

Liebowitz, Stan J.

Working Paper

A replication of four quasi-experiments and three facts from 'The effect of file sharing on record sales: an empirical analysis' (Journal of Political Economy, 2007)

Economics Discussion Papers, No. 2016-48

Provided in Cooperation with:

Kiel Institute for the World Economy – Leibniz Center for Research on Global Economic Challenges

Suggested Citation: Liebowitz, Stan J. (2016) : A replication of four quasi-experiments and three facts from 'The effect of file sharing on record sales: an empirical analysis' (Journal of Political Economy, 2007), Economics Discussion Papers, No. 2016-48, Kiel Institute for the World Economy (IfW), Kiel

This Version is available at:

<https://hdl.handle.net/10419/148420>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

Discussion Paper

No. 2016-48 | December 05, 2016 | <http://www.economics-ejournal.org/economics/discussionpapers/2016-48>

A Replication of Four Quasi-Experiments and Three Facts from ‘The Effect of File Sharing on Record Sales: An Empirical Analysis’ (Journal of Political Economy, 2007)

Stan J. Liebowitz

Abstract

The influential piracy paper by Professors Oberholzer-Gee and Strumpf, although mainly based on proprietary data, contained an “important complement” to the main results, consisting of four “quasi-experiments” using publicly available data. This replication examines all of these quasi-experiments, first, by narrowly using identical data and statistical methods, as well as in a broader sense by extending or augmenting the data or methods. This study concludes that none of the four quasi-experiments provide evidence in support of OS' hypothesis that file-sharing has not harmed record sales.

Note of the Editor: The findings should be viewed as tentative until the paper has completed the review process and been published as an article.

(Submitted as [Replication Study](#))

JEL Z1 O3 L8

Keywords Piracy; File Sharing; Oberholzer-Gee; Strumpf; Sound Recordings; Music; iTunes; Napster

Authors

Stan J. Liebowitz, ✉ School of Management, University of Texas at Dallas, USA,
liebowit@utdallas.edu

The author thanks the Center for the Analysis of Property Rights and Innovation for Financial Support.

Citation Stan J. Liebowitz (2016). A Replication of Four Quasi-Experiments and Three Facts from ‘The Effect of File Sharing on Record Sales: An Empirical Analysis’ (Journal of Political Economy, 2007). Economics Discussion Papers, No 2016-48, Kiel Institute for the World Economy. <http://www.economics-ejournal.org/economics/discussionpapers/2016-48>

Introduction

One of, if not the most influential article on the effects of digital piracy (file-sharing) on the sound recording industry, appeared as a lead article in the *Journal of Political Economy* in 2007.¹ Its authors, Felix Oberholzer-Gee and Koleman Strumpf (OS), using data obtained from a small pirate server, concluded that piracy had essentially no impact on record sales. Although I have elsewhere (Liebowitz, 2016b) discussed issues with their main data set and analyses, direct replication of their main regressions is not possible because OS never made their main data public.²

Nevertheless, OS used publicly available data when conducting four “quasi-experiments” described as “an important complement” to their main analysis. In this replication I examine these quasi-experiments and also several pieces of data that OS use to support their analysis. Note that additional details and results can be found in the Appendix, to which I will sometimes refer.

1. Does American Piracy Decrease Every Summer?

Here are OS describing their first quasi-experiment:

The first experiment involves variation over time. The number of filesharing users in the United States drops 12 percent over the summer (estimated from BigChampagne [2006]) because college students are away from their high-speed campus Internet connections. If downloads crowd out sales, we should observe that

¹ According to Google Scholar, Oberholzer-Gee and Strumpf (2007) has received more citations (1004 as of September 29th, 2016) than any other paper on piracy in the economics literature, of which I am aware. According to Web of Science (in May 2016), it was the fifth most cited paper that the *Journal of Political Economy* published in 2007.

² At a public forum in 2004, Koleman Strumpf stated they would make their data set available “soon” [<http://web.archive.org/web/20040515202021/http://cdbaby.net/fom/000004.html>, search for “soon”]. Four years after this statement, OS stated that they had signed a non-disclosure agreement, although they did not provide this agreement to reporters who asked to see it. See Glenn (2008) or Häring (2008).

the share of albums sold in the summer increases following the advent of file sharing. [2007, page 36]

OS purport to demonstrate that American piracy fell each summer, supposedly because American college students lost access to their high speed Internet connections when they went home for the summer.³ Thus the effects of piracy should have been weaker in the summer than during the rest of the year. OS then compared the yearly summer shares of record sales for four pre-Napster years and seven post-Napster, expecting the summer share of record sales to be higher in the post-Napster years if piracy hurts sales, everything else equal.

I would agree that this quasi-experiment would be a test of the piracy thesis *if piracy routinely fell in the summer*, as OS claim. Since a regular decline in summer piracy is the key requirement for this quasi-experiment, OS's examination of summer piracy is a worthy target for replication.

The BigChampagne data that OS use to measure piracy runs from August of 2002 to May of 2006. OS assume that what happens in the three complete summers during this 46-month period holds for all the summers during the full seven-year post-Napster period, because college students presumably went home every summer.

Figure 1 charts the monthly number of pirates using the BigChampagne data. The three complete summer periods in these data are denoted by circles around them.⁴ Visual inspection of Figure 1 would seem to indicate that only the first summer clearly reflects lower piracy than the non-summer months. Indeed, if one calculates piracy levels in the summer months relative to the other months in each of those three years, as was done in my prior replication (Liebowitz, 2007),

³ Note the contradiction between OS's claim here that American school holidays (which affect both secondary and college students) *reduce* American piracy when in their main regressions OS claim that German secondary school holidays *increase* German piracy. This claim also appears inconsistent with OS's main (first stage) regressions which imply that German summer school holidays would strongly *increase* American summer piracy.

