Review of Economic Research on Copyright Issues, 2012, vol. 9(2), pp. 31-54

MORE MUSIC IN MOVIES: WHAT BOX OFFICE DATA REVEALS ABOUT THE AVAILABILITY OF PUBLIC DOMAIN SONGS IN MOVIES FROM 1968-2008

PAUL J. HEALD, PEIBEI SHI, JEFFREY STOIBER AND QINGYAO ZHENG

Abstract. A previous empirical study suggested that as copyrighted songs transitioned into the public domain they were used just as frequently in movie soundtracks as when they were still legally protected. That study, however, did not account for the number people who viewed each movie in the theater. Since the debate over copyright term extension centers on the continuing "availability" of works as they fall into the public domain, a better measure of the availability of songs in movies would account for the relative box office success of the movies in which the songs appear. The present study collects box office data for hundreds of movies from 1968-2008 in which appeared hundreds of songs and concludes that public domain songs were heard by just as many people in movie theaters before and after they fell into the public domain.

1. Introduction

Advocates for extending the term of copyright for existing works have successfully convinced national legislatures that copyrighted works become less available after they fall into the public domain.[1] Indeed, if works disappear from public view or become less available to the public when protection ceases, then copyright's goal of advancing knowledge seems threatened and an argument can be made for retaining copyright status indefinitely, as suggested by Landes and Posner (2003). On the other hand, if works remain as available when they enter the public domain or increase in availability,

[1] See, for example, H.R. REP. No. 105-452, at 4 (1998) (House of Representatives report finding retroactive extension of copyright would "provide copyright owners generally with the incentive to restore older works and further disseminate them to the public.").

then the case for extension is weakened. Our data on the fate of music as it falls into the public domain complements other recent studies that suggest diminished availability is not a problem. The policy implications of these studies is clear – the burden of proving diminished availability should fall squarely on those seeking to justify a valuable expansion of their property rights. Mere assertions of empirically unfounded fears should no longer suffice to prompt legislative action.

Ideally, availability to the public would be measured in terms of purchases made by the public, but since sales data is kept privately, previous studies have adopted various proxies, such as in-print status for books or digitization rates for analog music. For example, the availability of public domain books has been measured by Heald (2008), who finds a positive public domain effect resulting in an increase in the number of editions and publishers once copyright protection ends. As regards sound recordings, Brooks (2005) finds that non-owners of copyrighted works have done a better job making old format sound recordings available in digital form, indicating that copyright owners are less efficient in effecting the digitization required to old music available to modern consumers. Policy-making based on the assumption that availability declines when works fall in to the public domain and no longer have an owner has yet to find an empirical basis. Collectively, the studies should shift the burden of proof to copyright owners.

In the precursor to the present study, Heald (2009) measured the frequency with which bestselling songs from 1909-23 appeared in movies from 1968-2007 before and after they fell into the public domain. Availability was measured in terms of movie appearances per song title without regard for the popularity of the movie in which it appeared. Using the behavior of copyrighted songs over the same period as a control group, Heald found no statistically significant difference between the number of uses of songs in movies before and after they fell into the public domain. But what if public domain songs tend to appear in obscure art films seen by only a small number of people while copyrighted songs tend to appear in blockbusters? Clearly, a better measure of availability in the music-in-movies context would account for the popularity of each movie in which a song appears. To measure popularity, the present study accounts for the number tickets sold, and therefore the number of viewers of each movie while it played in theatres. The data analyzed below, therefore, provides a more accurate sense of the "availability" of a song when it enters the public domain.

2. Data

In the United States, data collection was simplified because the works studied were subject to a standard length of protection, as opposed to one based on the life of the author. Prior to copyright term extension legislation passed in 1998, the copyright term was essentially 75 years for owners who properly filed for renewal. For example, works from 1915 fell into the public domain in 1992, and works from 1922 fell into the public domain in 1997. Works from 1923 and later have been prevented from falling into the public domain by the aforementioned 1998 legislation. Given these parameters, 601 bestselling songs 1909-1922 were selected from *Variety* magazine records complied by Matfield (1962) for analysis before and after they fell into the public domain. Each song was tracked in the www.imdb.com database of over 380,000 movies from 1968-2007, and every song appearance was recorded. Each song was observed during the time period that it was protected by copyright and during the time period that it was in the public domain. The gross box office receipts from http://boxofficemojo.com for each movie in which a song appeared was then divided by the average ticket price for its year of release, yielding a firm estimate of the number of tickets sold for that film, and therefore a firm estimate of how many moviegoers heard each song.

A set of 693 bestselling songs from the 1923-32 time period was also identified from *Variety* records in order to provide a control group for the analysis. None of these songs have fallen into the public domain, and they were tracked over the same period of time and in the same way as the public domain songs to account for possible anomalies in the IMDB database. For example, if the IMDB database contains proportionally more recent movies than older movies, all songs should show an uptick in absolute usage which might distort the picture for public domain songs if they were viewed in isolation.

The final dataset used for this project consisted of 257 songs which made 690 song appearances in a movie. Of those appearances, 137 were by a song when it was in the public domain, and 553 were copyrighted song while it was still protected by copyright. Many of these observations were of songs that appeared in multiple movies (257 songs appeared in at least one movie; 137 appeared in at least two movies; 87 appeared in at least three movies; and 60 appeared in at least four movies).

