

# **File-Sharing: Creative Destruction or just Plain Destruction?**

**Stan J. Liebowitz**

**School of Management  
University of Texas at Dallas**

**March 2005**

Ten years ago the term “file-sharing” was unknown. Then Napster arrived and quickly etched the term and itself into the public’s consciousness. It is easy to dismiss the media coverage, such as Time magazine putting the creator of Napster on its cover, as just another case of reporters romanticizing the impact of a new technology, yet Napster truly began a revolution in music listening.<sup>1</sup> Although Napster was effectively shut down as an unauthorized file-sharing service within two years of its birth, its progeny live on, and the habits of music listeners have been dramatically changed with the now familiar additions to the lexicon such as “ripping” files from CDs, listening to “MP3s” on “Ipods” and, of course, trading files online with others using programs such as “Kazaa” and “eDonkey.”

File-sharing is merely the most recent example in a long line of technologies that have lowered the cost of unauthorized copying by individuals. The photocopier, introduced by Xerox in 1959, greatly enhanced the ability of individuals to cheaply copy printed pages. Audio-taping, which became popular in the 1970s, greatly enhanced the ability of individuals to copy sound recordings, with dual-cassette decks intended for high-speed copying becoming commonplace. Videotaping, which became popular in the 1980s, allowed individuals to copy broadcasts and prerecorded movies, although there was greater difficulty copying the latter due to primitive copy protection built-in to most prerecorded movies. File-sharing is not only the most recent of these technological changes but by most estimates, as we will see below, the most heavily used.

The copyright industries reacted negatively to each of previous copying technologies. The publishing industry complained about photocopying, although an analysis by Liebowitz (1985) concluded that photocopying was beneficial to the industry. The movie and television industries brought suit to stop the videorecorder, but the Supreme Court’s ruling in the Betamax case went against these industries and a new market emerged, video rental and sales. In hindsight we know that, although unanticipated,<sup>2</sup> prerecorded video sales now provide the movie industry with revenues in excess of box office revenues even though real box office revenues per capita have more than doubled since the decision, indicating that the new technology led to a new market, not just a substitute market.<sup>3</sup>

---

<sup>1</sup> Shawn Fanning, the college student who created Napster was on the cover of Time Magazine, August 2, 2000.

<sup>2</sup> To be fair to the industry, an alternative technology, video laserdiscs, did not allow potential pirating since the players did not have recording capability and in the absence of VCRs might have brought all the playback advantages of the VCR without the possibility of copying.

<sup>3</sup> Box office gross was \$9.49 billion in 2003 according to the Motion Picture Association of America. According to Vogel, the movie studios net slightly less than 50% of the box office. According the Adams Media Research, revenues to the studios from the sale and rental of DVDs was \$11.38 billion, and from VHS tapes \$2.56 billion. Thus revenues earned by the movie studios from prerecorded movies are approximately three times as great as that from theatrical releases.

The software industry expressed similar concerns about unauthorized copying and early in its history experimented with anti piracy devices. The industry soon decided that, for business products at least, the anti-piracy devices did more harm than good, and the industry appears now to believe that organized piracy is more destructive and/or amenable to solutions than is personal copying.<sup>4</sup>

The final example is home audio taping to which the sound recording industry objected after such taping became popular. No less a luminary than Alan Greenspan, testifying for the industry stated: “At present...severe economic damage [is being done] to the property rights of owners of copyrights in sound recordings and musical compositions...under present and emerging conditions, the industry simply has no out...Unless something meaningful is done to respond to the...problem, the industry itself is at risk.”<sup>5</sup> Sales of sound recordings began a decade-long expansion just as this testimony was being given, once again making the claims of concern by the copyright industry appear unwarranted.<sup>6</sup>

Unlike the audio taping example, however, the sound recording industry experienced a dramatic swoon in sales at almost the same moment that file-sharing became widespread. The industry has blamed this decline on the rapid growth of file-sharing and in an attempt to stem the growth in peer-to-peer usage has sued thousands of individuals heavily engaged in file-sharing. These lawsuits have attracted a good deal of notoriety and discussion.<sup>7</sup>

This attention is partly due to the nature of these industries. The copyright industries attract far more attention than would be warranted by their share of GDP alone, even though their share of GDP

---

<sup>4</sup> Consumers objected to the original anti-piracy devices, mechanical dongles or discs, which quickly disappeared from the market. Recently Intuit tried to limit the printing of results from its TurboTax software to the single computer upon which the program was first installed although consumer complaints caused Intuit to remove this feature the next year. Video games, on the other hand, have had a longer history of copy protection.

<sup>5</sup> This is from Greenspan’s testimony on the Home Recording Act. Hearings before the Subcommittee on Patents, Copyrights and Trademarks, October 25, 1983.

<sup>6</sup> The reason for this increase in sales appears to be a new market opened up by audio cassettes—the mobile prerecorded music market. Prior to the cassette the only portable form of music was radio. One might argue that the industry would have benefited even more if unauthorized home audiotaping had been restricted, but at a practical level the industry was better off for the new technology.

<sup>7</sup> On the academic front, a large number of critics have voiced their unhappiness with copyright law and the entertainment industry. These copyright critics, sometimes associated with the concept of the ‘creative commons’ and the Electronic Frontier Foundation, argue that copyright laws are being used by the sound recording, movie, and software industries to thwart competitive forces that would otherwise open up the market to new competition. This is the thesis of Laurence Lessig’s recent book *Free Culture* which views the current controversies as extensions of long-running debates regarding the power of cartels to monopolize access to creative works. In this view of the world, file-sharing is a wealth enhancing innovation, likely to democratize the entertainment industry by allowing artists to broadcast and distribute their works without intermediaries such as record companies. In this view, file-sharing systems should be promoted and if necessary, copyright law should be altered to allow file-sharing to proceed apace.

is quite large.<sup>8</sup> The advertising industry is primarily based around these industries, as is the consumer electronics industry which would include all forms of stereo equipment, televisions, DVD players, VCRs, computers, and so forth. Enjoyment of sound recordings, movies, the Internet, and television occupies a majority of the time that individuals spend on leisure activities, with the average American watching four and a half hours of television and listening to more than three hours of music each day.<sup>9</sup> Thus even beyond their share of GDP, the impact of these industries upon the collective consciousness and behavior of consumers is very large.

Each of the previous copy-enhancing technologies brought forth some work by economists on the subject. See the surveys by Varian (2005) or Watt (2000, 2004) for discussions of some of this earlier work, much of which consisted of theoretical models created to explain the impacts of copying.<sup>10</sup>

One of the problems with analyzing these older technologies was the difficulty in generating measurements of unauthorized copying. One of the promises of the new technology is to provide better data, but as demonstrated below, this promise remains largely unfulfilled. These data sources provide sets of numbers that are, sadly, largely inconsistent with one another.

Supporters of file-sharing suggest that copyright industries are once again crying ‘wolf’ and that, like the new technologies of the past, file-sharing will be seen, in hindsight, to provide beneficial opportunities for all, including copyright owners.

The analysis below concludes that the industry is not crying wolf. The evidence seems compelling that the recent decline in sales can be properly attributed to file-sharing. This analysis also suggests, less strongly, that file-sharing has reduced sales by at least the decline that has occurred and possibly has vitiated an increase that might otherwise have occurred.

Naturally the current concern over the impacts of copying brought forth renewed interest by economists. Recent papers include those by Blackburn (2004), Boorstin (2004), Liebowitz (2004), Michel (2004), Oberholzer and Strumpf (2004), Peitz and Waelbroeck (2004), Rob and Waldfogel (2004), and Zentner (2004). All of these papers find some degree of harm brought about by file-sharing

---

<sup>8</sup> It is estimated (Siwek, 2002) to be between five and seven percent of the GDP according to a report produced for a copyright trade association.

<sup>9</sup> These values are taken from Table No.1125 ‘Media Usage and Consumer Spending: 1998 to 2006’ in the 2003 US Statistical Abstract.

<sup>10</sup> The US Office of Technology Assessment, in the late 1980s, commissioned several studies to examine the impacts of home taping. These studies were based on survey data, which can be of questionable value when it comes to measuring copying, and there was no clear consensus about the impact of audio taping. See US Congress 1989.

except for the paper by Oberholzer and Strumpf.<sup>11</sup> These authors use different methodologies and different data source. The data have not been subject to much examination and that is the topic to which we now turn.

## **I. The Brief History of File-Sharing and its Measurement**

File-sharing, simply put, allows one computer on the Internet to search for and access files on the hard drives of other computers that are connected to the Internet. Any individual on a file-sharing network can make available any file on their hard drive to all other members of the file-sharing network.