⁴ American university summers tend to take place during June through August, although OS also include May and September. I use their 5-month definition of summer vacation in all the calculations below.

it is only the first summer that has a notable decline. The average summer piracy decline is 12%, matching the value reported by OS.⁵

Nevertheless, OS have explained, in response to the prior replication, that they were comparing the summer decline *relative to a trend*. They stated “[a]t a time of rapid growth in the number of file sharers – the number of US users doubled between January 2004 and January 2006 – these summer months represent clear breaks from the growth trend in this period.”⁶

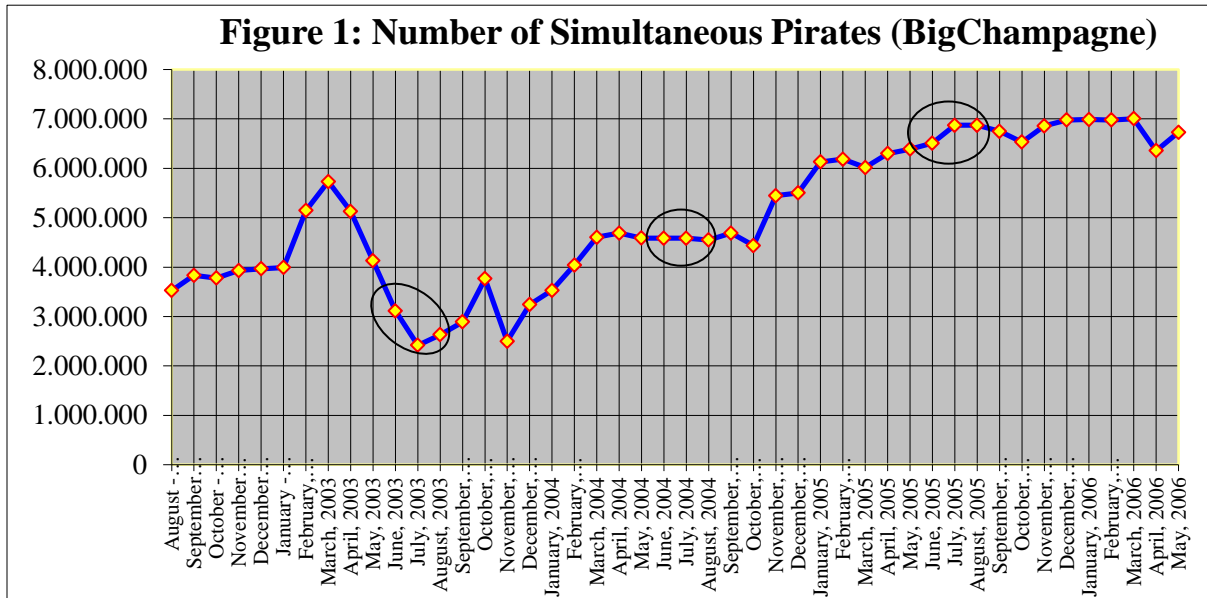


Figure 1 does not seem to comport any better with this latter explanation. It does not appear to reveal a “clear break” of an otherwise upward trend during the summer of 2005, since piracy in that summer is higher than in the previous winter/spring, and is basically the same as during the following winter. But let’s bring this claim more formally to the data.

In their response to my earlier replication, OS tell us that they regressed the number of pirates:

⁵ The first summer has a piracy drop of 38.8%, the second summer a drop of 0.2% and the third summer an increase of 3.8%.

⁶ This quote comes from a response by OS to an editor from an earlier submission of this replication.

in a specification which includes a time trend term and an indicator variable for summer. The regression results imply file sharing activity dropped 12% in the summer when we include a time trend and fell by 8% when we do not.

The singular term “indicator variable” suggests that they used a single dummy variable to represent all three complete summers. This seems strange, since using a single dummy would *not* allow an answer to the question of whether a summer piracy decline is a regular, routine occurrence *each* summer. Answering the latter question would require separate dummies for each summer.

| | | | | |
|--|-----------|-----------|-----------|-----------|
| Summer03 | | | -1.29e+06 | -2.12e+06 |
| | | | [-3.42] | [-5.67] |
| Summer04 | | | -611,668 | -556,702 |
| | | | [-4.14] | [-2.22] |
| Summer05 | | | 578,467 | 1.52e+06 |
| | | | [4.37] | [5.72] |
| Summer3yr | -449,687 | -384,482 | | |
| | [-1.82] | [-0.81] | | |
| Time Trend | 87,885 | | 74,084 | |
| | [10.8] | | [9.91] | |
| Constant | 3.11e+06 | 5.16e+06 | 3.43e+06 | 5.16e+06 |
| | [10.4] | [21.1] | [12.1] | [20.6] |
| Observations | 46 | 46 | 46 | 46 |
| Adjusted R-squared | 0.685 | -0.006 | 0.757 | 0.336 |
| Average # pirates in 21 non-summer months over 36 months | 5,082,597 | 5,082,597 | | |
| Average Summer Impact | -8.8% | -7.6% | | |
| Summer change 2003 | | | -30.6% | -50.3% |
| Summer change 2004 | | | -13.3% | -12.1% |
| Summer change 2005 | | | 9.0% | 23.6% |
| Average of 3 summers | | | -11.6% | -12.9% |

Nevertheless, it is easy to replicate the regression described above and the results (highlighted in red) are found in the first two data columns of Table 1 above. With the time trend included, the summer dummy is negative and of borderline statistical significance (8%), indicating that on

average piracy falls in the summer, but the average decline appears to be only 8.8%, not the 12% reported by OS.⁷

If we are going to test the claim that piracy declines in *each* summer due, say, to college vacations, it is necessary that we examine each summer individually. The two rightmost columns of Table 1 provide the coefficients from including separate yearly dummies (all statistically significant) for each of the three summers (highlighted in blue).

The results of these regressions confirm the intuition from Figure 1. The key finding is that the summer of 2005 has a significant *increase* in piracy. This runs counter to OS's claim that piracy routinely *falls* in each summer, thus destroying the logic of this quasi-experiment. Interestingly, the average decline is the same as was originally found just by taking the average three summer decline.

But note, also, that the 2003 decline might be due, at least in part, to special circumstances. In June of 2003 the organization representing American record companies, the RIAA, officially announced that it was gathering evidence for future lawsuits against individual pirates, and in September it began to bring suit against individuals. Bhattacharjee et al. (2006) demonstrate that this threat of lawsuits reduced piracy. Therefore, we cannot even be sure that the 2003 summer piracy decline was not partially or completely an artifact of *sui generis* piracy politics.

Finally, OS have argued, in response to an earlier version of this replication, that another data set exists which indicates that American piracy falls in the summer: “[a] detailed inspection of

⁷ There are various defensible methods that can be used to calculate the percentage summer piracy decline, so it is possible that some other method might match OS's stated 12% value, although none of three or four other methods I used provided a match. My calculation for the single summer coefficient in the table divides it by the average number of downloads during non-summer months for the three years with summers, although the full 46-month period was used for the regression. For the specification with three summer dummies, I divide each coefficient by the average piracy values for the 7 non-summer months in the same year, and average those three values.

packets on Internet2 shows that there are large dips in [piracy] traffic during the summers of 2003, 2004, and 2005.” The problem with this claim is that Internet2 traffic is unsuitable to be a measure of how piracy fluctuates during the year.