Alternative sources of data for song availability were considered. Tracking the radio airplay trends of songs after they fall into the public domain would certainly be of interest, but ASCAP, the major collecting society in the U.S., refused to provide data for the study. Likewise, historical data on sheet music sales is treated as confidential by music publishers. Since the Harry Fox Agency (HFA) provides some incomplete information on how often a song is "covered" by recording artists, we considered tracking cover statistics, but since the popularity of a song among consumers is likely the dominate reason why it is recorded multiple times, we hesitated to proceed in the face of such a significant confounding variable (we suggest in Part

5 that there is reason to believe that the songs from 1923-32 still under copyright have always been more popular). In addition, HFA public data is clearly incomplete. In the end, we found ourselves comfortable with the music-in-movies metric for two reasons. First, the choice of background music for a film is driven by multiple factors beyond consumer popularity (appropriateness for the setting, time period, a mood of the movie; linkage between story line and content of lyrics, etc.). The songs from 1923-32 would seem to have no inherent advantage among these considerations. Second, movie producers often must make an investment in recording the music chosen for a film. Although an existing recording maybe used, often an ensemble must be hired and a studio booked in order to exploit the work. Looking at music-in-movies, therefore, provided an opportunity to measure the extent that investors are willing to expend capital to create derivative works based on public domain material.

Finally, we should note that we did not check U.S. Copyright Office records to confirm the legal status of the songs from 1923-32 that we assumed to be protected by copyright. In order to remain protected, a song from that era had to be renewed by its owner in the 28th year after initial publication. U.S. copyright records before 1976 are not presently available on-line and must be searched manually (one prominent legal research firm charges $750 per title). Nevertheless, the vast majority of the bestselling songs from the 1923-32 period are, not surprisingly, owned by prominent music publishers who are repeat players in the game of registration and renewal. Their business models are built around consistent legal compliance, and we doubt more than a handful of the 693 songs from 1923-32 fell by inadvertence in the public domain due to failure to renew. For example, in a prior study of 173 bestselling novels from the same period, only 2 of the books had fallen into the public domain for failure to renew (Heald, 2008).

3. Methodology and Results: All Songs

As an initial pass through our data, we arranged the movies in order of box-office revenue. Then, for each movie, we calculate the average status of the songs contained in it by allocating to each public domain song the value 0 and to each song under copyright the value 1, and we calculate the average of these values for each movie. In this way, a movie that only has copyrighted songs will have an average status value of 1, a movie that only has public domain songs will have an average status value of 0, and if a movie contains both copyrighted and public domain songs then it will have an average status value between 0 and 1. An average status value closer to 1 implies that the movie is more dependent upon copyrighted rather than public domain songs, while the opposite is true when the average status value is close to 0. This generates the following table of frequencies of movies at each copyright status value:

Table 1: Frequencies of average copyright status

Av. status	0	0.22	0.33	0.5	0.62	0.67	0.73	0.8	0.92	1
Frequency	65	1	5	23	1	8	1	1	1	336

Clearly the data is highly bi-modal, with 76% of the movies having an average status of 1 (i.e. no public domain films), and about 15% of the movies having an average status of 0 (i.e. no copyrighted songs at all). That only leaves about 9% of all the movies with some songs of both status.

The interesting thing to consider then is whether or not it holds that the higher is the average song status variable, the higher is the box office revenue, in order to see if it holds that the more successful movies do indeed rely more heavily on copyrighted songs. The data is shown graphically in Figure 1, where again the bimodal nature of the data is easy to see. In the Figure, we also show the OLS regression line through the data.

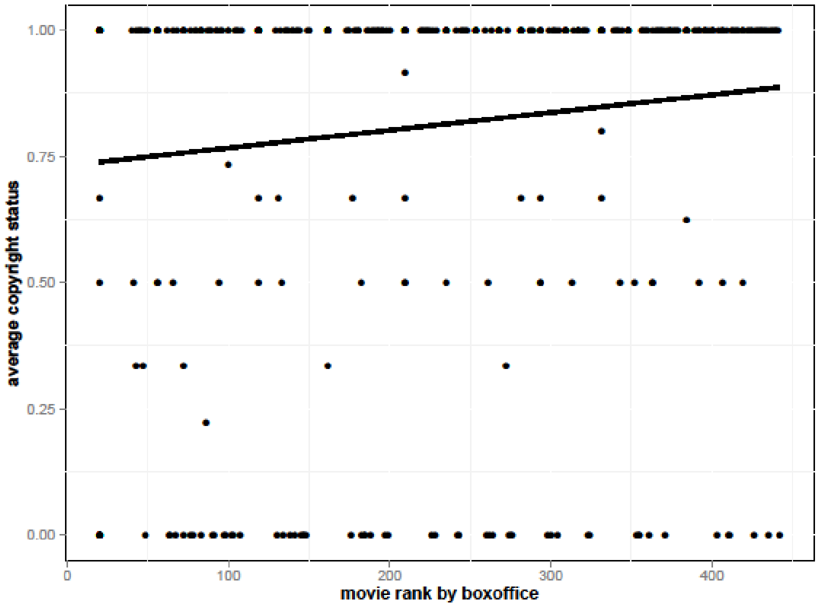

FIGURE 1. Average copyright status and box-office success

The intercept of the OLS regression is 0.7305, and the slope parameter is 0.0004.[2] Both are highly significant, however the adjusted R-square for the regression is very low at 0.01335. This low fit is of course due primarily to the bimodal nature of the data. However, as a first pass, this look at the data tells us that on average over all movies, the percentage of songs that are copyrighted is very high (between about 73% and 90%), and that number is greater for the more successful movies.