The term ‘file-sharing’ is actually something of a misnomer, however. Individuals do not ‘share’ the files that move back and forth on the Internet. They do not experience these files together nor are they likely ever meet or even know one another. Nor do they lend or trade the files among one another, since the files are never borrowed, exchanged, or given back. A more appropriate term might be ‘file-copying’ since that reflects what actually occurs. The industries whose products are most frequently copied on these networks would most likely prefer a term such as ‘file-stealing’ since that is how the copyright-based industries view such activity. Since file-sharing occurs anonymously and between strangers, the files that are ‘shared’ are products with general appeal and commercial value, not home movies or baby pictures.

The end result of file-sharing is that individuals who do not own and have not purchased a particular song or movie can nevertheless obtain that song or movie from unknown third parties.

Napster came into existence in 1999, created with the purpose of allowing music files to be shared over computer networks. It was, for all intents and purposes, shut down by a preliminary injunction against it in early 2001.<sup>12</sup> Into the void stepped numerous other file-sharing programs, particularly those which, unlike Napster, were not based on a central server.

Currently, file-sharing encompasses sound recordings, films and television programs, computer software, various forms of pornography, and other products that can be digitized. Because music files were easily compressed, relatively small, very popular, and the primary type of file downloaded, they appear to be the best candidate for assessing the impact of file-sharing itself. File-sharing threatens or

---

<sup>11</sup> The Boorstin paper does not find a negative result but that is due to his inclusion of a questionable dummy variable for the years with file-sharing, as discussed in Liebowitz (2005).

<sup>12</sup> A&M Records v. Napster, 239 F.3d 1004 (9th Cir. 2001).

promises, depending on your point of view, to do for the movie and software industries what it has done for the sound recording industry.

When Napster first came into existence, most downloaders would not have had in place the requisite CD burners that would allow listening to this downloaded music except from a computer. Nor did these downloaders necessarily have the hard drive space to store large numbers of high fidelity mp3 files, which are less compressed than lower fidelity files. For these reasons, MP3 files were not, at the turn of the twentieth century, terribly good substitutes for music purchased on a CD.

Over the last few years, however, the MP3 audio files that are traded on file-sharing networks have become much better substitutes for the music on prerecorded CDs.<sup>13</sup> It is increasingly possible to play MP3 files directly on devices such as CD/DVD players or iPods, meaning that listening to MP3 files is not confined to the computer. Downloaders now can simply create musical CDs by converting MP3 files into the typical CD format (wav files) since CD burners have become much more common.<sup>14</sup> Thus, file-sharing now produces files that are, for most music listeners, very good substitutes for the purchased CDs.

Video files have not yet reached this level of substitutability. Because of the large size of video files, the downloading of movies generally is measured in hours as opposed to the minutes it takes to download an audio file. A DVD movie tends to take up at least four gigabytes of space, compared to an audio CD's 700 megabytes. MP3 compression can reduce the size of a typical audio album to 50 megabytes while still retaining high quality audio. Compression routines can reduce the size of movies, but the video and audio quality of a movie compressed to 1 gigabyte is considerably lower than the original DVD quality.<sup>15</sup>

---

<sup>13</sup> This ignores the possible dilution of expected quality due to 'spoofer files' which look like music files but are actually empty files employed by the recording industry to increase the costs of downloading. It is also the case that anti-piracy countermeasures, such as spoofer files, might impact the measured relationship between file-sharing and CD purchases, but since we have no information on the nature of these countermeasures they are ignored in the following sections.

<sup>14</sup> In June of 2002, according to Ipsos/Tempo, 53% of American file-sharers had CD burners, which was more than twice as high as for the general population as a whole. According to Ipsos/Tempo, the penetration of CD burners for the general population increased by 42% from the first quarter of 2002 to the first quarter of 2004 (from 22% to 31%). If the growth in penetration for the population of downloaders was similar, this would have led to a penetration rate of 75% among those engaged in file-sharing.

<sup>15</sup> Compressed movies that can be directly played in DVD players tend to be VCD or SVCD formats, which represent approximately VHS quality and laserdisc quality. These movies are typically recorded on CDs, but it takes 2 or 3 of these CDs for each movie, making them far less convenient than DVDs for watching movies. The popularity of these formats reflects the dual constraints of Internet connection speed as well as the apparently limited number of DVD burners that have been available.

Similarly, DVD burners (and the software to bypass the built-in DVD copy protection) are not as ubiquitous as CD writers. Downloaded movies, therefore, are not particularly good substitutes for purchased or rented movies, although this is likely to change in the next few years.<sup>16</sup>

For analysts, digital file-sharing and its attendant measurements should have allowed a great improvement in understanding the impacts of copying since all prior estimates of copying were based on surveys of one sort or another and surveys have well known flaws. Survey questions trying to elicit information on file-sharing's impact on sales are likely to be unreliable for several reasons. Truthful respondents are not likely to know with any precision how much copying they do, how many originals they normally purchase, or how their purchases were changed due to the copying technology. Further, many respondents would be expected to be untruthful since they are being asked to acknowledge culpability in activities of questionable legality. Unlike surveys, the transfer of digital copies over the Internet should be amenable to counting by virtue of their computer linkage. File-sharing, therefore, appeared to hold the promise of allowing, for the first time, precise measurement of copying activities.

The reality of file-sharing measurements, however, does not yet live up to this promise. Given that there are numerous news reports mentioning the number of files being traded or the number of individuals using file-sharing networks, you would think getting solid and consistent measurements would be easy—but it is not.

Although there have been numerous news stories in the last few years reporting statistics on file-sharing, these reports generally cite the same few sources. At the time of Napster, Webnoize was most frequently quoted in the press, although it went belly-up at the end of 2001.<sup>17</sup> It is not clear what methodology was behind the analysis of Webnoize, but it reported that 2.79 billion files had been transferred at the peak month of Napster (February of 2001) and that by August of 2001 the number of files transferred on the four leading Napster replacements (FastTrack, Audiogalaxy, iMesh and Gnutella) had reached above 3.05 billion per month.<sup>18</sup> To put this in perspective, worldwide sales of music amounted to about 3 billion songs per month in the year 2000, so one might conclude from these

---

<sup>16</sup> See Lee Gomes "Coming Soon: Movies You Rent On the Web -- And Then Download" Wall Street Journal, August 30, 2004, Page B1.

<sup>17</sup> See for example, "Napster users turn down the volume: Downloads fall sharply in April." San Francisco Chronicle, Wednesday, May 2, 2001 by Benny Evangelista Available at: <http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2001/05/02/BU184607.DTL>.

<sup>18</sup> For example, see VNUnet.com 07 Sep 2001 "Music downloads on the rise" by John Gerald, at <http://www.infomaticsonline.co.uk/news/1125301>.

figures that the number of songs being downloaded on file-sharing networks was equivalent to the number of songs purchased in the legitimate market.<sup>19</sup>

The focus of analysis in this paper is on the American market for which there has been abundant reporting of recent statistics, although the numbers still tend to come from a few key sources. At the high end, there are claims that up to 60 million Americans have used peer-to-peer networks,<sup>20</sup> that perhaps as many as five billion music files are downloaded by Americans in a typical month (18 files for every man, woman, and child!),<sup>21</sup> and that perhaps 60% or more of all Internet bandwidth is taken up by file-sharing.<sup>22</sup> Although I am quoting the more striking numbers that have been put forward, even the more subdued estimates appear less reliable than we would like. Before discussing the actual measurements, however, a brief discussion of the methodologies is in order.

There are at least three types of macro data on file-sharing that have been collected. First, statistics on the number of participants in file-sharing activities are provided by several sources (e.g., comScore Media Metrix, Nielsen NetRatings, Big Champagne, PEW Internet and American Life Project). ComScore and Nielsen examine the number of users of particular file-sharing programs, such as Kazaa or BitTorrent.<sup>23</sup> Big Champagne measures the number of users of file-sharing networks.<sup>24</sup> Second, the number of files made available to others by a panel of file-sharers has been measured (Bhattacharjee et. al., 2004, or BGLM). Third, the number of files actually transferred has been estimated by one company (NPD).

Each of these types of data can be generated using various methodologies.

---

<sup>19</sup> This statistic comes from the IFPI in their brochure: “2000 Recording Industry World Sales”, April 2001. They report 3.5 billion albums per year. If we assume 10 songs per album, this works out to 2.91 billion songs per month.

<sup>20</sup> The Electronic Frontier Foundation reports that 60 million Americans use file-sharing software, but it is not clear where that estimate comes from. In May of 2003 ComScore/Media Metrix reports total users of file-sharing software at 45 million users, but this could include double counting of users who are on multiple networks. The comScore estimate arrives at its number by taking the number of households that operated a file-sharing program on their computer in the month and multiplies this value by the number of people in a household, so it is probably too generous a number.