According to Internet2’s website, “Internet2 operates the nation’s largest and fastest, coast-to-coast research and education network.” OS are clearly aware of this since they define Internet2, in their 2009 article (p. 28), as “the U.S. high-speed network which primarily connects universities.” But American university activities slow during the summer, with most students and many faculty members on vacation. Therefore, Internet2 usage, particularly piracy, declines in the summer for idiosyncratic reasons having nothing to do with overall changes in piracy. Internet2 piracy statistics, therefore, cannot legitimately be used to resurrect their quasi-experiment from the failed replication using the same BigChampagne data that they used.

2. East Coast Sales versus West Coast Sales

OS argue that due to the time zone differences, East Coast Americans have greater access to European pirate-files than do West Coast Americans. The supposed reason for this is that prime-time American East Coast pirating occurs between 1am and 5am in Europe, whereas prime-time West Coast pirating takes place between 4am and 8am in Europe. OS take this to mean that Europeans are more likely to be sleeping, and thus have their computers off, during prime piracy times on the American West Coast compared to the East Coast. OS argue that if piracy harmed record sales, the East Coast should suffer a larger sales decline than the West Coast because East Coast users have more access to European files.

In their words:

A second experiment considers spatial variation...Because of time zone differences, such [piracy] transfers are easier for East rather than West Coast users. [2007, page 36]

OS present some statistics that they take as supporting their conclusion that piracy has not harmed East Coast sales relative to West Coast sales. Here is their empirical summation:

In 1998, the last year in the pre-P2P [pre-piracy] period, the share of album sales in the eastern time zone was 43.9 percent. This share has hardly moved since then. In 1999–2002, the mean was 43.5 percent and the range was 42.7–44.0 percent. This is consistent with some common national factors, rather than file sharing, driving sales trends.

OS note that the share of East Coast sales has remained in a narrow range and conclude that this small variation somehow demonstrates that sales in the East did not fall relative to sales in the West. OS do not compare actual market share changes in these two (out of four) U.S. time zones. Surely, to test their hypothesis, it would seem to matter whether the shares fell on the East coast relative to the West Coast after piracy became popular in 2000, since that is what the experiment claims to test.

Using the same Nielsen SoundScan ‘album-sales-by-city’ data for the same years as OS, plus an additional year’s data, 2003, allows a comparison of yearly East Coast and West Coast market shares. After examining the data, I find only small changes in East Coast market shares in the pre-Napster and post-Napster periods, just as OS did, so in that sense the replication largely confirms their claim about East Coast market shares hardly changing.⁸

But I also examined and compared the changes in sales for the East and West Coasts before and after piracy’s birth, which is shown in Table 2 (with the raw data in the Appendix), for various starting and ending years.

| | East | West | Extra East Decline |
|-----------|---------|---------|--------------------|
| 1998-2002 | -8.92% | 0.83% | -9.75% |
| 1999-2002 | -11.86% | -2.40% | -9.46% |
| 1998-2003 | -12.53% | -2.59% | -9.94% |
| 1999-2003 | -15.35% | -5.71% | -9.64% |
| 2000-2003 | -18.88% | -11.11% | -7.77% |

⁸ In the appendix I provide the raw data.

When the data are arranged to actually provide an answer to the question this quasi-experiment posed, the results are contrary to OS's conclusion, and are entirely consistent with piracy being harmful to sales. The sales declines in the East are greater than the declines in the West,⁹ meaning that market shares rise in the West and set in the East.¹⁰ The decline in East Coast sales was almost ten percentage points greater than the decline in the West Coast (equivalently, a 70% larger decline), a surprisingly large number. The implication, of course, is that if the slight increase in piracy due to enhanced access to a sliver of non-sleeping Europeans caused a sales decline of ten percentage points in the East (relative to the West), then the overall impact of piracy should be considerably larger.

This replication fails to support OS's conclusion.

I should note that I do not consider this to be evidence supporting a claim that piracy harms sales. I find the suggestion that a variation in the small number of Europeans awake late at night could measurably influence piracy behavior of Americans to be sufficiently farfetched to attribute these results to some other unknown factor.

3. The relationship between monthly changes in file-sharing and record sales.

The next quasi-experiment is a simple regression relating American record sales to the number of American pirates, with monthly fixed effects, over a 46-month period. In their words:

⁹ Although not shown, the market share of the combined middle time zones rose relative to the East but fell relative to the West, also consistent with the hypothesis that OS reject. Calculation of correct market shares for the middle time zones was not possible for OS given the way they treated a large category ("others") of unallocated sales that could not be fit into DMAs. OS put those sales into the middle time zones, which is an arbitrary procedure that obviously would incorrectly overstate the market shares for the middle time zones. I removed those "others" sales from the analysis, which allows the calculation of a correct market share for the middle time zones while not changing the shares of East and West Coast DMAs relative to each other.

¹⁰ Market share in the East drops consecutively each year after 2000, from 52.1% to 50.73%. They rise in the West, each year after 2000, from 16.95% to 18.09%. See the appendix for details.

When we use monthly data from August 2002 to May 2006 (N=46) and define Sales and Users in millions (with respective sample means of 56.0 million and 5.0 million), the estimated [coefficient] = -0.427 with a robust standard error of 0.33. There is little evidence that growth in the number of users has had a statistically or economically significant effect on sales. [page 37]

Table 3 contains the reported OS results along with what should be an identical regression on identical data, and a regression that also includes a simple measure of the health of the economy (the full results with coefficients for monthly fixed effects is in the Appendix). The OS results are found in the first column and the second column contains a seemingly exact replication (the full regressions with monthly dummies are in the Appendix).