We now go on to attempt to explain the relationship between box office success and song status more formally. We initially approach the data by grouping movies by year, and looking at the sum of tickets sold as depicted in Figure 2.

One can see that copyrighted songs are seen in a movie by more people than public domain songs in every year except 1997. At first glance this

[2]Of course, the slope parameter is a very small number only because the horizontal axis is measured in the millions while the vertical is on a scale of 0 to 1. What is of essence is that the slope is positive and significant.

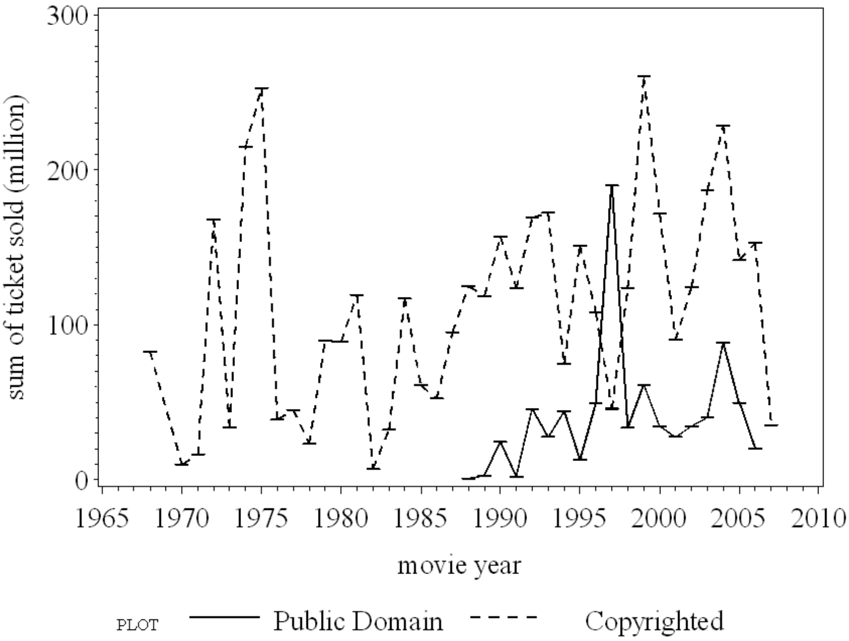

FIGURE 2. Sum of tickets sold by year

difference appears very pronounced, but the effect diminishes in light of the number of public domain songs available to be measured in each of those years, as seen in the next two figures. We start with Figure 3, which depicts in absolute terms the number copyrighted and public domain songs available to appear in movies in a given year.

This plot shows the legal status of songs in the dataset between 1968 and 2008. Between 1968 and 1983, all the songs are copyrighted. In 1984 (1909+75 years), the first songs begin to fall into the public domain. The number of public domain songs gradually increases over time, until 1998, when the copyright law changed. After 1997, due to the Copyright Term Extension Act, no new songs move into the public domain.

One can compare the total number of tickets sold in a year more fairly, therefore, if one scales the number of tickets by the number of available songs in that year. Since 257 songs, all copyrighted at the time, appeared in