<sup>21</sup> According to IDATE, there were either twelve or sixty billion audio files downloaded in the US in 2003, depending on which of two seemingly inconsistent statements you wish to believe. See [http://www.idate.fr/an/qdn/an-03/IF282/index\\_a.htm](http://www.idate.fr/an/qdn/an-03/IF282/index_a.htm). The breathless prose goes on to predict that by 2008 broadband users will download an average of 4300 audio files per year, which seems a quite outlandish prediction.

<sup>22</sup> For example, see “ISPs reel from P2P bandwidth hogs” Reuters May 23, 2003 which reports the 60% figure. Also, IDATE reports that “According to virtually all the experts in this field, P2P represents on average between 50% and 60% of all broadband traffic during the daytime, and as much as 80% to 90% of all night time traffic.”

<sup>23</sup> This tends to overstate the decline if file-sharers were migrating away from monitored programs toward programs that were not monitored. Companies such as comScore try to update their list of programs to keep up-to-date, but their measurements will always lag somewhat behind the behavior of users.

<sup>24</sup> There are often several different software programs that use the same network. These networks are not identical to the programs that use these networks. For example, Kazaa, Grokster, Kazaa Lite (a competitor to Kazaa) and iMesh all use the ‘FastTrack’ network but there are also other networks, such as Gnutella and DirectConnect, which are used by a different set of programs.

The most common methodology takes a panel of users and follows their file-sharing behavior over time. Nielsen NetRatings, NPD, and comScore/Media Metrix collect data based on voluntary panels and then adjust the data to represent the entire population. Members of the voluntary panel agree to allow their behavior to be monitored by a program installed on their computers which relays information back to the monitoring organization. Members of voluntary panels usually receive some small compensation as a reward for being members of the panel. These panels range from a few thousand to the hundreds of thousands. Alternatively, it is possible to create a panel whose members are unaware of being monitored, as BGLM have done.<sup>25</sup>

Big Champagne's method of measuring the activity on file-sharing networks is proprietary and therefore difficult to judge.<sup>26</sup> It does not have a panel of users but instead claims that it can measure the number of users on file-sharing networks, although it is not entirely clear whether these are downloaders, uploaders, or both. Big Champagne claims to measure the number of users of these networks every few minutes, calculating peak and average values for the month. The peak monthly values, however, are only about 60% above the average values in the last half of 2003 and only about 30% higher in the first half of 2004. This relatively small difference between peak and average seems highly implausible, and is a possible source of concern about the reliability of the Big Champagne numbers.<sup>27</sup>

Finally, there are surveys regarding Internet usage. For example, dueling surveys in the hearings that led to the preliminary injunction shutting down Napster reached opposing opinions about impact

---

<sup>25</sup> These authors measured the number of files being shared by approximately 2000 users of Kazaa over 52 weekly intervals. This study found a large decline in the number files made available after the RIAA lawsuits, consistent with the comScore and Nielsen results. They also found an upward trend in the number of files available per person at the time the lawsuits were announced followed by a decline in the number of files available per person. The BGLM study has several weaknesses. If users switch to another program, change their Kazaa logon id, or stop downloading files, it would show up as a decline in file-sharing. There is no attempt to replenish the sample or to determine if the sample is representative of Internet users as a whole. This makes their measurements very likely to understate the total number of files being shared over time although the average number of files shared per remaining user need not be biased.

<sup>26</sup> For a critique of Big Champagne's methods see <http://onlinetonight.net/archives/001415.html>.

<sup>27</sup> Eric Garland of Big Champagne argues, in correspondence with this author, that this small difference between peak and monthly values is due to the fact that many Internet users keep their machines and file-sharing software running 24 hours a day. This is possible, but doesn't seem likely. For one thing, broadband penetration (according to Nielsen/NetRatings) was only 35-42% in the last half of 2003 and 43-50% in the first half of 2004. Dial-up users seem most unlikely to keep their phone lines tied-up 24 hours a day. Also, informal polling of my students indicates that even those users with broadband tend to shut off their computers at night with a frequency of 50%. And those students who keep their computers on tend not to keep their file-sharing software running all the time. If these figures are accurate, it implies that at least 75% of users turn off their file-sharing software at night or when they are at school or work so that peak usage should be at least two or three times as high as average use.

of file-sharing.<sup>28</sup> The Pew Internet and Society project has conducted numerous surveys on Internet usage in the last few years, as has Ipsos/Tempo.

Unfortunately, each data source, regardless of data type or methodology, suffers from one or more imperfections.

All panel-based data sources, for example, are open to the criticism that the panel might not reflect the user population. In the case of Internet panels, a charge could be leveled that individuals willing to have their computer monitored by third party software might not be reflective of the overall file-sharing community.<sup>29</sup>

A second criticism that can be leveled at many of these data sets is that they measure proxies for the variable of interest, and not the variable itself. Clearly, if our goal is to estimate the impact of file-sharing on sales of CDs, the number of files downloaded by file-sharers is the statistic that we would want to use since it is the downloaded file which is a substitute for the original.<sup>30</sup> The number of users or the number of files available for downloading are merely useful proxies. But the number of individuals who use a file-sharing program may be far less than perfectly correlated with the number of files traded. It is conceivable, for example, that there might be large increases or decreases in the number of files exchanged even when the number of individuals using file-sharing software is constant (e.g., as they shift to broadband).

There is an additional important concern that holds for most of the measures of file-sharing. To the extent that video and computer files might form an increasing share of the file-sharing universe, measures of overall file-sharing might not properly reflect the downloading of *music* files. If video files are becoming increasingly important in the file-sharing universe, as seems to be the case, then overall file-sharing statistics might overstate the file-sharing of audio files. According to an OECD report (2004) using Big Champagne data, audio files dropped from 63% to 49% of total files from 2002 to 2003 although these Big Champagne numbers appear to be in conflict with other data.<sup>31</sup>

---

<sup>28</sup> See Liebowitz (2002) for a detailed description and critique of these surveys.

<sup>29</sup> This criticism loses some force from the fact that most file-sharing software (e.g., Kazaa Media Desktop) includes 'spyware' and 'adware' which monitor the usage of the computer to better gauge the commercial value of the computer user to vendors of various products. This means that many file-sharers have already allowed third party software on their computer.

<sup>30</sup> This is not quite right since consumers normally purchase CDs containing approximately ten songs and it is very difficult (prior to the creation of legitimate download sites) to purchase individual songs since the number of singles released by the record labels is quite small. CD's therefore, are less than perfect substitutes for the download of individual songs.

<sup>31</sup> It is not clear whether this measure is supposed to be based on the number of files available on hard drives, the space taken up on hard drives, the bandwidth taken up in transfers, or the number of files being transferred. I find it implausible

Finally, several of the measures that are based upon the usage of particular file-sharing software, such as comScore and Nielsen, are liable to provide misleading results if file-sharers shift from one program to another. Even though comScore and Nielsen try to add new programs to the list of ‘monitored’ programs, they inevitably must react with a lag, leading to an underestimate of usage if users are switching to new, unmonitored programs.<sup>32</sup> Although Big Champagne, based as it is upon networks, is susceptible to the same type of problems if new networks appear, it is likely to be a less serious problem since there are fewer networks than programs.

Although the weaknesses of survey data have already been discussed, surveys sometimes provide the only source of some data. Results from the Pew Internet and American Life project, for example, indicate that downloaders tend to be young and male. Their results also indicate that downloaders are more likely to be poor and less educated, but whether this is an independent factor or just a covariate with youth is unclear.<sup>33</sup>

Our interest, of course, is in the extent of file-sharing activity. As a benchmark we can start with some Pew survey results reported in Table 1.<sup>34</sup> In October of 2002, according to the survey, 19% of the adult population downloaded music.<sup>35</sup> Of those aged 18-29, 41% downloaded music. These numbers are consistent with a view that file-sharing is a very popular activity.

---

that the number of *movie* files transferred is as high as half the number of audio files (which is the Big Champagne statistic reported by the OECD). Lyman and Varian (2003), in their table 8.9, report that although shared video files take up twice as much hard drive space as shared audio files, there were ten times as many audio files residing on the hard drives of computers in 2003. Whether this inconsistency is due to misreporting by the OECD or an actual difference in measurement with Big Champagne is uncertain.

<sup>32</sup> Nevertheless, when new programs are added to the list they often seem to be marginal. ComScore, for example, added BitTorrent, eMule and Kazaa Lite to its list after January of 2003, yet its measurements of the usage of these newer programs generally indicated that that omission of these programs would not have led to major errors.

