over the infringer to be held liable under historical principles of secondary liability. See Yen, *supra* note 97; see also White, *supra* note 97, at 811, 814-21 (discussing common law safe harbors as applied to YouTube).

¹⁰⁰ Since liability requires actual knowledge, it is possible that such knowledge could come from a source other than the copyright owner (e.g., YouTube's own Content ID program).

YouTube. Roughly half of the compositions (1919-22) are in the public domain, and roughly half (1923-26) are not. Seventy percent of the public domain compositions had songs on YouTube while 77% of the copyrighted compositions from 1923-26 had songs on YouTube. The public domain compositions averaged 9.8 uploaded songs, and the copyrighted compositions averaged 14 uploads. Copyright status seems to provide little impediment to the availability of these 385 old songs on YouTube. One further point: attributing the increased availability of songs from the period 1923-26 to their copyright status is almost certainly premature. Previous data had suggested that songs on the more recent side of the 1923 divide were intrinsically more popular.¹⁰¹ Indeed, the songs from the period 1923-26 averaged 112,000 total views, while the songs from 1919-22 averaged only 39,000 total views. In the YouTube market, copyright status may simply not matter, at least for songs of more than a certain age.

In sum, between iTunes and YouTube, old music--at least hits from the past--seems to be quite readily available to consumers. However, as noted in Part II, the present difficulty of taking a random sample of songs from iTunes and analyzing them by date of initial publication makes it impossible to know whether the availability curve for music in general suffers significantly from the missing works phenomenon.

C. Impediments to Availability

Given iTunes and YouTube data, it appears that book publishers are not making their back catalogs as available as are music publishers. As already noted, only 27% of bestselling books from 1923-32 have been published as eBooks,¹⁰² while 84% of bestselling music from the same era can be bought in digital form on iTunes. This paper cannot offer a complete

¹⁰¹ See Heald, *supra* note 19 at 37 (While songs from 1913-32 were all under copyright, the songs from 1923-32 were significantly more likely to appear in movie soundtracks.).

¹⁰² See *supra* note 84 and accompanying text.

investigation of the differences in the music and book-publishing industries, but several explanations for the difference merit consideration.

1. Print Publishing Business Models.

Printing physical books is costly, and until a publisher adopts a print-on-demand format, it will be willing to incur the expense of printing and storing books only if it foresees a pre-set minimum demand for the edition. One recent estimate suggests that publishing business models require a residual demand of 500-1000 books in order to justify keeping a book in print.¹⁰³ Such minimums render many older books ineligible for reprinting. Even when sufficient demand is foreseen, a 1979 change in tax law by the Court in *Thor Power Tool v. Commissioner*¹⁰⁴ provided additional motivation for keeping inventory in book warehouses to a minimum. The decision changed the rules for writing down business inventory and required publishers to assign for tax purposes the full sales price on all their books, regardless of the rate of present sales or prospects for future full-price sales. According to one commentator, “*Thor Power* eliminated a tax dodge, and thereby made it more expensive for publishers to carry inventory from year to year. As a result, publishers have cut print runs in order to minimize inventory. They have also become quicker to dispose of inventory — i.e., pulp it — before the end of the fiscal year.”¹⁰⁵ The decision may help explain the sudden drop in book titles available on Amazon in the 1980’s as compared to the 1970’s. In any event, rigid business models and changes in tax law suggest one reason why hard cover books are not more frequently re-printed.

2. *Boosey & Rosetta*

¹⁰³ See Smith, Telang, and Zhang, *supra* note 75 at 2.

¹⁰⁴ 439 U.S. 522 (1979) (limiting the write down a taxpayer can take on inventory that is unlikely to sell quickly or at its current market price).

¹⁰⁵ Kevin O’Donnell, *How Thor Power Hammered Publishing*, <http://www.sfga.org/2005/01/how-thor-power-hammered-publishing/> (last visited, Jan. 28, 2014) (detailing the effect of *Thor Power* on the size of book publisher inventories).