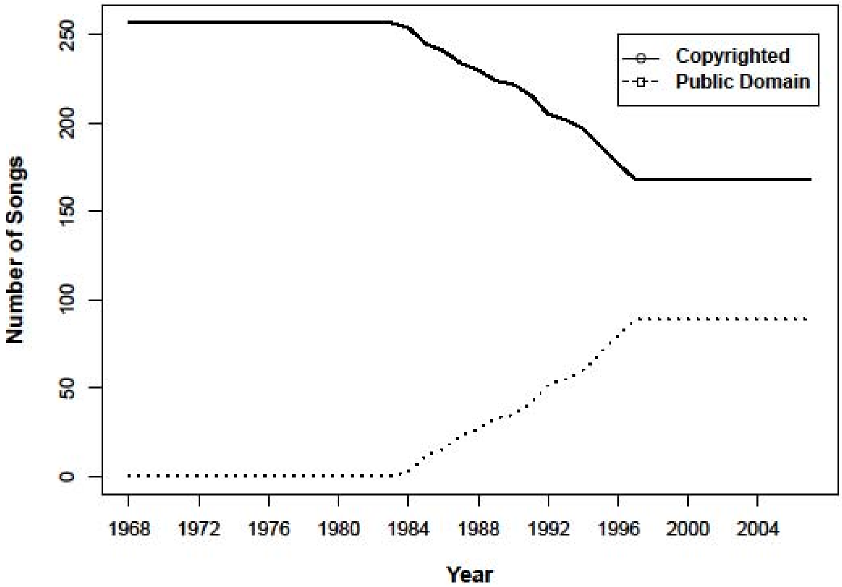

FIGURE 3. Status of songs by year

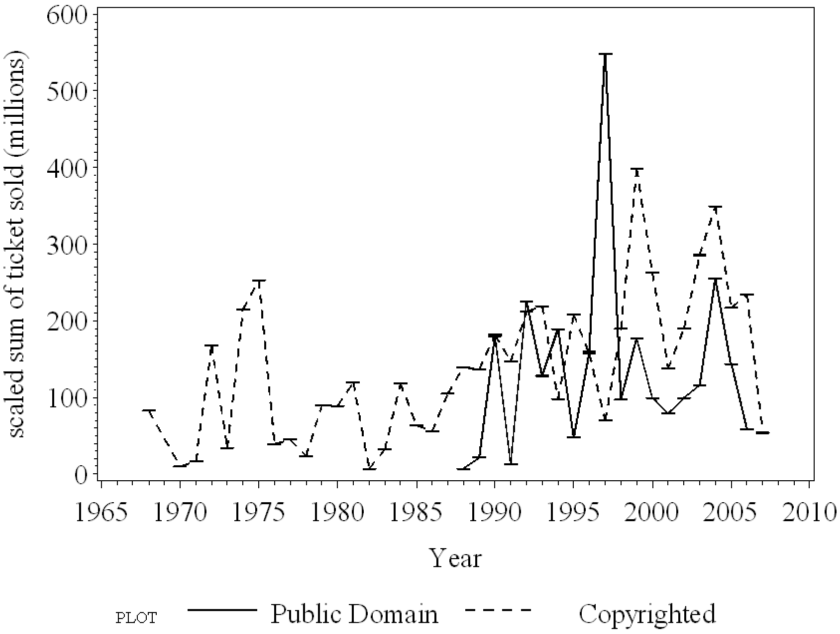

FIGURE 4. Scaled sum of tickets sold by year

movies from 1968-1983, this will be the baseline. Each year will have its sum of tickets sold multiplied by 257 divided by the number of songs available, to get an estimate of how many tickets might have been sold if more songs were available. For example: in 1990, around 50 songs were in the public domain. To compare the number of tickets sold in 1990 with other years, multiply by 257/50 which is approximately equal to 5. Performing this scaling adjustment for every year, for both copyrighted and public domain songs, yields the graph in Figure 4.

The lines are now much closer, without much difference between the total numbers of tickets sold, once the increased number of songs that are copyrighted are accounted for. This scaling method is useful for illustrating the problem caused by measuring a moving target like the legal status of songs, but it is not used in the model adopted later in the paper.

Looking at the mean of tickets sold really does not tell a complete story. One obvious problem is that number of tickets sold is heavily skewed. For this reason, a simple t-test is probably not appropriate. By taking the natural log of the data instead, a much more normal distribution is obtained, and this new variable, the log of tickets sold, can be the target of a simple linear regression model. Log of tickets sold is the response variable, and predictor variables are the copyright status (1=copyrighted, 0=public domain), year when the movie was released, year when the song was released, interaction between copyright status and movie year, and interaction between copyright status and song year.[3]

The copyrighted variable is negative, but not significant, meaning no significant difference between a copyrighted or public domain song. The interaction terms, which included copyright status, were also not significant.

---

[3]The interaction effect between two variables is indicated by *, e.g. Copyrighted*Movie year. The interaction effect in a linear model describes the simultaneous effect of two variables on response variable when the effect is not additive.

Table 2: Linear regression on log of tickets sold

Parameter	Coef. Estimate	Std. error	t-value	p-value
Intercept	173.3184	78.4361	2.21	0.0275
Copyrighted	-97.2150	83.4958	-1.16	0.2447
Movie year	-0.0470	0.0317	-1.48	0.1389
Song year	-0.0410	0.0310	-1.32	0.1867
Copyrighted*Movie year	0.0214	0.0324	0.66	0.5089
Copyrighted*Song year	0.0287	0.0345	0.83	0.4063

The first question one would like to answer is whether or not one sees an increasing trend of availability over time for songs in either song group. Here one can model the year-effect based on grouping data by adding up tickets sold in each movie year. Looking back at Figure 2, we can observe that in each song group, there is an increasing trend of sum of tickets sold as year goes by. To quantitatively confirm it, we fit a regression model to each song group respectively, with the following model:

$$\text{sum tickets sold} = \beta \times \text{movie year} + \varepsilon$$

We then find that the parameter estimates for public domain and copyrighted show that the public domain model results are not significant, the movie year coefficients for both models are positive. Therefore, one sees increasing usage over time in both song categories, which gets us closer to the ultimate question of whether copyrighted songs more available than public domain songs.

In order to make a fair comparison, one should use the mean of tickets sold for regression analysis, using mean-of-tickets-sold = sum-of-tickets-sold ÷ number-of-movies. But a graph of the mean of tickets sold, like the scaled sum graph in Figure 4, is erratic with no clear trend. To address this, one can fit a linear model as follows:

$$\text{mean ticket sold} = \beta_1 \times \text{copyright} + \beta_2 \times \text{movie year} + \varepsilon$$

In this model, the copyrighted coefficient is positive, but not significant. The R-square statistic is only 0.07, which indicates that the covariates in this model are not adequate to explain our response variable. One therefore needs to consider a more complex interaction model.

If one fits the interaction model to the data:

$$\text{mean ticket sold} =$$

$$\beta_1 \times \text{copyright} + \beta_2 \times \text{movie year} + \beta_3 \times \text{copyright*movie year} + \varepsilon$$

The interaction model turns out to fit better than the simple linear model. All variables in the interaction model are significant, with a larger R-square (about 0.15) than the simple linear model.

One should recall the assumption of the interaction model here. A model with interaction indicates that the regression lines for the two groups of songs have both different slope and intercept, and one can calculate the year in which the two regression lines cross each other. Dividing copyright coefficient by copyright*movie year coefficient, this particular year is around 2000, which tells us that before 2000, the availability for public domain songs was generally higher than copyrighted songs, and copyrighted songs became somewhat more available after 2000. Therefore, after considering both models above, one has enough evidence to conclude that public domain songs are, if not better than, at least as available as copyrighted songs, where availability is measured by the number of viewers hearing the song in a movie.