<sup>33</sup> See page 4 of the Pew Internet Project and ComScore Media Metrix Data Memo authored by Lee Rainie, Mary Madden, Dan Hess and Graham Mudd, “RE: The impact of recording industry suits against music file swappers”. January 2004

<sup>34</sup> To construct this table I multiplied the percentage of respondents answering this question in the affirmative, which was only asked to those with Internet access, by the number of respondents claiming to have Internet access.

<sup>35</sup> If this question “Do you ever...” were interpreted in the past tense, such as “have you ever” we would not expect the numbers to ever fall as long as respondents are being truthful. The decline that does occur can be taken either as evidence that the question is not interpreted as ‘have you ever’ or that some respondents might have begun to lie due to fear of prosecution. It is also possible that some respondent might have stopped downloading but still answer the question in the affirmative because they interpret the question to mean ‘have you ever’.

| Table 1: Percentage of Adult Population answering yes to question: "Do you ever download music files onto your computer so you can play them at any time you want?" Source: PEW Internet Project |             |             |        |        |             |
|--|-------------|-------------|--------|--------|-------------|
|  | July-Aug-00 | Aug-Sept-01 | Oct-02 | Nov-03 | May-June-04 |
| Overall  | 11%         | 15%         | 19%    | 9%     | 13%         |
| 18-29  | 25%         | 36%         | 41%    | 23%    | 31%         |
| 30-49  | 11%         | 16%         | 21%    | 9%     | 11%         |
| 50-64  | 3%          | 6%          | 8%     | 4%     | 6%          |
| 65+  | 2%          | 2%          | 3%     | 2%     | 2%          |
| Men  | 12%         | 19%         | 22%    | 12%    | 17%         |
| Women  | 10%         | 13%         | 16%    | 7%     | 9%          |

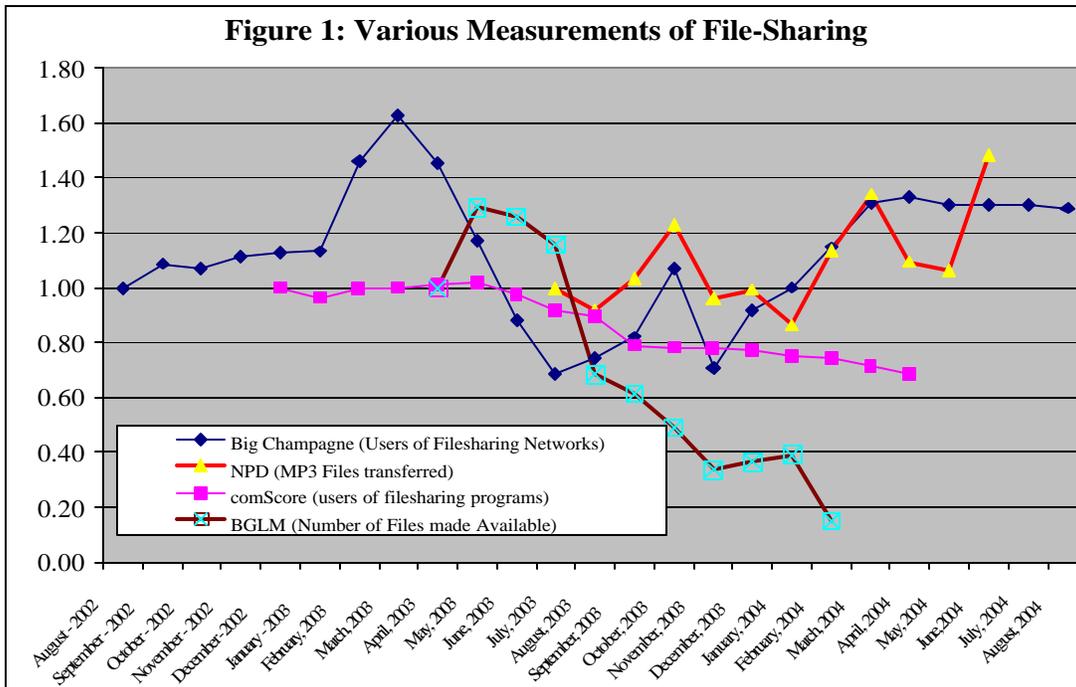
What do the other data sources have to say about the number of file-sharers? ComScore claims that there were 40 million unduplicated users of file-sharing software during January of 2003, which is in general agreement with the PEW numbers. Big Champagne measures the number of ‘simultaneous’ users on file-sharing networks at numerous intervals during the day, averaged over the month. In January of 2003 this average usage was just shy of 4 million users. It is difficult to compare Big Champagne numbers with comScore since comScore essentially measures, by way of analogy to television and radio, the reach per month and Big Champagne tends to measure average audience. Reach figures are always much higher than average audience figures. Another possible cause of the difference between the two is that Big Champagne counts computers whereas comScore attributes the entire household’s population to each computer using file-sharing software.

How do these non-survey data correspond to one another in terms of trends? The answer is: “not very well at all.” Figure 1 plots data from four of these data sets. Each data set has a different starting date. Since the numbers reported in these data sets are generally not strictly comparable, I normalize each of the data sets so that its starting measurement is 1.

For the two year period, these data sets provide different views about the changing size of file-sharing activity. The trend found in Table 1 based upon the Pew surveys indicated a sharp decline in 2003 followed by a small increase in the first half of 2004. There is some confirmation of this 2003 decline in Figure 1. Both Big Champagne and comScore data reveal a large decline (25%-50%) during mid 2003, although the Big Champagne data indicate that the number of file-sharers increased during 2004 while the comScore data indicates a continuing decline during the first half of 2004. Data from BGLM show an even sharper decline beginning in mid 2003 (as do Nielsen statistics which are not shown).<sup>36</sup>

---

<sup>36</sup> I do not report details of the Nielsen statistics since I only have two data points.



The degree of correlation between these various measures is disturbingly low, as reported in Table 2. Even if these variables measure different aspects of file-sharing, it is troubling that the correlations are not higher than they are. Big Champagne and comScore each purport to measure the number of American file-sharers, yet the correlation between their estimates is negative.

| Measurement Provider | NPD   | comScore | Big Champagne | BGLM |
|----------------------|-------|----------|---------------|------|
| NPD                  | 1     |          |               |      |
| comScore             | -0.44 | 1.00     |               |      |
| Big Champagne        | 0.67  | 0.26     | 1.00          |      |
| BGLM                 | -0.16 | 0.96     | 0.09          | 1    |

NPD purports to measure the actual music files being downloaded. If they have succeeded in measuring this it should be the gold standard of these measures. Since the NPD data only began in the summer of 2003 the series is too short to provide the basis for much analysis.<sup>37</sup>

Are the NPD data accurate? We do not know. But if they were we would be able to draw some interesting inferences. If the NPD data were correct, then the Big Champagne data would appear to be the most reliable of these other non-survey data sets, based on its higher correlations with NPD. The

<sup>37</sup> These NPD numbers are not comparable to earlier NPD numbers that were cited in various published articles. Apparently the earlier NPD numbers included files that were burned from CDs along with the files that were downloaded. These earlier NPD numbers had shown a large decline during the summer of 2003, which is when the RIAA lawsuits occurred. See for example Harrison (2003) at <http://www.nwfusion.com/newsletters/fileshare/2003/0922p2p2.html>

Pew survey data also indicate an increase from November 2003 until June 2004 that is also quite similar to what NPD finds, although there are not enough data points to calculate a useful correlation.<sup>38</sup>

Do these data allow us to draw any conclusions about the historical trend in organized file-sharing? Clearly, organized file-sharing stood at zero in 1998. By 2001, which was both the zenith and the nadir of Napster, it had grown quite large. The only series that I have seen that goes back this far is from the Pew surveys, and it indicates an increase in the popularity of file-sharing, from the summer of 2000 until October of 2002, on the order of 50%. All the data sets indicate a substantial drop during early and mid 2003, with some data sets indicating the decline was closely tied to the onset of the RIAA lawsuits and others seeming to show little coincidence of the timing of the two events, which I discuss in more detail below.

The variation in measurements of the absolute size of file-sharing activities is even greater than variations in measurements of the trend. If the Webnoize statistics on Napster use had been accurate, and if the US contributed 30% of the world usage of Napster, this would have implied about 1 billion files per month downloaded by Americans.<sup>39</sup> These are lower estimates than the more recent 5 billion file per month statistic reported by IDATE, which was mentioned at the beginning of this section. But both of these are in stark contrast to the numbers reported by NPD.