Business model minimum print runs and tax law do little to explain why book publishers have not embraced eBooks as a low cost and convenient way to make their back catalogs available. Two prominent contract cases, however, may help explain why music publishers have been so active and book publishers relatively dormant.

In *Boosey & Hawkes Music Publishers, Ltd. v. The Walt Disney Company*,¹⁰⁶ the Second Circuit considered the claim that Disney had exceeded its license to use Stravinsky's *The Rite of Spring* in "one motion picture" when it converted *Fantasia* from film to video format in the early 1990's. At the time Stravinsky licensed the music to Disney (1939), the home video format did not exist, and Disney had not included broad language in the agreement claiming the right to exploit *Fantasia* and its soundtrack in non-film formats that might be developed in the future. According to Boosey, Stravinsky's licensee, "one motion picture" meant the movie could be exploited in the only format known at the time, acetate-based film produced for viewing in theaters. Despite the lack of any language about exploitation in as-yet-to-be-developed technologies, the Second Circuit found for Disney. Converting old music to new formats did not require the licensee to negotiate a new license with the copyright owner.

The case seems to apply to music publishers who want to convert vinyl albums into a digital format that can be sold on iTunes. A song on vinyl and a song on a CD or on iTunes are consumed in similar ways, and they are produced for the same purpose, although Boosey argued that the film format referenced in the Disney license anticipated a public viewing in a theater, whereas the new video technology enabled the quite different context of private home viewing. *Boosey* is very strong precedent that in the case of music, the conversion from vinyl or tape to .mp3 format does not require the renegotiation of a license with the copyright owner. Music publishers can proceed with the digitization of their back catalog without competing to re-sign

¹⁰⁶ 145 F.3d 481 (2d. Cir. 1998).

authors or hiring lawyers to renegotiate and write new contracts. Research has revealed no cases holding that music publishers must renegotiate in order to digitize their vinyl back catalogs.

The situation for book publishers is substantially the opposite. In the landmark case of *Random House v. Rosetta Books*,¹⁰⁷ the Second Circuit held that Random House had to renegotiate deals with its authors in order to publish their hard copy books in eBook format. Without Random House's permission, William Styron's estate and Kurt Vonnegut had granted permission to Rosetta to publish their works as eBooks. The contracts stated that Random House had the right to publish their works "in book form," but that phrase was not read to encompass the work in digital form. In other words, if Random House wanted to publish Vonnegut and Styron in eBook format, it would have to renegotiate the contracts in competition with other interested publishers. The decision, of course, relies heavily on the language of the individual contracts, but Random House's form book contracts were likely not substantially different from those used by other book publishers. In any event, research reveals no subsequent opinions freeing publishers from the need to find rights holders, compete, and negotiate in order to publish their eBooks.

Risk aversion is probably hindering the development of a rich body of case law on the issue of old media and new technological formats, but the story told by the divergent opinions in *Boosey* and *Rosetta Books* has significant power to explain why so many more old hit songs have found their way to iTunes than old books have found their way to the Kindle Store. If book publishers have to renegotiate and music publishers do not, then one would expect to see more digital versions of older music than digital version of older books.

3. Digitization Costs

¹⁰⁷ 150 F. Supp.2d 613 (S.D.N.Y. 2001), *aff'd* 283 F.3d 490 (2002).

Another advantage that the music industry may have is the lower cost of digitization. A vinyl album or audio master tape can be converted directly to a consumable digital form and be made available almost immediately. A book, on the other hand, can be scanned quite easily, but in order to be marketed as a professional-looking eBook (as opposed to a low quality, camera-like image of the original book), the scanned text needs to be manipulated with word processing software to reset the fonts and improve the appearance of the text.¹⁰⁸ And given that the best optical scanning software can leave glitches in the text, an eBook intended for mass market distribution should also be proofread for typographical errors.¹⁰⁹

4. Comparative Marketplace Efficiencies and Information Costs.

The market for older music may also be more robust than the market for older books. Songs seldom play for more than four minutes, and they can be previewed in part on iTunes and often listened to for free in their entirety on YouTube. They are easily discovered on the radio or on on-line services like Pandora and easily and quickly consumed, often for less than a dollar. In 2012-13, approximately consumers paid over \$16 billion to download songs from iTunes, while Amazon reported only \$4 billion in eBook sales.¹¹⁰ The comparative attractiveness and efficiencies present in the music marketplace may provide more of an incentive for music publishers to digitize, as opposed to book publishers.