It is somewhat difficult to categorize this attempted replication. The coefficient in the replication is 48% larger than OS’s coefficient, although they should have been quantitatively identical.¹¹ The qualitative results are largely the same—a negative relationship between piracy and record sales that is not statistically significant. This direct replication fails, therefore, but it does not change the qualitative conclusion. [**The Appendix contains the raw data which in principle, could allow researchers to understand the cause of the different coefficients, if OS were to make their data available.]

| Table 3: Replicating Coefficients on Sales with Monthly Fixed Effects | | | |
|---|-------------|--------------------|------------------|
| | OS Reported | Direct Replication | Add Unemployment |
| Number of Pirates (Coefficient) | -0.43 | -0.638 | -2.54 |
| t-stat (robust) | -1.29 | -0.96 | -2.16 |
| Unemployment Rate | | | -6,326,399 |
| t-stat (robust) | | | -1.87 |
| Implied impact on sales, 2003-2005 | -76,134,826 | -113,720,825 | -452,886,321 |
| Share of decline due to piracy, relative to 2000 | -21.8% | -32.6% | -129.8% |

A very straightforward broadening of the replication would be to include some measure of the health of the economy. To this end, a monthly unemployment rate was included in the

¹¹ Stata, SPSS, and Excel each generated identical coefficients.

regression, seen in the last column of Table 3.¹² The coefficient on piracy quadruples in size while achieving statistical significance (4%). The unemployment rate is also of borderline significance (7%) with a negative sign, implying that record sales fall when the economy worsens, as we would expect.¹³

OS claim that the coefficient on the number of pirates is economically unimportant. To address this claim, the next to last row of Table 3 contains the predicted decline in record albums during the 2003-2005 time-frame, and the last row calculates the predicted decline's share of the overall decline that had occurred during the 2003-2005 period relative to the peak sales in 2000, just as Napster was ramping up.¹⁴ The OS result implies that somewhat more than one fifth of the actual sales reduction was due to piracy and the directly replicated results imply that approximately one third of the actual sales reduction was due to piracy. When the unemployment rate is added, the entire decline in album sales could be attributed to piracy,¹⁵ a result that happens to match what the majority of the economic studies have found about the impact of piracy on record sales (Liebowitz, 2016a). I do not believe that declines of this size are economically unimportant.

¹² These values come from the U.S. Bureau of Labor Statistics. I used the non-seasonally adjusted values since the purpose is to measure the overall strength of the economy, and if it is always a little stronger during the Christmas season we would want to include that extra strength. Using the seasonally adjusted values would slightly increase the absolute value of the coefficients and t-stats for both the number of pirates and unemployment rate.

¹³ OS have argued, in their response to an earlier version of this replication, that multicollinearity between the number of pirates and the unemployment rate makes these coefficients unreliable (the correlation is -.68). Their concern has some merit, but the VIFs are 5.91 and 4.78 for unemployment and piracy and respectively. These VIFs are moderately high, but in a range normally thought not to be indicative of a serious multicollinearity problem (typically thought to require a VIF above 10).

¹⁴ The predicted decline in sales due to piracy is calculated as the product of the average number of pirates over this interval (2003-2005) and the piracy coefficient. In the appendix I provide these measurements separately for each year. The overall sales decline is based on Nielsen SoundScan average record sales over this period, compared to 2000, which was the peak year of sales using SoundScan statistics.

¹⁵ A value over 100% means that there would have been an increase in sales from 2000 to the 2003-5 period, except for the negative consequences of piracy.

In conclusion, a narrow replication was 48% larger than the coefficient reported by OS although it should have been identical, and a slightly broader replication indicated a considerably larger economic impact of piracy that is clearly inconsistent with their findings.

4. The relationship between genre piracy and genre sales

The final quasi-experiment examines whether genres of music that are most prone to piracy suffer larger sales declines than less piracy prone genres.¹⁶ Here is OS's explanation:

A third experiment, which also provides a test of the dropout hypothesis, is to see whether download [piracy] intensity influences long-run sales effect [sic] of file sharing growth after explicitly controlling for trends in [radio] music format popularity. [page 36]

OS found that genres with high piracy intensities were less likely to suffer sales declines than genres with low piracy intensities, although this result was statistically insignificant. They provided few details of the underlying analysis other than to say that they ran regressions on a genre's sales changes controlling for genre piracy intensity and changes in the genre's popularity as measured on radio.

These missing details are important because it appears to be impossible for OS to have correctly done what they claim to have done for the genres used throughout their paper. Throughout their paper they include 11 "genres" into which albums were classified when in fact several of the "genres" are merely categories of sound recordings and not musical genres.¹⁷ Four

¹⁶ See Liebowitz (2008) for a detailed examination of the impact of piracy on musical genres over a five-year period across 100 cities, where those genres with the greatest piracy intensity were also the genres that lost the most sales.

¹⁷ OS's record genres include: Catalog(ue), Alternative, Hard, Jazz, Latin, New, R&B, Rap, Current, Country, Soundtrack.

of those categories (*catalog, new artists, current hits, and soundtrack*)¹⁸ are based on sales volume, age of recording, experience of the artist, or whether the music was in a movie.

The reason it does not seem possible to correctly do what they claim to have done is that there are no radio station music genres that match these four record categories. For example, consider the record category “new artists.” A radio station may play new rap artists, or new country and western artists, or new rock artists, or even new classical or jazz artists, but these stations would also play established artists and there are no radio stations that play only new artists from every musical genre. Further, there are no radio genres that match the ‘new artists’ category (a list of radio genres is found in the Appendix). Nor are there radio stations playing only soundtracks of every type of music. Ditto for “current hits” and albums older than 18 months (catalog) of any style of music. Since there are no radio station genres that match these record categories, it does not seem possible to control for radio format popularity trends for these four “genres,” although that is what OS claim to have done. OS do not specifically mention whether they include these four genres in their genre regressions, although it appears that they did.¹⁹ Nor do they tell us what radio genres they “matched” to these four record categories that do not have similar radio genres.

It should be noted that even for actual musical genres, it is surprisingly difficult to match radio genres to sound recording genres.²⁰

¹⁸ As an example of the heterogeneity with any of these four genres, consider the soundtrack from the movie “Frozen” and the soundtrack from the movie “Straight Outta Compton.” Although both are successful soundtracks, one is a Disney movie for children and the other is a biography of the first gangsta rap group. There are no radio station genres which would play both albums.

¹⁹ It appears that the four genres are not all removed because OS state on page 37 that the mean value of pirate downloads per album *title*, across genres, is 61.2 (which, strangely, does not match the 57.1 value in their Table 3), which is very different than the average value of 35.9 if the four non-musical genres are removed, using the numbers in their Table 3.