In July of 2003, for example, NPD reports that there were 169 million music files downloaded in the United States. This is less than *one twentieth* the monthly estimates reported by IDATE and less than one fifth the numbers reported by Webnoize for a period two years earlier. Since the methodology behind NPD's numbers is fairly transparent, whereas that of Webnoize is nonexistent, and since some of IDATE's prognostications appear rather incredible, it is tempting to accept the NPD data as correct. Without additional confirmation from other sources, however, we should refrain from drawing such a conclusion.

Nevertheless, the NPD data indicate that file-sharing activity, instead of surpassing the legitimate music business in size, is actually considerably dwarfed by the legitimate market (legitimate market

---

<sup>38</sup> Before we crown Big Champagne and Pew as the data kings for the period prior to NPD, however, we must repeat the already mentioned expected weaknesses of surveys such as Pew, and the problems of Big Champagne. Besides the concern of the small difference between peak and average values mentioned above, Figure 1 also indicates a very large decline (more than half) in peer-to-peer usage in the early to mid part of 2003, beginning prior to the RIAA lawsuits, following an inexplicably large increase, and each change occurred for no apparent reason, casting further doubt on the validity of the Big Champagne statistics.

<sup>39</sup> NUA reports that in August of 2001 there were 515 million Internet users worldwide and 166 million in the US and Canada. See [http://www.nua.ie/surveys/how\\_many\\_online/](http://www.nua.ie/surveys/how_many_online/).

purchases were running at approximately 900 million songs per month in 1999 although they are now approximately 650 million).<sup>40</sup>

The absolute size of file-sharing activities has an importance quite separate from the information found in the trends of file-sharing activities. Prior to NPD data, downloads appeared much more numerous than any possible decline in sales, thus it has been perfectly reasonable to conclude that a large portion of downloaded files did not replace the purchase of originals. If the NPD figures are correct, however, this conclusion would need to be reexamined. The decline in CDs is approximately 200 million units. Assuming there would have been no growth in the absence of file-sharing, this reduction in CDs represents approximately the same number of files as reported by NPD.

This also opens up an entirely new possibility that would cast file-sharing in a far more negative light. If, as previously thought, a majority of files being shared are not replacing sales, then the majority of files are generating new value to some consumers. Under those circumstances consumer surplus might increase sufficiently to overcome any loss of surplus due to a decline in the quantity of new albums brought about by reduced revenues. If, on the other hand, the majority of downloaded files are directly substituting for a sale, then not only would file-sharing be harming producers and causing a decline in albums, but it would be producing little if any additional surplus since most downloads would merely substitute for an original purchase and the activity of downloading would merely create a transfer from producers (who no longer receive payment) to these consumers. The likelihood of decreased total surplus is much higher in this case.

## **II. Economic Theory of File-Sharing's Impact**

In the last few decades, economists have come to understand that unauthorized copying of originals need not have negative impacts on copyright owners. The question becomes the relative strength of potential competing forces. In the case of file-sharing, however, it appears that the relative strength of these competing forces is more likely to have a negative impact on sellers than a positive impact.

On the one hand, the unauthorized downloading of a copyrighted file can easily be seen as a substitute for the purchase of that copyrighted work (song or CD). The replacement of a downloaded

---

<sup>40</sup> If the NPD figures are accurate and if claims of Internet congestion caused by file-sharing were also accurate then we might expect that the Internet would not have had the capacity to handle a full switch of the legitimate CDs purchases to online sites such as iTunes. This is an interesting possibility to ponder although not one to take to seriously since the data do not seem terribly reliable and network capacities are continually rising

copy for a purchased original is easily understood to have a negative impact on sales, particularly if the copy is identical or close in quality to the original. When a downloaded copy is a good substitute for a purchased original there is little reason for the listener to purchase the copyrighted work unless the continued use of a copy provides some disutility due to, say, the possibility of being punished for violating the law or the guilt from not supporting one's favorite artists.

This substitution effect can only work to reduce the effective demand in the market facing the seller of sound recordings, and can only harm the financial position of the sound recording companies.

The degree to which downloaded audio files can substitute for purchased originals depends on several factors including: a) the audio quality of the download relative to a purchased original; b) the ability of the downloaded file to substitute for a purchased original in ways other than audio quality (posters, information about the artists, lyrics); and c) the ability to listen to the download song in the same and as wide a variety of locations as can be done with an authorized original.<sup>41</sup>

It is hard to imagine that this substitution effect does not play an important role for some reasonable subset of the downloading population. Simple observation of acquaintances and family members generally provides clear anecdotal evidence that this substitution effect is not zero.

A second possible impact of file-sharing argues that users might merely use downloaded songs to sample from available music to help guide their purchases. The consumers get a chance to discover those songs and albums that most closely match their tastes then adjust their purchases in accordance with this new information. Although this idea can be traced back several decades, and was originally referred to as the *exposure* effect,<sup>42</sup> it is currently called the *sampling* effect, as it was in the Napster case. Note that sampling is less plausible for movies than for music since movies are only usually viewed a few times whereas the same song can be listened to repeatedly.

This sampling hypothesis is usually associated with a claim that the seller will benefit if consumers are allowed to become more familiar with the product before they purchase it.<sup>43</sup> As I have

---

<sup>41</sup> Claims have sometimes been made that to counter the substitution impact producers merely need to alter their business models to offer additional value items with the original CD that cannot be emulated over file-sharing networks, or that they need to lower price (see xx). Either of these actions, however, can be expected to lower the profit of the seller (assuming that profits were maximized originally) and as such will harm to the seller. The fact that certain tactics might be able to reduce the harm does not refute the argument that harm occurs.

<sup>42</sup> See Liebowitz 1981, 1985.

<sup>43</sup> This claim has been made, for example, by Digital Mogul and CENTRISS based on a survey they conducted, reported in a story available at <http://www.centris.com/announcements/musicannounce.htm>

argued elsewhere (Liebowitz, 2004), however, sampling in the market for music files would likely lead to a decrease in revenues in this market.

Using an insight of Jack Hirshleifer (1971), it is useful to consider the market for ‘music listening services’ as the ultimate product that is consumed, with the CD or audio file as the transmission mechanism for fulfilling that demand. The basic idea is to notice that albums that have more prescreening by consumers will contain, on average, music that provides greater music listening services than albums purchased without the information found in a prescreening. This prescreening twists the demand for albums clockwise since the initial albums purchased will contain more valued music listening services but by the same logic a smaller number of albums can satiate the demand for music listening services.

The elasticity of demand for music listening services will determine whether the number of albums purchased increases or decreases. Because the production of music CDs has very low variable costs this means that the production of music listening services has very low variable costs.<sup>44</sup> Record companies, to the extent that they maximize profits, will price music listening services at a level where the elasticity of demand is close to one. As is normally the case, competition between record companies should cause the *market* elasticity of demand for music listening services to be less than the elasticity facing individual firms.<sup>45</sup>

Since the effect of sampling (which generates more music-listening services at a constant CD price) is to effectively lower the price of music-listening services in an inelastic market, the impact of sampling should be to lower the revenues. With a price per CD that is independent of the sampling effect, this implies that sampling will cause the quantity of CDs to fall.<sup>46</sup>

---

<sup>44</sup> Although it is merely a theoretical convenience in models of markets like software or music to commonly assume a zero marginal cost of production, and this assumption is often not supported by the facts, variable costs do appear to be quite low in the case of sound recordings due to the peculiarities of the contracting. Although the artists normally receive a royalty that is expressed as a function of sales, those royalty payments are usually paid up-front as a non-refundable advance, so for most units sold, marginal royalties paid by the producers are effectively zero. Promotional costs for CDs are also usually also taken out of up-front advances, removing another potential variable cost from the variable cost column.

<sup>45</sup> There is an added potential complication here. A different view might be that firms produce multiple competing CDs in an analogy to a multiproduct firm. For such a model to be appropriate, one must believe that the music listening services from CDs are imperfect substitutes for one another. Multiproduct pricing would imply that the price for CDs would be raised above the level that would exist if the firm produced only a single CD. Of course, we still have competition between record companies and as long as the competitive effect outweighs the multiproduct effect, the elasticity in the market, to the extent that we can define the market, will be less than one. My thanks to Mike Ward for bringing up this point.

<sup>46</sup> Support for this view of sampling comes from analyses of cable television in many parts of the world. Cable should have increased the probability that viewers would find a program to their liking during any viewing hour, just as sampling would allow listeners to find a better album. Previous research examining cable’s impact of increased viewer choice on total time viewing television tended to find that that providing more choice to consumers did not increase the time they spent viewing

The sampling effect could well occur. Nevertheless, because sampling is unlikely to increase the sales of the industry, the existence of even a large subset of the population engaged in sampling is not likely to counterbalance the negative impacts of the substitution effect.