CONCLUSION

Whatever the reasons for differences in the book and music publishing industries, the lack of availability of books from the post-1923 portion of the 20th century is startling. Senator Orrin Hatch argued in defense of the 1998 term extension that maintaining the availability and

¹⁰⁸ See *supra* note 68.

¹⁰⁹ <http://ocr-software-review.toptenreviews.com/>.

¹¹⁰ Market data available at www.statista.com (fee paid service).

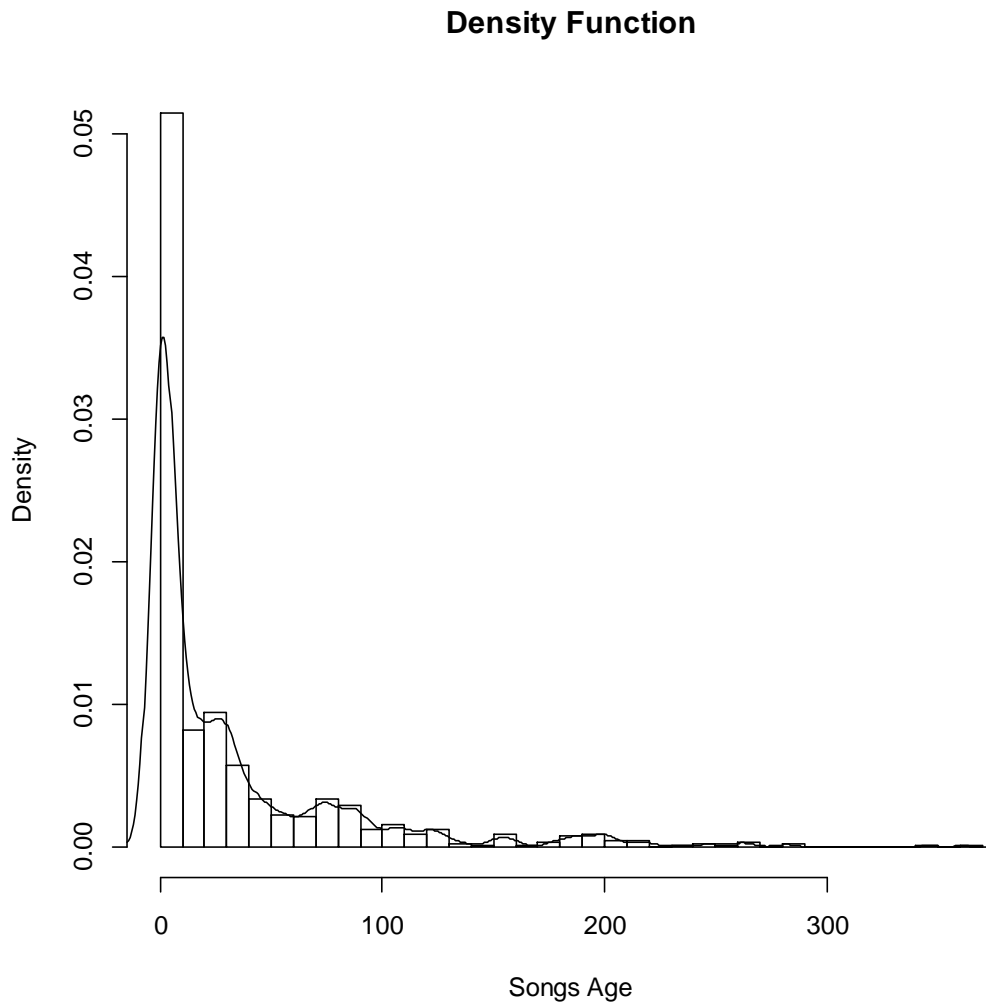
distribution of works is at the heart of the meaning of “progress” in the Copyright Clause of the Constitution.¹¹¹ He was absolutely correct about the purpose of copyright, but utterly wrong about how to solve the problem of missing works. Copyright term extensions have clearly prevented the development of a market for re-printing the massive number of “missing” works from the 20th century. If availability matters, then further attempts to extend the copyright term should be resisted, not encouraged. Copyright was not designed by the framers of the Constitution as a means by which Congress could make books disappear.