²⁰ For example, one of the OS-chosen sales genres is “jazz” but although there is a radio “jazz” subcategory of New AC/Smooth Jazz, most listening measurements in that category are zero. On the other hand, “new adult contemporary”, which is also a subcategory of “NewAC/Smooth Jazz” appears to be the closest to jazz, although it

A separate issue is the OS measure of piracy intensity using their proprietary piracy data (the quality of which I have questioned elsewhere²¹). OS construct a variable intended to measure piracy intensity which they name “downloads per album” but this is apparently a ratio of the number of pirate downloads per album *title*, in a genre,²² not downloads per album *sold*. I create an alternative measure of piracy intensity (which I refer to as “OS corrected”), which is downloads per albums sold. The Appendix contains a discussion of why the latter is the superior variable, and also addresses how all the piracy intensity variables used here differ from more ideal measures. The correlation between the OS variable and my “corrected” version of this variable, for all OS genres, is only 6.5%, although it increases to 62.3% when the four non-music record categories are removed.

Putting these difficulties aside, we can try to replicate their regressions using the seven actual musical genres in their data. I broaden the replication by also examining the period 2000-2005, which appears to be a more appropriate time frame because file-sharing (beginning with Napster) did not become prominent until 2000²³ and because SoundScan’s measured sales of sound recordings peaked in 2000, making this period more representative of the post Napster regime of declining record sales. One further broadening of the replication is to use an additional source of

is not clear how close. Similarly, the OS “hard” genre (related to “metal”) is not a radio category, with the closest appearing to be the “new rock” subcategory under “Alternative” or the “active rock” subcategory of “Rock.” A complete list of American radio genres is in the Appendix.

²¹ See Liebowitz (2016b).

²² As seen in OS’s Tables 1 and 3, their album sales averages are more than a thousand-fold larger than their measure of downloads, implying that the mean value of downloads per sold album would be much less 1. Therefore, the average value that OS provide (61.2 downloads per album) implies that the number of album *titles* per genre is in the denominator.

²³ I provide, in the Appendix, details on Napster’s size that make clear that Napster was not economically important until the summer of 2000.

genre piracy intensity data, and I was able to acquire 2005 data on piracy intensity from the firm NPD.^{24,25}

Table 4 presents simple correlations between changes in genre sales and piracy intensity. In all cases, genres with higher piracy intensities experienced larger percentage declines in sales, and although the correlations are fairly large, due to having only seven observations the correlations would need have an absolute value greater than .75 to be statistically significant.²⁶

| Sales change | NPD | OS Corrected | OS Original |
|--------------|--------|--------------|-------------|
| 99-05 | -0.231 | -0.355 | -0.284 |
| 00-05 | -0.488 | -0.407 | -0.305 |

The actual replicated regressions linking piracy intensity and genre sales changes (controlling for radio genre audience changes) are found in Table 5. Higher piracy intensity is linked to lower sales in all six instances, although the coefficients are only statistically significant in one case (using NPD data during 2000-2005 period), which should not be surprising with only seven observations.

The results of the regressions can be summarized in an intuitive fashion, as found in the bottom two rows of Table 5, which calculates the change in genre sales as piracy levels increase under two hypotheticals. In the first hypothetical (Economic Impact I), this calculation is performed under the assumption that the piracy intensity level increases from zero to the highest piracy intensity genre (which differs for the three measures of genre piracy intensity).²⁷ The first

²⁴ NPD data do not necessarily reflect a representative sample of internet users because its web users voluntarily allowed themselves to be monitored. There is no reason to think that this problem would bias the relative amount of piracy *across genres*, however.

²⁵ NPD provided data for nine genres but we could only match six to the OS genres and seven to the SoundScan sales genres. OS's data did not include "classical" and NPD's data did not include "Latin."

²⁶ OS have wondered why the NPD correlations here are weaker than in my 2007 working paper. The answer is that in 2007 I equated their "hard" genre with NPD's "rock" but I later decided that NPD's "metal" genre was probably a better match.

²⁷ The highest piracy-intensity value for the original OS measurement of downloads per album title is associated with "alternative", for the corrected OS measure of downloads per unit sales it is "hard", and for NPD it is "rap".

hypothetical gives an idea of the loss (of revenue) that might occur as other genre's reach the same piracy-intensity as the most pirate-intense genre in the early 2000s. Five of the six values are unambiguously large and economically important, and the sixth is not trivial.

| Timeframe | 99-05 | 00-05 | 99-05 | 00-05 | 99-05 | 00-05 |
|----------------------------------|------------------|------------------|-----------------|-----------------|--------|--------|
| Genre Piracy Intensity | -0.078 | -0.072 | -0.12 | -0.10 | -0.033 | -0.100 |
| Robust T-stat | -0.73 | -0.86 | -0.85 | -0.9 | -1.09 | -2.9 |
| % Change in Genre Radio Audience | -0.089 | -0.13 | 0.15 | 0.095 | -0.15 | -0.20 |
| Robust T-stat | -0.45 | -0.72 | 0.33 | 0.22 | -2.7 | -3.02 |
| Observations | 7 | 7 | 7 | 7 | 7 | 7 |
| Adjusted R-squared | -0.36 | -0.31 | -0.36 | -0.24 | 0.19 | 0.41 |
| Measure of Piracy Intensity | OS normalized | OS normalized | OS corrected | OS corrected | NPD | NPD |
| Economic Impact I | -30.7% | -28.4% | -58.2% | -48.5% | -9.4% | -28.4% |
| Economic Impact II | -5.1% | -4.7% | -9.6% | -8.0% | -4.0% | -12.0% |

The second hypothetical calculates the expected loss when a genre, starting from zero piracy, achieves the average piracy rate for all genres.²⁸ The results from the second hypothetical, found in the bottom row of Table 5, are considerably smaller than the first hypothetical since it measures the piracy decline for the average genre. Although smaller, note that these average predicted declines are a fairly large portion of the actual decline in sales that took place from 2000-2005, which was 16.6%,²⁹ so the piracy-induced component would be an economically important portion of the decline.

The conclusion that the genres with the highest piracy rates suffered the largest sales losses is fairly consistent across specifications (although we cannot have a great deal of confidence in these results), revealing that it makes little difference which time period was used, whether the original OS piracy intensity numbers or the corrected OS piracy intensity numbers were used, or

²⁸ The average NPD value is only an approximation since we only have data for 67% of the albums sales. Also, the difference between average genre and maximum genre differs between the NPD data and the OS data.

²⁹ The equivalent value for 1999-2005 is 13.3%.

whether the NPD values were used. This is in contrast to OS, who found a positive (statistically insignificant) relationship.

I view these results as economically quite different than the OS genre results, even if the small number of observations makes it difficult to have normal levels of confidence in the results.