A third claim about the impact of file-sharing concerns the role of network effects. Models such as Takeyama (1994) demonstrate that under certain conditions, unauthorized users of an intellectual product might create sufficient additional value to the purchasers of legitimate copies that sellers might benefit from the unauthorized use.<sup>47</sup>

There are several issues that arise in the context of possible network effects for sound recordings. The first is whether there are network effects at all for music listening and what the nature of those effects might be. A related issue is whether such network effects, if they exist, would merely shift demand among different sound recordings or whether they would work to alter the size of the total market.<sup>48</sup> In other words, even if individuals cared about the music that others were listening to, would this cause them to purchase more music in total or merely to have their purchases mimic the purchases of others?

I am doubtful that network effects are likely to be important for overall record sales although they are likely to be very important in the sales of individual recordings. I base this conclusion in part on conjectures about the nature of network effects in these markets, in particular the insight that any network effects of importance would have to be based on all forms of listening to music, not to the listening of prerecorded music. Economic models of the impact of network effect of CD sales treat MP3 downloads as a new, less expensive form of music listening for individuals who had been unable or unwilling to purchase CDs prior to the advent of file-sharing. A decline in the price of music

---

television. See Liebowitz (1982) who compares the link between viewing hours and cable penetration across different Canadian metropolitan areas and finds an insignificant but sometimes negative relationship. Also see Weimann (1996) who examines viewers in Israel after the introduction of a multi-channel cable system where previously there had been but a single public channel (a more extreme increase in choice than would normally be found). After a year, there was virtually no difference in changes in viewing between a group with cable and a control group that did not receive cable (the cable group increased its viewing by 16 minutes over the control group). There are several papers looking at the impact of cable on different European countries and reaching largely the same conclusions that can be found in Becker and Schoenbach (1989).

<sup>47</sup> For example, if individuals are familiar with using a unauthorized copy of a spreadsheet, then their employers, who purchase legitimate copies, might place higher values upon purchasing spreadsheets since training costs are lower

<sup>48</sup> Were network effects to leave the revenues of the sound recording industry unchanged while at the same time increasing usage, the net effect would be to decrease appropriability. Although we cannot pretend to know with certainty whether the prior level of appropriability was efficient, it would be wrong to conclude that the impacts on society were only the salutary increases in consumer surplus from those file-sharers who get to consume the product but would have been priced out by the market. Indeed, the usual assumption would be that the reduced appropriability would cause an underproduction of new copyrighted works.

listening would be expected to increase the quantity as long as price is the binding constraint. Ignored in these analyses is the fact that there is already a zero priced and very popular alternative available—radio. In fact, published estimates are that individuals have for many years spent approximately three times as much time listening to radio as they did prerecorded music. It is true that the zero-priced mp3 alternative is now of higher quality and more in tune with the tastes of the listener than was radio, but this does not imply any increase in overall listening (the candy bar story, again). Thus, there is no reason to believe that MP3s would provide any new network effects that might increase record sales from those individuals whose purchases are not negatively impacted by the substitution effect.<sup>49</sup>

This conclusion is further supported by the finding that radio play has historically had little or no positive impact on record sales although it greatly influences which songs are purchased (Liebowitz 2004a). Since music listening on radio should exhibit both sampling and network effects in a manner similar to downloading, the fact that there was no positive impact on sales would imply that network effects were not particularly strong.

The final impact of copying that might apply to file-sharing is indirect appropriability. This is a concept coined in Liebowitz (1985) and discussed for the case of file sharing by Boldrin and Levine (2004), Klein and Murphy (2002), and Liebowitz (2002). The basic idea is that originals from which copies are made might undergo an increase in demand as those making copies of originals capture some of the value from those receiving the copies.

If, for example, everyone who purchased a CD made one cassette to play in their automobile, then the demand for the original CD would increase by the value of being able to make the tape and the sellers could capture some of this higher value by increasing the market price of the CD. This value is captured indirectly since there is no direct payment made for the copy.

In order for indirect appropriability to work, however, one of two conditions must hold. First, the variability in the number of copies made must be small, as in the example above. The second possibility is for the seller to be able to identify those originals from which the most copies are made and then charge higher prices for those originals, as journal publishers charge higher prices to libraries.

Because there is great variability in the copies made from each original on file-sharing systems and the sellers of originals cannot identify which originals are going to be used on file-sharing systems, the mechanisms that allow indirect appropriability to function will not work.

---

<sup>49</sup> See Liebowitz 2005 for somewhat greater detail.

### III. Using Natural Outliers to test the Theory

Economists do not often get to conduct controlled experiments. Instead we are usually relegated to performing regression analysis in an attempt to learn something about causation in a world flooded with noise. Sometimes, however, exogenous changes in a market occur in such a way, and with sufficient force, that we are provided an opportunity to learn something that might otherwise be very difficult or largely impossible to measure accurately.

Generalizations from a single or small number of observations of this type are quite natural. If you imbibe from a bottle with unknown contents labeled “drink me,” fall unconscious, and next wake up feeling very peculiar and much smaller, it is natural to assume that something in the bottle caused the change. Even with only one observation it is reasonable to attribute a great deal of certainty to the cause and effect. Similarly, if a volcano explodes with the most ferocious eruption in recorded history, and the following summer is the coldest in recorded history, it is reasonable to assume that the volcano caused the unique weather even if hitherto there had been no examination of a linkage between volcanoes and weather.<sup>50</sup>

What happens in each of these instances is the confluence of two unusual events. The further each event is from the norm, the more confidence we can have that there is a causal linkage. The reason is quite simple. When an event occurs with a magnitude that is very far from the norm, it is a very unlikely event. If two very unlikely events occur in succession, it is more likely that they are related than it is for them to occur independently.

The causes of such large events are normally exogenous, although they do not need to be. When they are exogenous, they can usually be classified as a natural experiment. The unique size of such large events makes them outliers, so I use the term ‘natural outlier’ to describe such a situation. The usage of natural outliers improves our ability to measure the impact of such events since the typical changes in other factors are relegated to the position of background noise.

This type of logic is applied all the time. When the first airplane flew into the World Trade Center, it could have been just a highly unlikely accident. When a second plane did the same thing a few minutes later it was obvious to all that these two events were related.

---

<sup>50</sup> I am referring in this paragraph, of course, first to Alice in Wonderland and second to the eruption of Mount Tambora in 1815 and the following “year without a summer.”

Something like this, of course on a much smaller scale, has occurred in the case of file-sharing. There are actually two sets of shocks that provide a unique window into the impact of file-sharing, although the first is a natural outlier while the second is not.

The first shock is the explosive growth of file-sharing itself. The creation of the technology making file-sharing possible was exogenous to conditions in the sound recording market. Never before had a copying technology and its usage spread so quickly through the economy. By growing so rapidly, the yearly changes in copying became magnified relative to the noise that occurs in yearly sales of CDs. At the same time, as we will shortly elucidate, the sales of sound recordings had a more dramatic decline than any other in our data set spanning thirty years.

The second exogenous shock was the well publicized bringing of lawsuits against file-sharers by the Recording Industry Association of America (RIAA). These lawsuits were brought with the intention of reducing file-sharing. There is now fairly compelling evidence that these lawsuits had at least a somewhat chilling effect on file-sharing in the United States, reversing the growth of file-sharing. This reversal provides another opportunity to examine the impact of file-sharing. This shock is not large enough to qualify as a natural outlier however, as discussed below.

### ***A. The Advent of File-Sharing***

Data on the sales of recorded music in the United States are available on a yearly basis from the Recording Industry Association of America (RIAA) as well as from private data collection companies such as SoundScan. Information on sales of individual albums or sales in metropolitan areas usually is available, for purchase, from SoundScan. The RIAA reports information on all shipments whereas SoundScan reports sales based in retail outlets and barcode readers.<sup>51</sup> Although the two sources of data should provide similar results most of the time, SoundScan data do not include information on non-retail outlets such as record clubs, which were responsible for 25% of all units sold in 1999. The recent decline in sound recording sales has fallen disproportionately on non-retail units (non-retail outlets accounted for 42% of the total decline in units that has occurred since 1999 and 27% of the decline in

---

<sup>51</sup> RIAA 'shipments' are not identical to sales since there can be returns for items which do not sell. Data from SoundScan report on actual sales of bar-coded albums. There are several advantages in using RIAA data beside that mentioned in the text. First, RIAA data are publicly available to all researchers without any required payment. Second, RIAA data include quantities and revenues (although the latter are based on list prices) whereas SoundScan data report quantities only.

revenues), so that there is a danger in underestimating the impact of file-sharing by using SoundScan data (used by Blackburn, Boorstin, Oberholzer and Strumpf, and later in this paper).<sup>52</sup>

Figure 2 represents the per capita sale of full-length albums sold in the United States since 1973 using RIAA data for the entire market.<sup>53</sup> The recent decline begins in the year 2000, which is one year after the arrival of Napster. Clearly, there are yearly fluctuations in this series. The use of full-length albums reduces somewhat the recent measured decline in units since singles have experience a steeper decline, but I have chosen to remove singles because their decline appears to be part of a much longer secular trend quite separate from file-sharing.