¹¹¹ See Orrin D. Hatch & Thomas R. Lee, “*To Promote the Progress of Science*”: *The Copyright Clause and Congress’ Power to Extend Copyrights*, 16 HARV. J. L. & TECH. 1, 7 (2002) (“the founding-era understanding of “progress” clearly extends to the dissemination or distribution of existing artistic works”). Cf. U.S. CONST., art. 1, §8, cl. 8 (“Congress shall have the power To . . . promote the Progress of Science and useful Arts, securing for limited Times to Authors and Inventors the exclusive Right to their respective Writings and Discoveries.”); L. RAY PATTERSON, COPYRIGHT IN HISTORICAL PERSPECTIVE (1968); (discussing historical understandings of the word “science”).

APPENDIX A: PUBLIC DOMAIN SONGS IN MOVIES

(Analysis of Peibei Shi, Ph.D, Statistical Consulting Office, University of Illinois)

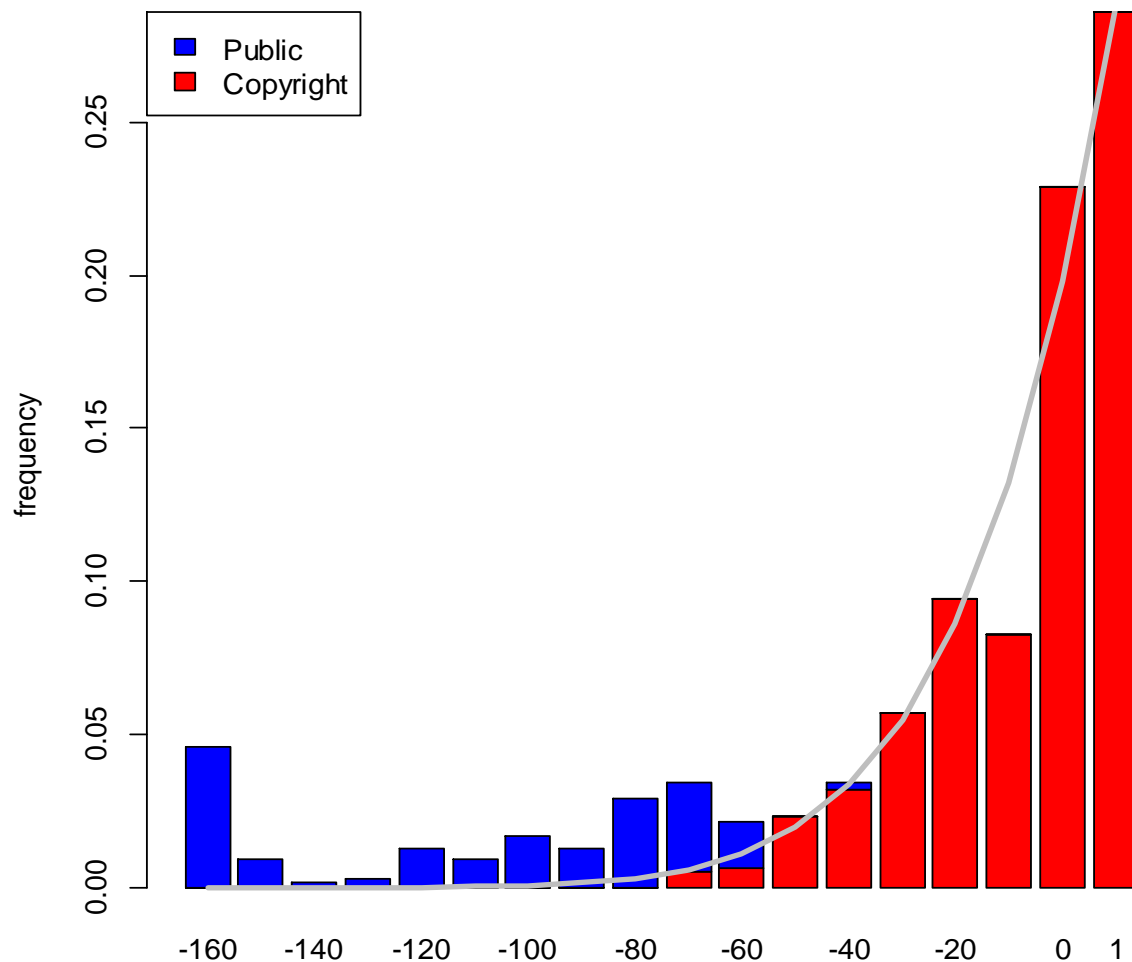
In order to determine whether the increase in public domain songs seen in Figure 5 is statistically significant, we first fit a density curve and test if there is a local mode around 60, and then fit a unimodal and bimodal curve, and use a likelihood ratio test.