5. Data Replication

In this section we compare three of OS's data claims made in support of their thesis, against primary sources.³⁰

OS made two factual claims disputing whether record sales had experienced a serious decline (with the two claims noted by bracketed numbers):

there are several trends that are inconsistent with the view that P2P [piracy] now displaces sales on a large scale...[1] music sales have been flat or even rising in major markets with a quickly growing file-sharing population. For example, [2] in 2005 retail music sales rose in four of the five largest national markets. [page 39]

Because OS provide no reference for these factual claims (and the JPE apparently did not ask), we cannot directly source-check their factual claims but instead we must go to our own direct sources. The primary, perhaps only, source for international comparisons of record sales by year is the IFPI (International Federation of the Phonographic Industry).

With regard to the first claim about there being major markets with flat or rising sales, we first need to define "major." I would think that major national markets are presumably larger than Switzerland, which has a population similar to that of New York City. I think most economists would define rising sales as an increase in real revenues.

³⁰ In my earlier 2007 working paper I discussed other claimed facts that I believe to be incorrect, but the demonstration in some of those cases was quite lengthy, so I merely refer the reader to Liebowitz (2007).

Using IFPI data and looking at the top 10 markets (Switzerland is number 10), reveals that real retail sales rose in 0 of the top 10 markets during 2000-2005.³¹ The IFPI data also reveal that real retail sales rose in only 1 of the top 5 (or 10) markets between 2004 and 2005, not the 4 markets claimed by OS.³² The Appendix provides more details.

Later, OS claim:

in the United States the entire drop in 2005 album sales is due to losses at a single firm [Sony-BMG]...If file sharing were responsible for the observed sales decline in the United States, we would not expect this activity to affect the products of only a single firm. (page 40)

Again, OS provide no reference to support their statement. Data from an online news source,³³ combined with SoundScan data on U.S. album sales in 2004 and 2005, allow a reconstruction of unit sales for the four major record companies and the large group of independent record labels which we can treat as a virtual fifth record company. These calculations (found in the Appendix) reveal that 2005 sales fell at four of the five record groups (all except UMG), and that Sony-BMG's decline was only 44% of the aggregated declines in the industry, not the 100% implied by OS.

These data replications clearly fail.

6. Conclusion

None of the economic conclusions from OS's quasi-experiments hold up under replication. Several factual claims made about the industry also could not be replicated. It appears that the

³¹ Nominal retail sales rose in only 1 out of the top 10 markets, The same results hold whether the starting year is 1999 or 2001.

³² Nominal retail sales rose in 2 of the top 5 (or 3 of the top 10) markets.

³³ The source can be found here:

http://web.archive.org/web/20070117062246/http://www.undercover.com.au/news/2006/jan06/20060105_universals.html

evidence from this “important complement” to their main regressions does not support their main conclusions.

References

Bhattacharjee Sudip, Gopal Ram D., Lertwachara Kaveepan, and Marsden, James R. "Impact of Legal Threats on Music Sharing Activity: An Analysis of Music Industry Legal Actions" *Journal of Law and Economics*, 49:1 91-114.

Glenn, David. "Dispute Over the Economics of File Sharing Intensifies", *Chronicle of Higher Education*, July 17, 2008.

Häring, Nobert "The Economists' War about Downloading" [in German] *Handelsblatt*, March 3, 2008 <http://www.handelsblatt.com/politik/konjunktur/oekonomie/nachrichten/journal-of-political-economy-in-der-kritik-der-download-krieg-der-oekonomen/2929688.html>

Liebowitz, Stan J., How Reliable is the Oberholzer-Gee and Strumpf Paper on File-Sharing? (September 1, 2007). Available at SSRN: <http://ssrn.com/abstract=1014399>.

Liebowitz, Stan. J. "How much of the Decline in Sound Recording Sales is due to File-Sharing?" *Journal of Cultural Economics*, 40(1), Feb 2016a, pp. 13-28.
<http://link.springer.com/article/10.1007/s10824-014-9233-2>

Liebowitz, Stan. J. "Why the Oberholzer Strumpf File-Sharing Article is not Credible" *Econ Journal Watch*, 13(3) September 2016b, pp 373-396.
<https://econjwatch.org/articles/why-the-oberholzer-gee-strumpf-article-on-file-sharing-is-not-credible>

Oberholzer-Gee, Felix and Koleman Strumpf. "The Effect of File Sharing on Record Sales: An Empirical Analysis" *Journal of Political Economy*, 2007, 115:1 1-42.
<http://www.jstor.org/stable/10.1086/511995>

APPENDIX

1. EAST COAST AND WEST COAST

Calculating market shares of music albums by U.S. time zones merely requires putting the SoundScan cities (DMAs) into their appropriate time zones, although this is not as simple as just assigning DMAs to states.³⁴

The table in the text shows the change in sales over different periods of time. These are the yearly raw numbers, including the shares of the East and West Coasts.

| | East | West | East Share | West Share |
|------|---------|---------|------------|------------|
| 1998 | 328,917 | 105,287 | 51.96% | 16.63% |
| 1999 | 339,882 | 108,768 | 51.84% | 16.59% |
| 2000 | 354,671 | 115,373 | 52.10% | 16.95% |
| 2001 | 342,753 | 115,302 | 51.78% | 17.42% |
| 2002 | 299,574 | 106,158 | 50.96% | 18.06% |
| 2003 | 287,700 | 102,556 | 50.73% | 18.09% |

The middle time zones (Central and Mountain) had market share increases after 2000, but not as large as those in the West. This is also consistent with the hypothesis that OS reject.

2. MONTHLY PIRACY AND RECORD SALES

Here are the full regressions, including the monthly dummies, that lead to Table 3:

| Dependent Variable is monthly unit sales of albums (including singles) | | | | |
|--|----------|--------|-----------|--------|
| | Coef | T-stat | Coef | T-stat |
| # of Filesharers | -0.6378 | -0.96 | -2.54 | -2.16 |
| Unemployment Rate | | | -6326399 | -1.87 |
| December | 6.16e+07 | 11.9 | 5.67e+07 | 9.2 |
| November | 1.23e+07 | 3.54 | 7.64e+06 | 1.76 |
| October | 1.18e+06 | 0.26 | -4.58e+06 | -0.76 |

³⁴ There are 5 states where portions of the state are in either an Eastern or Western time zone *and* some additional time zone (Idaho, Indiana, Kentucky, Oregon and Tennessee), so that for DMAs in these states, the time zones need to be separately checked.