The decline that begins in 2000 is clearly the largest that has occurred in these data. In absolute terms (units per capita) it is more than twice as large as the next closest decline and in percentage terms it is nearly fifty percent larger than the next largest decline, which may also have been impacted by copying.<sup>54</sup> Another way to gauge the size of the drop is to count how many years back one would need to go to equal the new lower sales level. In the case of the current drop, one needs to go back to

<sup>52</sup> The yearly decline in retail versus non-retail outlets (based on RIAA figures for each channel) is illustrated below for the yearly change in total units sold. Almost identical results obtain if revenues are used. Clearly the two channels have not been impacted equally. It might be suggested that retail sales held up for other reasons, perhaps the move away from record stores to mass merchants, but the deviation between the two types of outlets begins just as file-sharing is starting in earnest.

|            | 1998   | 1999  | 2000   | 2001    | 2002    | 2003    | overall (99-03) |
|------------|--------|-------|--------|---------|---------|---------|-----------------|
| retail     | 3.98%  | 2.32% | -9.33% | -7.04%  | -7.83%  | -2.59%  | -24.32%         |
| non-retail | 11.39% | 6.21% | -0.10% | -19.00% | -21.84% | -23.80% | -51.80%         |

It would be useful to know more about the types of individuals using each channel and in particular whether non-retail outlets such as record clubs contained a disproportionate number of individuals who substituted downloaded files for purchased files.

<sup>53</sup> Data on quantities were reported beginning in 1973. For prior years only the industry revenues were reported.

<sup>54</sup> The second largest decline occurred in 1978-82. Long playing vinyl records were in decline at that time, being replaced by prerecorded cassettes. This decline coincided with a fairly deep recession and the possibility that blank cassettes were being used to copy records. As I explain in Liebowitz (2004), cassettes ultimately allowed, for the first time, portability of prerecorded music, leading eventually to a large increase in sales that overcame any negative impacts of copying.

1987—17 years—before encountering a year with lower sales per capita.<sup>55</sup> The second largest decline ended in 1982 and erased the (much smaller) gains of the prior 7 years.

This recent decline is sufficiently striking that it would appear that something unusual has occurred in the last few years. Such a large change would be caused by either an unusually powerful factor or an unusual confluence of established factors.

The peak year in sales happens to occur just as Napster comes into the picture in September of 1999. Although there doesn't appear to have been a tremendous growth in downloading beyond that achieved during Napster's heyday, CD burners were becoming continuously less expensive and more popular during the following years.

The pattern of file-sharing's birth and rapid growth followed immediately by the unusually large decline in the sound recording market supports a claim that file-sharing is responsible for the decline in sales. Add to this the predictions of economic theory that file-sharing should lead to a decline in sales and we have what appears to be a very strong case that file-sharing is the cause of the decline.

### ***B. Market Turnaround and the RIAA Lawsuits against File-Sharers***

The 2004 RIAA half-year numbers indicate that sales of albums are up for the first time since 1999. Overall unit sales of albums are up by 5.1% from the first six months of 2003.<sup>56</sup> Revenues, on the other hand are up only 3.9% in nominal terms and 1.5% after inflation because real record prices dropped by 3.5%.<sup>57</sup>

Does the information revealed in this sudden, albeit small, turnaround for the sound recording industry represent a departure from the negative linkage between file-sharing and industry revenues? The answer, obviously, depends on whether this turnaround in record sales was matched by a turnaround in file-sharing.

---

<sup>55</sup> 1991 is almost as low, but this still would be 13 years.

<sup>56</sup> The RIAA publicized a more optimistic statistic, which was that unit sales to retail outlets were up by 10.2% (10% for full length albums) but I am using statistics for the complete sound recording market (including record clubs and other direct sales not through retailers).

<sup>57</sup> This decline, using RIAA list prices, is consistent with the reported transaction price decline measured by NPD (4%) during the first quarter of 2004 relative to the prior year period. See New York Times "Most Wanted: Drilling Down/Compact Discs; Falling Prices" Ian Austen, Section C, Page 9. Universal Music, the largest record company, lowered its list prices by 25% in the second half of 2003, but there was much resistance among retailers about the details of the decline and the price declines were often not passed on to consumers. See "CD prices hit sour note with retailers, buyers" Michael McCarthy; USA TODAY; Dec 8, 2003; pg. B.01.

The evidence, once again, supports the hypothesis that file-sharing is detrimental to the sound recording industry. This is because all the data sources discussed above agree that file-sharing has declined somewhat from the first half of 2003 to the first half of 2004 whereas it appears to have been increasing prior to that period. Table 3 below reports the amount of the declines.

| Table 3: Estimated Change in File Sharing, first half of 2003 to first half 2004 |        |
|--|--------|
| Big Champagne  | -4.4%  |
| ComScore   | -36.3% |
| PEW  | -28.8% |

Given the wide variation in the trends and amount of measured file-sharing usage, it is no surprise that there is quite a variation between these estimates. As discussed below, the main reason for this decline appears to be the file-sharing lawsuits brought by the RIAA in the second half of 2003. One might expect, therefore, that the Pew survey results would overstate the true decline as cautious file-sharers become reluctant to truthfully reveal the extent of their activities. ComScore might also be expected to overstate the decline since they might not track new file-sharing programs which promise greater anonymity to downloaders.<sup>58</sup>

The impact of the RIAA lawsuits has been a policy issue of interest in its own right. A brief history of these lawsuits goes like this. The recording industry announced, with great fanfare, plans to bring lawsuits against file-sharers on June 25, 2003.<sup>59</sup> On September 8, 2003, the RIAA brought what they referred to as the “first wave” of lawsuits against 261 individuals.<sup>60</sup> Since then there have been additional waves of lawsuits which in the fall of 2004 number almost five thousand. Since the RIAA had no reasonable likelihood of bringing lawsuits against each of the millions of file-sharers, the mere threat of a lawsuit needed to be sufficient to convince users to stop file-sharing. Therefore, these lawsuits were brought in an attempt to garner maximum attention and publicity. According to news reports, these lawsuits generally are settled with a payment to the RIAA of several thousand dollars.

For the purposes of measuring the impact of the RIAA lawsuits on file-sharing, there are several characteristics we would wish for in any data set being used:

---

<sup>58</sup> Although see the caveat in footnote 32.

<sup>59</sup> On January 21, 2003 a judge ruled against Verizon’s attempt to reject an RIAA request, under the Digital Millennium Copyright Act, to identify a particular subscriber of Verizon’s ISP service. While this was under appeal the RIAA moved to quickly take advantage of the ruling to inexpensively acquire the names of various file-sharers. A later decision in Verizon’s favor came on December 19, 2003.

<sup>60</sup> This chronology is taken from the RIAA website <http://www.riaa.com/news/newsletter/press2003.asp>.

- a. Data must be limited to the US since RIAA lawsuits only affect American users
- b. Data need to compare periods before and after the lawsuits
- c. Since RIAA lawsuits were limited to music files, the number of transferred *music* files is the datum of interest, not overall peer-to-peer usage.

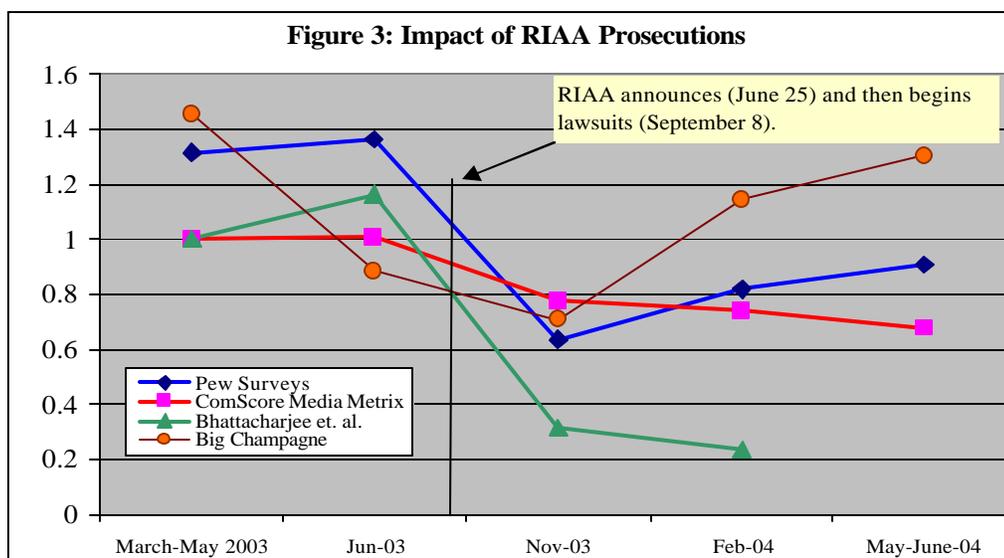
The NPD data meets condition (a) and (c), but not (b), making it of no use for measuring the impact of the lawsuits. Big Champagne and comScore both meet conditions (a) and (b), but not (c). The Pew survey data meet all three conditions. BGLM meets only condition (b).