We use a dip test to test for unimodality of the data, namely if the data comes from one model, or the mixture of two or more models. Employing Hartigans' dip test for unimodality, we see that $D = 0.0318$, $p\text{-value} = 1.202e-05$. Because the $p\text{-value}$ is quite small, we reject null hypothesis and conclude that the potential model is at least bimodal. The model comes with at

least two peaks and therefore indicates that the mere aging of songs over time is not adequate to explain the data.

We conduct further analysis based on Figure 5 and fit a non-linear least square curve to the bar (shown as the grey line) and use this curve to describe the “underlying” trend of use of music in movies (Grey line: frequency= a).



To test if the copyright status has an influence on the usage of the music in movies, we use the gray line value as a fitted frequency (if the decreasing trend is true) and look at the difference between fitted frequency and actual frequency (a/k/a residuals):

Group	Actual Frequency	Fitted Frequency	Residual
-160	0.0457	0	0.0457
-150	0.0089	0	0.0089
-140	0.0013	0	0.0013
-130	0.0025	0	0.0025
-120	0.0127	0.0001	0.0126
-110	0.0089	0.0002	0.0087
-100	0.0165	0.0006	0.016
-90	0.0127	0.0013	0.0114
-80	0.0292	0.0029	0.0263
-70	0.0343	0.0059	0.0284
-60	0.0216	0.011	0.0106
-50	0.0229	0.0197	0.0032
-40	0.0343	0.0334	0.0009
-30	0.0572	0.0546	0.0026
-20	0.094	0.0863	0.0078
-10	0.0826	0.1323	-0.0497
0	0.2287	0.1977	0.031
1	0.2859	0.2888	-0.0029

If the usage of music in movies follows the decreasing trend as music gets older, the residuals would be approximately half positive and half negative. But we can see the residual from range -160 and -20 are all positive, indicating that part of the data does not follow the assumed model.

We then take the residuals in range [-150, -40], 12 residuals in all, and conduct a test to see whether the residuals are significantly larger than 0 or not (residual at group -160 not included because the “-160” group is an accumulative group):

A sample t-test shows $t = 4.1612$, $df = 11$, and $p\text{-value} = 0.000793$, so we adopt the alternative hypothesis that the true mean is greater than 0. We then conduct a Wilcoxon signed rank test and find the $p\text{-value} = 0.0002441$. In addition, we adopt the alternative hypothesis where the true location is greater than 0. Both the t-test and non-parametric test indicate that the residuals are significantly greater than 0 providing evidence to conclude that the change of copyright status (from copyrighted to public domain) breaks the decreasing trend of music’s usage in movies.