Although the models above try to reduce the variance between years by grouping tickets sold, the plot still fluctuates heavily, which might hinder observation of the real trend. Therefore, one might incorporate a smoothing

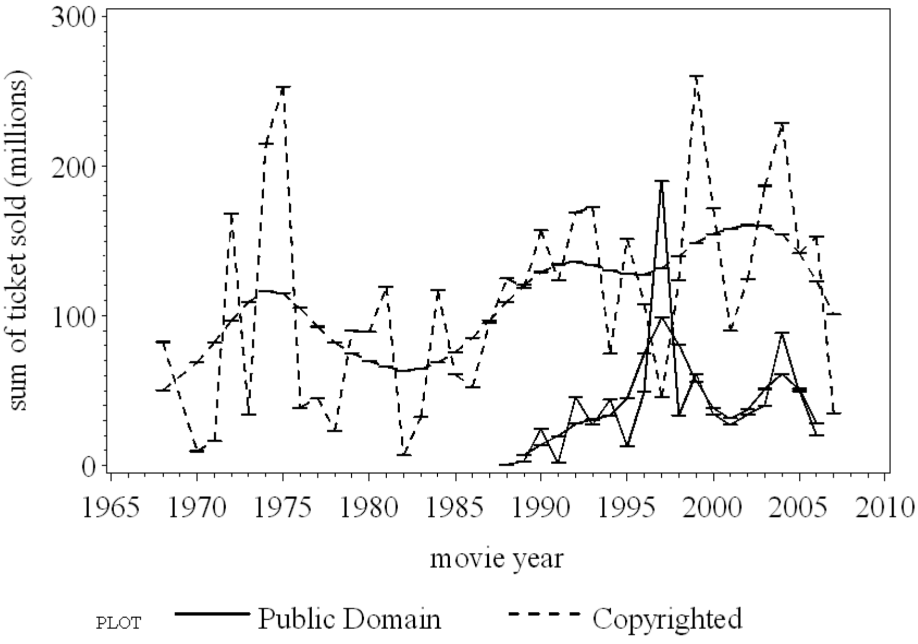

Figure 5. Sum of tickets smoothing curve

spline method to smooth the tickets sold curve and thus reduce variance with data. The models already considered are fitted using the "smoothed value," which will better illustrate the non-obvious trend behind the heavy fluctuations.

Two plots in Figures 5 and 6 are the smoothing sum of tickets sold and mean of tickets sold plots respectively. The dashed lines are the original lines, while the solid lines are the smoothed lines. Models targeting the smoothed mean of tickets sold both with and without interaction are shown in Figure 5.

In the simple model: smooth (mean ticket sold) $= \beta_1\times$ copyright $+\beta_2\times$ movie year $+\varepsilon$, the R-square value increases from .0731 to .2249. In the interaction model: smooth (mean ticket sold) $= \beta_1\times$ copyright $+\beta_2\times$movie year $+\beta_3\times$copyright*movie year $+\varepsilon$, the R-square value increases from .1475 to .3755. Comparing these two models, the latter one seems more objective.

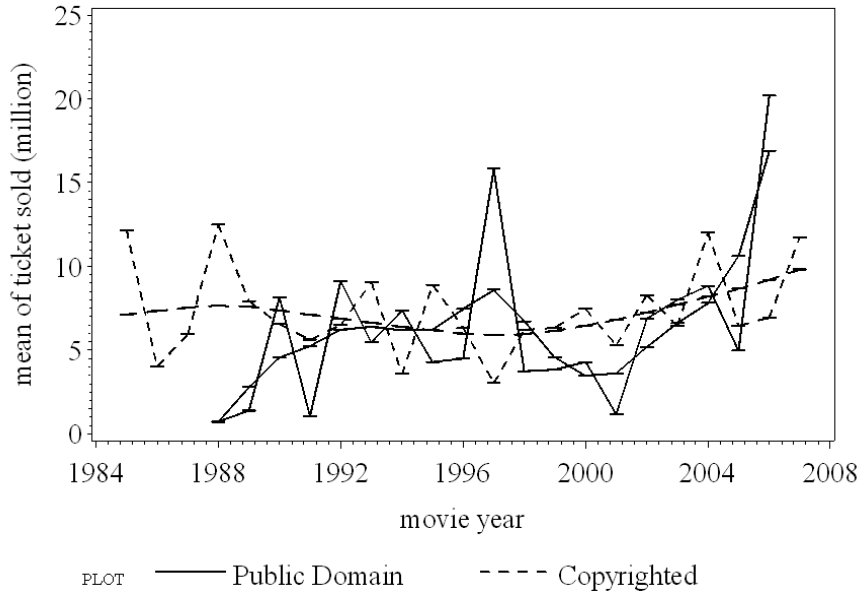

FIGURE 6. Mean tickets smoothing curve

Again, the two regression lines in interaction model cross at around 2000. We draw the same conclusion as with the previous analysis using original data, finding no difference between copyrighted songs and those in the public domain. This finding is consistent with Heald (2008) which analyzed song availability in terms of the number of appearances that the observed songs made in movies. Now, after accounting for the number of moviegoers attending those movies, we conclude that a change in legal status from copyright to public domain has no negative affect on a song's availability measured by number listeners, as opposed to number of movie appearances.