| | | | | |
|--------------------|-----------|-------|-----------|-------|
| September | -875,156 | -0.18 | -6.01e+06 | -1.01 |
| August | -1.15e+06 | -0.24 | -5.13e+06 | -0.99 |
| July | -1.20e+06 | -0.34 | -3.27e+06 | -0.89 |
| June | 4.21e+06 | 0.74 | 2.98e+06 | 0.54 |
| May | -2.24e+06 | -0.68 | -6.73e+06 | -1.42 |
| April | -519,282 | -0.11 | -4.39e+06 | -0.87 |
| March | 4.09e+06 | 0.79 | 3.48e+06 | 0.78 |
| February | 2.81e+06 | 0.87 | 3.14e+06 | 0.99 |
| Constant | 5.23e+07 | 11 | 9.94e+07 | 3.89 |
| Observations | 46 | | 46 | |
| Adjusted R-squared | 0.878 | | 0.883 | |

Here are the predicted piracy-induced reduction in sales for individual years:

| Additional implied impact of piracy on sales | OS Reported | Direct Replication | Add Unemployment |
|--|-------------|--------------------|------------------|
| Implied impact on sales, 2003 | -19,078,501 | -28,497,114 | -121,083,694 |
| Share of drop due to piracy, 2003 | -14.8% | -22.1% | -94.0% |
| Implied impact on sales, 2004 | -23,586,017 | -35,229,887 | -149,691,116 |
| Share of drop due to piracy, 2004 | -26.4% | -39.4% | -167.4% |
| implied impact on sales, 2005 | -33,470,308 | -49,993,824 | -212,422,800 |
| Share of drop due to piracy, 2005 | -14.6% | -21.8% | -92.7% |

3. Genre PIRACY and Genre Sales:

Piracy Intensity Measure

The genre piracy intensity measure, for our purposes, should reflect of number of recorded pirated songs listened to as a share of all recorded songs being listened to by those consumers who would have listened to those songs if piracy did not exist. If piracy has no effect on sales, this ratio should be close to zero since pirated songs would only be listened to in order to decide whether to purchase the song or not, and that decision should not take too many repeated plays. Pirate listeners who would not have purchased the song without piracy should be excluded from these calculations since their behavior does not directly affect record sales.

For obvious reasons, we cannot construct an ideal measure of piracy intensity. OS have information from their sample of pirates on the number of pirated downloaded songs and the national sales for their sample of albums, classified by genre. NPD has data on the share of users in their sample who listen to pirated songs and the share who listen to purchased songs.

These variables allow the creation of rough measures of piracy intensity. OS use “number of downloads” [number of pirated files] and appear to divide it by the number of album *titles* in their sample of albums in a genre [footnote 22 discussed why this appears to be the case] although it is not clear that this was their intention. I create a variable which I believe better fits their variable name “downloads per album” by using album sales in the denominator, not album titles.

Downloads per album title is not a good choice for a pirate intensity measure because the number of album titles may not reflect the number of albums sold. Here is a simple example to illustrate the point:

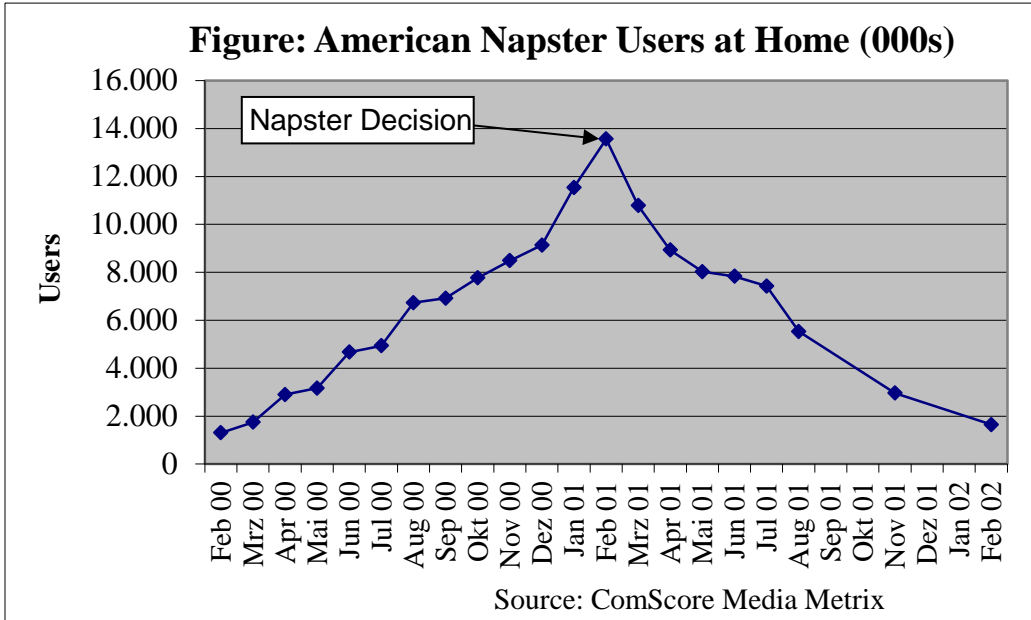
Genre A has 1 million albums sold, 1 million total pirated albums, and 1,000 titles (each selling 1000 units, on average). Genre B has 1 million albums sold, 1 million total pirated albums, and 2 titles (each selling half a million on average). Each genre has the same number of albums sold. Each genre has the same number of pirated albums. Each genre, therefore, has the same piracy intensity, about 50%. Yet, the OS measure of piracy intensity would indicate that Genre B has 500 times the piracy intensity as Genre A because the number of pirated files per album title is 500 times greater.

Finally, these measures take the form a/b , where, if we could do it correctly, a is a measure of piracy (on the part of those who would otherwise purchase the music) and b is a measure of sales. A more correct variable would take the form $a/(a+b)$ [where $a+b$ equals the size of the potential market], and this ratio would be linearly related to piracy intensity as it would be expected to influence sales, assuming that piracy reduces sales. The variable actually used, a/b , grows more rapidly as piracy grows than the variable $a/(a+b)$, if piracy reduces sales (b). In other words, piracy increases not only increase the numerator of a/b , but also decrease the denominator. Thus, the measures of piracy being used will show too high a level of piracy for genres with high levels of piracy and the coefficients relating sales changes to piracy intensity will be *understated* relative to the more appropriate variable $a/(a+b)$. This understatement of the size of the results discussed in the text merely lends more credence to the claim that this evidence supports the hypothesis that piracy decreases sales.