APPENDIX B: SONGS WITH UNIQUE NAMES FROM 1913-32 ON ITUNES

YEAR	ITUNES	SONG TITLE	COMPOSER(S)
1913	y	When You're All Dressed Up and No Place to Go	Burt & Hein
1913	n	When You Play in the Game of Love	Goodwin & Piantadosi
1913	y	Where Did You Get That Girl?	Kalmar & Puck
1913	y	You're a Great Big Blue Eyed Baby	A. Seymour Brown
1913	n	You've Got Your Mother's Big Blue Eyes	Irving Berlin
1914	n	When You're Wearing the Ball and Chain	Smith & Herbert
1914	y	When You Wore a Tulip and I Wore a Big Red Rose	Mahoney & Wenrich
1914	y	Wein, Du Stadt meiner Traume	Rudolf Siczynski
1914	y	You're More Than the World to Me	Branen & Solman
1914	n	You Planted a Rose in the Garden of Love	Callahan & Ball
1915	n	We'll Have a Jubilee in My Old Kentucky Home	Goetz & Donaldson
1915	n	What a Wonderful Mother You'd Be	Goodwin & Piantadosi
1915	y	When I Leave the World Behind	Irving Berlin
1915	y	You'd Never Know the Old Home-Town of Mine	Johnson & Donaldson
1915	n	You'll Always Be the Same Sweet Girl	Sterling & Von Tilzer
1916	y	What Do You Want to Make Those Eyes at Me For?	McCarthy, Johnson & Monaco
1916	n	When the Black Sheep Returns to the Fold	Irving Berlin
1916	y	Where Did Robinson Crusoe Go with Friday on Saturday Night?	Lewis & Young; G. Meyer
1916	y	Yacka Hula Hickey Dula (Robinson Crusoe, Jr.)	Goetz, Young & Wendling
1916	y	You Can't Get Along with 'Em or Without 'Em	Clarke & Fisher
1917	y	When Yankee Doodle Learns to Parlez Vous Francais	Hart & Nelson
1917	y	Where the Black-Eyed Susans Grow (Robinson Crusoe, Jr.)	Radford & Whiting
1917	y	Where the Morning Glories Grow	Kahn & Egan; Whiting
1917	y	The White Peacock	Charles Tomlinson Griffes
1917	n	Whose Little Heart Are You Breaking Now?	Irving Berlin
1918	y	We Don't Want the Bacon-What We Want is a Piece of the Rhine	Carr; Russell & Havens
1918	y	When Alexander Takes His Ragtime Band to France	Bryan, Hess & Leslie
1918	n	When You Look Into the Heart of a Rose	Gillespie and Methven
1918	y	Why Do They All Take the Night Boat to Albany	Young & Lewis; Schwartz
1918	y	Would You Rather Be a Colonel With an Eagle on Your Shoulder	Mitchell & Gottler
1919	y	Wait Till You Get Them Up in the Air, Boys	Brown & Von Tilzer
1919	n	What'll We Do on a Saturday Night--When the Town Goes Dry	Harry Ruby
1919	y	The World is Waiting for the Sunrise	Lockhart & Seitz
1919	y	You're a Million Miles From Nowhere	Lewis & Young; Donaldson
1919	y	Your Eyes Have Told Me So	Kahn & Van Alstyne
1920	y	Who Ate Napoleons with Josephine When Bonaparte Was Away?	Alfred Bryan & Goetz
1920	n	The Wooing of the Violin (Some Colonel)	Smith & Herbert
1920	y	The Wreck of the "Julie Plante"	Drummond & O'Hara
1920	y	You Oughta See My Baby	Turk & Ahlert
1920	y	A Young Man's Fancy	Anderson; Yellen & Ager
1921	y	The Wang, Wang Blues	Mueller, Johnson & Busse
1921	y	When Big Profundo Sang Low "C"	Bohannon & Botsford
1921	y	When Buddha Smiles	Freed & Brown
1921	y	When Francis Dances with Me	Ryan & Violonsky
1921	y	When the Honeymoon Was Over	Fred Fisher
1922	n	Throw Me a Kiss	Hirsch, Buck, Stamper & Yvain
1922	y	Toot, Toot, Tootsie! (Bombo)	Kahn, Erdman & Russo