4. Methodology and Results: Songs in Four or More Movies

As a final exercise, we chose to break out the set of songs that movie directors found the most attractive, those songs that appeared in at least four movies from 1968-2007. Heald (2008) focused primary attention on this group of songs because in most cases the political and economic debate over copyright term extension has focused on the works that hold their value best over time. Even Landes and Posner (2003) state that works without any market value should not be protected by copyright. We view movie directors' choices as an objective indicator of enduring value. The new target ("high value songs") consists of 348 song appearances in movies, 65 by public domain songs and 283 by copyrighted songs. After an initial analysis, we will also consider whether within the group of high value songs there may relative differences in popularity between the songs from 1909-22 and those from 1923-32 as measured by Billboard Chart data. If one group of songs was significantly more popular when they were all protected by copyright, then we must isolate any comparative popularity effect unrelated to any change in legal status.

Moving directly to the regression where log tickets sold is dependent on copyright, movie year and song year, for these high value songs, we see in Table 3 that the copyright indicator turns out to have a positive effect on the amount of tickets sold. The small p-value suggests the effect it has is significant, even with the small sample size. All variables in the table have 1 degree of freedom.

Table 3: Linear regression of log of tickets sold, high value songs

Variable	Coef. Estimate	Std. error	t-value	p-value
Intercept	70.7545	59.2316	1.19	0.2332
Copyrighted	0.9105	0.4375	2.08	0.0382
Movie year	-0.0169	0.0131	-1.29	0.1969
Song year	-0.0189	0.0292	-0.65	0.5175

Regression summary statistics: Root MSE = 1.4359, R-Square = 0.0423, Dependent Mean = 1.2330, Adj R-Sq. = 0.0333, Coeff Var. = 116.4543.

However, the analysis gave us quite small R-square and Adj-R-square here, which means these three variables might not predict the behavior of log tickets sold very well. While this model would indicate that copyrighted songs are more popular, the poor model fit means that the model may not be reliable.

Next, the interaction model is adopted, but none of the predictors (movie year, copyright status, or song year) significantly predict ticket sales. The Interaction Model explains log tickets sold using as independent variables movie year, song year, copyrighted*movie year and copyrighted*song year. This is reported in Table 4.

Table 4: Linear regression, interaction model

Parameter	Coef. Estimate	Std. error	t-value	p-value
Intercept	228.3814	124.7017	1.83	0.0680
Copyrighted	-206.6870	142.7403	-1.45	0.1486
Movie year	-0.0443	0.0443	-1.00	0.3181
Song year	-0.0727	0.0558	-1.30	0.1938
Copyrighted*Movie year	0.0300	0.0464	0.65	0.5174
Copyrighted*Song year	0.0769	0.0656	1.17	0.2420

Summary statistics: R-Square = 0.0487, Coeff Var. = 116.4320, Root MSE = 1.4357, log tickets Mean = 1.2330

Performing the prior grouping analysis with only the songs that appear in four or more movies generates graphs that look similar to Figure 2 (see Figure 7). A rough increasing trend can be observed by year for both public domain songs and copyrighted songs, although the effect is not as obvious as that of full data. Obviously, the sum of tickets sold of copyrighted songs is higher than that of public domain songs because there are a larger number of them. Looking once again at mean tickets sold instead, one can see at some points public domain songs peak higher than the copyrighted songs.

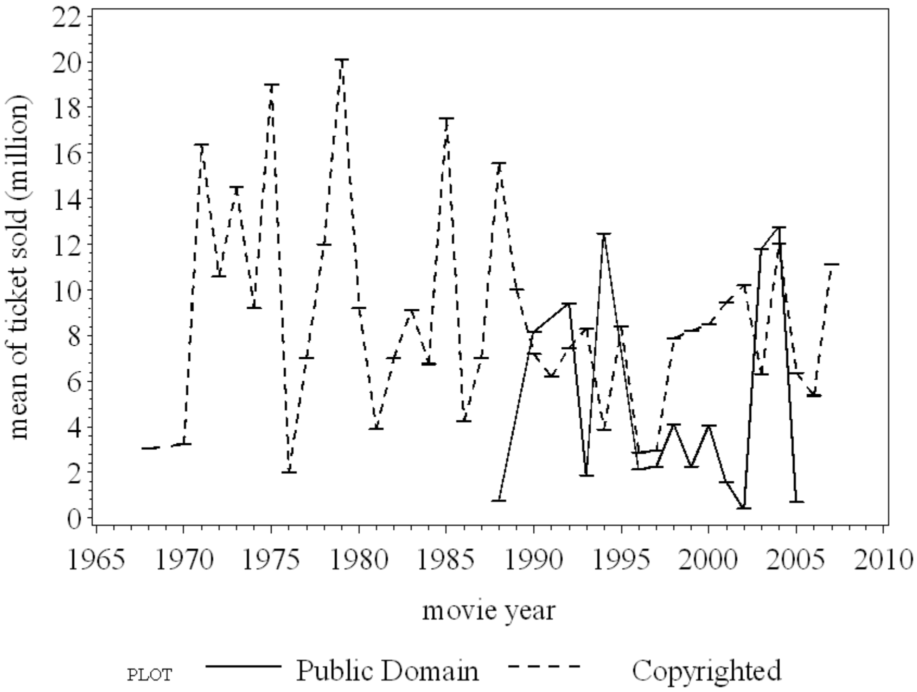

FIGURE 7. Mean tickets sold

Table 5: Linear regression, interaction model

Parameter	Coef. Estimate	Std. error	t-value	p-value
Intercept	-6.1789	416.5307	-0.01	0.9882
Copyrighted	123.0833	480.5996	0.26	0.7994
Movie year	0.0056	0.2085	0.03	0.9788
Copyrighted*Movie year	-0.0601	0.2407	-0.25	0.8042

Summary statistics: R-Square = 0.1365, Coeff Var. = 59.2128, Root MSE = 4.0742, mean tickets Mean = 6.8807

When one fits the model based on mean tickets sold, the parameter estimate (3.01712) is positive ($p = 0.03$), which again indicates that copyright songs are seen in movies by more people. The R-square and Adj R-square slightly improve, which means copyright and movie year predicts better the mean tickets sold.