The history of Napster

This figure, taken from my 2006 article, demonstrates the growth of Napster. Napster was not even tracked by ComScore until February of 2000, when it had 1.3 million users. It peaked at just about ten times that level in February of 2001 (after which pirates moved to other pirate services). Obviously, the number of users when Napster started in 1999 would be much less

than in 2000. Time Magazine put Napster on its cover in October of 2000. The RIAA brought legal action against Napster in *December* of 1999 in order to nip its piracy problem in the bud.



RADIO STATION GENRES

Here is a list of radio station genres and sub genres according to Arbitron, the leading organization measuring the size of radio audiences. It should make clear how difficult it can be trying to match a record category, of which there are not as many, to the correct radio genre, since the two have different terminologies.:

| | | |
|-------------------------|-----------------------|---------|
| | | CL HITS |
| | | OLDIES |
| | | RHY OLD |
| | Rock | 70SROCK |
| | | ACTROCK |
| | | AOR |
| | | CL ROCK |
| | Spanish | LAT URB |
| | | MEX REG |
| | | SP ADUL |
| | | SP CONT |
| | | SP LANG |
| | | SP NEWS |
| | | SP OLD |
| | | SP REG |
| | | SP RELG |
| | | SP TROP |
| | | SP VRTY |
| | | TEJANO |
| Adult Contemporary | | |
| A/C | | |
| HOT A/C | | |
| MOD A/C | | |
| SOFTA/C | | |
| Adult Hits | | |
| ADLT HT | | |
| Alternative | Urban | |
| AAA | | URB A/C |
| ALTERN | | URB OLD |
| NW ROCK | | URBAN |
| Contemporary Hits Radio | News/Talk/Information | |
| CHR | | ALLNEWS |
| POP CHR | | ALLSPRT |
| RHY CHR | | NWSTALK |
| Classical | | TALK |
| CLASSCL | | |
| Country | Remaining Formats | |
| CLCNTRY | | CHLDRAD |
| COUNTRY | | EDUCAT |
| NWCNTRY | | ETHNIC |
| Adult Standards | | OTHER |
| EZ LIST | | R&B |
| MOR | | |
| NOSTLGA | Religious | |
| VARIETY | | CHRSTN |
| New AC/Smooth Jazz | | CO INSP |
| JAZZ | | GOSPEL |
| NAC | | RELGOUS |
| Oldies | | SGOSPEL |
| 70 HITS | | |
| 80 HITS | | |

4. Data claims made by OS

Leading Markets with sales increases 2000-2005 and 2004-2005

OS claim to have found leading markets with increases in sales over the period 2000-2005. Below are listed the nominal and real sales changes for the ten leading markets over the period 2000-2005.

| 2000-2005 Change in Retail Revenue | | |
|------------------------------------|--------|---------|
| | Real | Nominal |
| USA | -29.9% | -12.6% |
| Japan | -10.7% | -13.6% |
| UK | -14.4% | 1.6% |
| Germany | -42.8% | -32.5% |
| France | -26.2% | -14.5% |
| Canada | -36.7% | -21.7% |
| Australia | -29.7% | -8.3% |
| Italy | -41.0% | -25.8% |
| Spain | -56.2% | -34.6% |
| Netherlands | -46.1% | -30.2% |
| Switzerland | -19.5% | -19.5% |

None of the top 10 markets have an increase in real retail revenue over this period. These data come from the IFPI annual publication "Recording Industry in Numbers."

OS also report that music sales rose in 4 of top 5 markets in 2005. The table below, using the same IFPI data on retail sales, reveals that 1 of the top 5 markets had a real increase in revenue and 2 of the top 5 had a nominal increase in revenue. An additional five countries (ranked by size) are shown to indicate none of the other top 10 markets had real sales increases that year although Italy had a nominal increase.

| 2004-05 Change in Retail Revenue | | |
|----------------------------------|--------------|-----------------|
| | Real Revenue | Nominal Revenue |
| USA | -5.30% | -1.90% |
| Japan | 1.94% | 1.64% |
| UK | -5.46% | -2.66% |
| Germany | -1.56% | 0.44% |
| France | -4.24% | -2.54% |
| Canada | -4.32% | -2.12% |
| Australia | -12.33% | -9.63% |
| Italy | -1.73% | 0.07% |
| Spain | -8.47% | -5.07% |
| Netherlands | -18.27% | -16.57% |

2005 Sales Declines Supposedly at only a Single U.S. Record Company

| | Market Shares | | Sales (000s) | | Sales Change |
|-------------------|---------------|---------|--------------|---------|--------------|
| | 2004 | 2005 | 2004 | 2005 | |
| UMG | 29.59% | 31.71% | 201,475 | 207,449 | 5,974 |
| SONYBMG | 28.46% | 25.61% | 193,781 | 167,542 | -26,239 |
| WMG | 14.68% | 15.00% | 99,954 | 98,131 | -1,824 |
| EMI | 9.91% | 9.55% | 67,476 | 62,477 | -4,999 |
| OTHERS | 17.36% | 13.18% | 118,202 | 86,224 | -31,978 |
| Industry Sales | 680,889 | 654,206 | | | |

Market shares are from the news story in footnote 33 which references SoundScan data for the market shares of leading record companies. Sales figures for each company are derived from Nielsen's SoundScan's measure of U.S. sales. I multiplied the industry yearly sales by the market share statistics for firms to derive each firms' sales. The last column calculates the change in sales, which is negative for all firms except for Universal (UMG).

Conflict of Interest Disclosure Form

Please state the sources of financial support for the research.

Center for the Analysis of Property Rights and Innovation (at UTD)

Please identify each interested party from whom you (or a close relative or partner) have received financial support summing' to at least U.S.\$10,000 in the past three years:

Society of Composers, Authors and Publishers of Canada

Law and Economics Consulting Associates Ltd, Australia

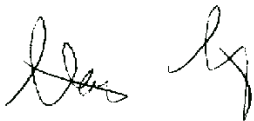
Please disclose any paid or unpaid positions as officer, director, board member of relevant nonprofit organizations or profit-making entities held by you (or by a close relative or partner):

Director, Center for the Analysis of Property Rights and Innovation

Has another party had the right to review the paper prior to its circulation?

No

Signature:



Stan Liebowitz

November 24, 2016

Please note:

You are most sincerely encouraged to participate in the open assessment of this discussion paper. You can do so by either recommending the paper or by posting your comments.

Please go to:

<http://www.economics-ejournal.org/economics/discussionpapers/2016-48>

The Editor