1922	y	Way Down Yonder in New Orleans	Creamer & Layton
1922	y	When the Leaves Come Tumbling Down	Richard Howard
1922	y	You Remind Me of My Mother (Little Nellie Kelly)	George M. Cohan
1923	n	Two Little Magpies	John Barnes Wells
1923	y	When It's Nighttime in Italy, It's Wednesday Over Here	Kendis & Brown
1923	y	When You Walked Out Someone Else Walked Right In	Irving Berlin
1923	y	Who'll Buy My Violets?	Goetz & Padilla
1923	y	Yes! We Have No Bananas	Silver & Cohn
1924	y	Ritual Fire Dance	Manuel De Falla
1924	n	Sometime You'll Wish Me Back Again	E. Austin Keith
1924	y	West of the Great Divide	Whiting & Ball
1924	y	When You and I Were Seventeen	Kahn & Rosoff
1924	y	Where the Lazy Daisies Grow	Cliff Friend
1925	y	That Certain Party	Kahn & Donaldson
1925	y	Ukelele Lady	Kahn & Whiting
1925	y	Waters of Perkiomen	Dubin & Klickmann
1925	y	Who Takes Care of the Caretaker's Daughter While the Caretaker's	Chick Endor
1925	y	Yes Sir, That's My Baby	Kahn & Donaldson
1926	y	Tamiami Trail	Friend & Santly
1926	y	A Tree in the Park (Peggy-Ann)	Hart & Rodgers
1926	y	When the Red, Red Robin Comes Bob, Bob, Bobbin' Along	Harry Woods
1926	y	Where'd You Get Those Eyes	Walter Donaldson
1926	y	Where do You Work-a John?	Weinberg, Marks & Warren
1927	y	There's a Cradle in Caroline	Lewis & Young; Ahlert
1927	n	There's Something Nice About Everyones	Terker & Bryan; Wendling
1927	y	Thou Swell (A Connecticut Yankee)	Hart & Rodgers
1927	y	The Varsity Drag (Good News)	De Sylva, Brown, & Henderson
1927	y	What Do We Do On a Dew-Dew-Dewy Day	Johnson, Tobias & Sherman
1928	y	Where is the Song of Songs for Me	Irving Berlin
1928	y	Where the Shy Little Violets Grow	Kahn & Warren
1928	y	You're the Cream in My Coffee (Hold Everything)	DeSylva, Brown & Henderson
1928	y	You Took Advantage of Me (Present Arms)	Hart & Rodgers
1928	y	You Wouldn't Fool Me (Follow Thru)	DeSylva, Brown & Henderson
1929	y	The Wedding of the Painted Doll	Freed & Brown
1929	y	When It's Springtime in the Rockies	Woolsey, Taggart & Sauer
1929	y	When the Organ Played at Twilight	Wallace; Campbell & Connelly
1929	y	You Don't Know Paree (Fifty Million Frenchmen)	Cole Porter
1929	y	Zigeuner (Bitter Sweet)	Noel Coward
1930	y	So Beats My Heart for You	Ballard, Henderson & Waring
1930	y	Swingin' in a Hammock	Seymour & O'Flynn; Wendling
1930	y	The Waltz You Saved for Me	Kahn; King & Flindt
1930	y	Would You Like to Take a Walk?	Dixon & Rose; Warren
1930	y	You Brought a New Kind of Love to Me	Fain, Kahal & Norman
1931	y	When Yuba Plays the Rumba on the Tuba	Herman Hupfeld
1931	y	Where the Blue of the Night Meets the Gold of the Day	Turk, Crosby & Ahlert
1931	y	You Didn't Have to Tell Me--I Knew it All the Time	Donaldson
1931	y	You Forgot Your Gloves (The Third Little Show)	Eliscu & Lehak
1931	y	You Try Somebody Else	DeSylva; Brown & Henderson
1932	y	Underneath the Harlem Moon	Gordon & Revel
1932	y	Willow Weep for Me	Ann Ronell
1932	y	Wintergreen for President	Gershwin and Gershwin
1932	y	You're an Old Smoothie	DeSylva, Whiting & Brown