The interaction model, in which mean tickets is explained by copyrighted, movie year and copyrighted*movie year, is less helpful (see Table 5). None of the predictors are significant, and therefore it is not an ideal model.

Overall, the analysis is not inconsistent with the conclusion that for the set of high value songs (those used in four or more movies) copyright status plays a weak positive role. However, several other variables should be considered before endorsing such a finding. Factors other than legal status are likely to influence whether a song is chosen for a movie or not. One relevant factor might be the age of the song. The copyrighted songs from 1923-32 are on average more than ten years "newer" than their counterparts from 1909-22 that fell into the public domain. More importantly, the relative popularity of the songs from each 10-year era might influence movie director's choices. A variable attempting to capture relative popularity of each song group is introduced below.

5. New Variable: Charting on Billboard

The mid-to-late 1920's produced some of the most popular music in American history, a phenomenon which may suggest that the songs from 1923-32 are inherently more popular than their counterparts from 1909-22 and therefore more attractive for inclusion in movie soundtracks. For example, among the famous song writers of the era, the complete data set includes 26 hits by Ira Gershwin (all after 1922), 10 hits by Cole Porter (9 after 1922), 50 hits by the prolific Bud de Sylva (42 after 1922), 22 hits by Rodgers and Hart (all after 1922), and 26 hits by Oscar Hammerstein (all after 1922). A less anecdotal suggestion of disparity in popularity between the two groups of songs is found in the number of the songs from 1909-22 that appeared in two or more movies while they were still protected by copyright. Only 20 of 835 songs from 1909-22 appeared in two or more movies while they were still copyrighted. Over the parallel time period (ending in 1997 when all of the songs from 1909-22 had fallen into the public domain), 75 of 694 songs from 1923-32 appeared in two movies or more, indicating they may simply have been more popular to begin with. Looking casually at data collected from *Billboard* magazine, the top 17 songs from 1909-22 (those that appeared in 4 or more movies overall), charted in the Top 100 an average of 4.05 times while they were still protected by copyright, while the top 17 songs from 1923-32 have charted in the Top 100 an average of 7.16 times.

We can use *Billboard* chart information more formally by looking at all the high value songs and measuring how many times each song showed up in *Billboard* Top 100 music chart. This new variable is denominated "board time." Board time data is discrete count data, which can be skewed distributionally, so one needs to check the scatter plot between the response variable and board time before putting it into model. We find the scatter plots of board time to be focused on the left side, so a transformation is needed, and since there are many zero values of board time, log-transformation here is no longer proper. A square root transformation for the board time variable called "sq board" improves the situation and will therefore be used in the model analysis.

Table 6: Linear regression on log of tickets

Variable	Coef. Estimate	Std. error	t-value	p-value
Intercept	101.1875	26.1923	3.86	0.0001
Copyrighted	0.7062	0.2194	3.22	0.0014
Movie year	-0.02670	0.0065	-4.17	<.0001
Song year	-0.0244	0.0136	-1.79	0.0743
sq board	0.1478	0.0478	3.09	0.0021

Summary statistics: R-Square = 0.0752, Adj R-Square = 0.0694, Coeff Var. = 111.3515, Root MSE = 1.3787, Dependent Mean = 1.2381

Whether the songs from 1922-32 are more popular can be approached using a simple model in which log tickets is explained using as independent variables copyright, movie year, song year and sq board. The results are reported in Table 6 (all variables have 1 degree of freedom).

The question of the popularity of the 1923-32 songs can also be approached by using a more complex interaction model, in which copyrighted is explained using as independent variables movie year, song year, sq board, copyrighted*movie year and copyrighted*song year. This is reported in Table 7 (again, all variables have 1 degree of freedom):

Table 7: Linear regression, interaction model

Variable	Coef. estimate	Std. error	t-value	p-value
Intercept	178.6144	77.9521	2.29	0.0223
Copyrighted	-86.5977	83.0333	-1.04	0.2974
Movie year	-0.0496	0.0315	-1.57	0.1160
Song year	-0.0411	0.0308	-1.33	0.1829
sq board	0.1459	0.0480	3.04	0.0025
Copyrighted*Movie year	0.0233	0.0322	0.72	0.4702
Copyrighted*Song year	0.0213	0.0344	0.62	0.5368

Summary statistics: R-Square = 0.076825, Coeff Var. = 111.4291, Root MSE = 1.379628, log tickets Mean = 1.238122

The R-squares for both models are not particularly high. First of all, the sq board variable for both models are significant, which indicates that it is a valid variable for model interpretation. The simple linear model shows a positive and significant coefficient for copyright variable. Therefore, the data tends to support the idea that public domain songs are less popular than copyrighted songs as measured by *Billboard* data. The interaction model does not support this conclusion, as copyrighted is not a significant variable. These conflicting results are similar to what the models showed without the *Billboard* time variable, and make it difficult to state confidently whether copyrighted or public domain songs are more popular.

Table 8: Linear regression, log of tickets for high value songs

Variable	Coef. estimate	Standard error	t-value	p-value
Intercept	70.1568	41.5539	1.69	0.0922
Copyrighted	0.9033	0.3499	2.58	0.0102
Movie year	-0.0175	0.0090	-1.95	0.0522
Song year	-0.0182	0.0218	-0.83	0.4057
sq board	0.1484	0.0726	2.04	0.0417

Summary statistics: R-Square = 0.0636, Adj R-Square = 0.0532, Coeff Var. = 107.1717, Root MSE = 1.3830, Dependent Mean = 1.2904

Table 9: Linear regression, interaction model, high value songs

Variable	Coef. estimate	Std. error	t-value	p-value
Intercept	259.4708	120.8064	2.15	0.0324
Copyrighted	-216.0828	128.7944	-1.68	0.0943
Movie year	-0.0471	0.0426	-1.10	0.2699
Song year	-0.0861	0.0540	-1.59	0.1119
sq board	0.1606	0.0729	2.20	0.0282
Copyrighted*Movie year	0.0302	0.0436	0.69	0.4892
Copyrighted*Song year	0.0817	0.0589	1.39	0.1665

Summary statistics: R-Square = 0.071049, Coeff Var. = 107.0416, Root MSE = 1.381290, log tickets Mean = 1.290423

The *Billboard* variable may nonetheless be added to the models used earlier to measure the effect of public domain status on the availability of the sub-set of songs that appeared in four or more movies. We now compare the results of incorporating the *Billboard* variable into the simple model and the interaction model.

First, the simple model in which log tickets is explained using as the independent variables copyrighted, movie year, song year and sq board. The results of the linear regression of this model are displayed in Table 8. And second, the interaction model, in which log tickets is explained using as independent variables copyrighted, movie year, song year, sq board, copyrighted*movie year, and copyrighted*song year (Table 9). In both cases, all variables have one degree of freedom.

The simple model shows all variables except song year to be significant. The coefficient for copyrighted is positive, meaning the copyrighted movies may have more availability. The interaction model may be more appropriate however. That was the model that seemed to fit the best on the full dataset, and in the interaction model copyrighted is not a significant variable. Therefore, after adding the *Billboard* chart information, it is fairly clear that copyrighted songs are not more available than public domain songs. The initial finding of a weak positive effect of copyright law on availability cannot be endorsed.

6. Conclusions

For both copyrighted songs and public domain songs, availability measured by ticket sales of movies using the songs increases by year. The lower availability of songs in the public domain in absolute terms is due to the fact that there are fewer of them to be measured at significant times. As illustrated graphically, copyrighted songs and public domain songs behave differently. However, the evidence does not show that availability is significantly influenced by the legal status of a song. Although we lack of a perfectly satisfactory model that predicts outcomes consistently, we prefer the interaction model. The difference between the standard regression model and the interaction model is that the standard model assumes that copyrighted and public domain songs have the same slope and difference intercepts, while the interaction model assumes both the slope and intercept might be different. We prefer the interaction model to the standard regression model based on r-squared, which is a measure of how well the model fits the data.[4] Neither model finds a statistically significant difference between copyrighted and public domain songs. When analyzing songs that appear in at least 4 movies, one can discern a weak indication that copyrighted songs are more popular, but this effect mostly disappears after accounting for the popularity of the songs themselves by looking at chart data from *Billboard*.

In general, the inclusion of box office data and *Billboard* data into the analysis of song availability supports the conclusion reached by Heald (2009) that songs remain just as available to the public after they fall into the public domain. Since copyright term extension incurs significant costs on consumers, lobbyists advocating for added protection bear a heavy burden of showing the negative consequences of works falling into the public domain. So far, no empirical work supports the primary rationale for term extensions – that public domain works are less available to the public – while this study, along with studies by Brooks (2005) and Heald (2008; 2009), support the opposite conclusion.

References

Brooks, T. (2005), Survey of Reissues of U.S. Sound Recordings, available on http://www.clir.org/pubs/reports/pub133/contents.html.

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

Heald, P. (2009), "Bestselling Musical Compositions (1913-32) and Their Use in Cinema (1968-2007", *Review of Economic Research on Copyright Issues*, 6; 31-60.

[4] If we have reservations about the models, it is due to the fact that the r-squares are pretty low overall; meaning even our best model does not predict strongly. However, this should not pose a significant problem because the data is so volatile that it would be very difficult to achieve a higher r-squared. The unreliability (low r-square) is much more pronounced in the 4+ models than the overall models.

Heald, P. (2008), "Property Rights and Efficient Exploitation of Copyrighted Works: An Empirical Analysis of Public Domain and Copyrighted Fiction Bestsellers", *Minnesota Law Review*, 92; 1031-63.

Mattfield, J. (1962), *Variety Music Calvacade: 1620-1961. A Chronology of Vocal and Instrumental Music Popular in the United States*, New Jersey: Prentice-Hall.

Paul Heald is Professor of Law, University of Illinois, and Professorial Fellow, Centre for Intellectual Property Policy and Management, Bournemouth University. Piebei Shi, Jeffrey Stoiber and Qingyao Zheng are degree candidates in the Statistics Department of the University of Illinois.