There has been some debate about the efficacy of those lawsuits. Some have argued that there has been a considerable decline in usage coinciding with the lawsuits, while others have argued that the lawsuits have had virtually no impact (Karagiannis et. al., 2004).<sup>61</sup>

ComScore, Pew, and BGLM all show that file-sharing was level or increasing prior to the lawsuits and that there was a sharp decline after the lawsuits were announced and commenced.. This can be easily seen in Figure 3 (and also in Figure 1). Using monthly data, ComScore and BGLM both show extremely close timing between the announcement of lawsuits and the decline. Pew doesn't allow that level of precision. Big Champagne data, on the other hand, indicate that the decline in file-sharing began before the announcement of the lawsuits. The very large drop in file-sharing reported in the Big Champagne data is mysterious, but the continued decline after the lawsuits is consistent with a negative impact of the lawsuits on file-sharing activities.

---

<sup>61</sup> For an overview see John Borland "RIAA lawsuits yield mixed results" December 4, 2003, CNET News.com, which provides quotes from both sides, available at [http://news.com.com/RIAA+lawsuits+yield+mixed+results/2100-1027\\_3-5113188.html](http://news.com.com/RIAA+lawsuits+yield+mixed+results/2100-1027_3-5113188.html) . The Karagiannis et. al. paper does not limit its analysis to the US and bases its results on a single one hour observation prior to the lawsuits and a single hour observation after the lawsuits. Despite their attempts to make some calculations to control for variations, these measurements are too sparse, in my opinion, to allow any conclusions to be drawn.



Note that Big Champagne and Pew show file-sharing beginning to grow again during 2004 (consistent with NPD) although current levels are not yet back to the peak pre-lawsuit levels. ComScore and BGLM show a continuing decline in file-sharing. If these data sets continue to provide divergent measurements, it will make it difficult to test future linkages between file-sharing usage and record sales unless we can crown one the winner.

Overall, the evidence appears to be that the lawsuits had an initial negative impact on file-sharing, as hoped by the RIAA, but after this initial decline the impact of the lawsuits appears to be wearing off. For our main purpose of determining the relationship between file-sharing and the sales of sound recordings, however, we do not need to know why file-sharing declined from the first half of 2003 until the first half of 2004, merely that it did.

#### IV. Investigating the Background Noise

There are obviously other possible factors that could have had an impact on record sales. Although a natural outlier will submerge background noise I cannot claim that the magnitude of the sales decline is large enough to make the background noise irrelevant. I have examined some of these alternative factors in prior work (Liebowitz 2004) and will summarize much of that work here. In order for these alternative variables to be able to ‘explain’ or to have caused the rapid change in record sales that began in 2000, these other variables would need to have markedly changed at approximately the same time.

The variables that I examined were price, income, music quality (measured by concerts and radio listenership), the markets for substitutes and complements, the opening of new markets, and whether consumers might have stopped replenishing their audio cassette library with CD versions of the same music. That paper, in summary, found that list prices, adjusted for inflation, have been virtually constant for the last decade, ruling out price as an explanation of the sales decline; real GDP is only weakly related to record sales and the mild recession of 2001 was insufficient to account for even a small part of the four year decline; trends in movie box office receipts and videogame receipts did not change on or around the year 2000; the increased portability of prerecorded music brought about by audiocassettes (and CDs) helped lead to a large and sustained increase in sales of prerecorded music; and there was no noticeable impact of librarying.

Some researchers, such as Koleman Strumpf, have suggested that DVD growth is the primary suspect for the fall in CD sales.<sup>62</sup> In Liebowitz (2005) I noted that prerecorded movie sales (measured in units) per capita had evidenced a sharp acceleration beginning in 2000, consistent with this hypothesis. I also noted, however, that there was a similar although somewhat less steep increase in the per capita units of prerecorded movies in the early 1990s, a period of robust growth in the CD market, and suggested this inconsistency weakened or ruled out the DVD hypothesis.<sup>63</sup> I used units of DVD and prerecorded videocassettes in the earlier paper since these data were publicly available on the MPAA web page.

Since then I have purchased data on both sales and *rentals* of prerecorded movies, which should allow for a more detailed understanding of the market for prerecorded movies.<sup>64</sup> It turns out that the increase in DVD unit sales that has occurred in the last few years came largely at the expense of the video rental market. The data are represented in Figure 4.

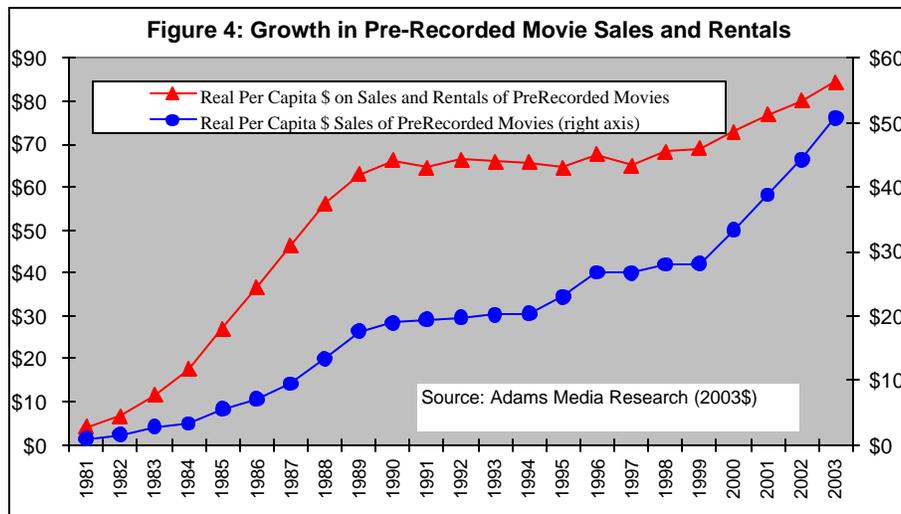
---

<sup>62</sup> See, for example, this story in the July 22, 2004 edition of the Guardian "Listen to the flip side" by Suw Charman. She reports asking Professor Strumpf what caused the decline in CD sales: "Over the period 1999 to 2003, DVD prices fell by 25% and the price of players fell in the US from over \$1,000 to almost nothing," says Strumpf. "At the same time, CD prices went up by 10%. Combined DVD and VHS tape sales went up by 500m, while CD sales fell by 200m, so a possible explanation is that people were spending on DVDs instead of CDs." The article is available at <http://www.guardian.co.uk/online/story/0,,1265840,00.html>.

<sup>63</sup> This can be seen in Figure 2. In the 2004 paper I also examined the possibility that the constraint that binds consumers might be one of time and not one of income. Evidence on the time spent viewing prerecorded movies, however, indicated that moderate changes in such viewing time would have a very small impact on music listening, which takes up considerably more time on the part of consumers.

<sup>64</sup> Prerecorded movies might substitute for CDs in a budget constraint for entertainment expense, which is most appropriately represented by expenditures made for prerecorded movies. It might also substitute in a time constraint, the time taken to view movies, where units would be appropriate. It is not clear that either DVD units or DVD expenditures have much impact on CD sales, but the possibility of DVDs impacting the budget constraint seems to have a greater affinity for economists.

Figure 4 compares real per capita consumer expenditures on prerecorded movies from 1981 until 2003, broken down by rentals versus sales. We know that record sales began their steep dive in 2000. That matches nicely with a sharp increase in expenditures on the sale of prerecorded DVDs and VHS tapes beginning in the year 2000, as represented by the lower line with circles as the markers (with the amounts read off the right axis). Expenditures on both video sales and rentals (the line with triangle markers, with values read off the left axis), by way on contrast, show little in the way of such an acceleration the year 2000. What happened, as the arithmetic tells us, is that rental expenditures have not been keeping pace with sales (in fact they have fallen 18% since 1999). Since consumer expenditure on video rentals should be highly substitutable for expenditures on video sales, it turns out to be very misleading to focus only on the sales of prerecorded video.



Further, the increase in the combined video expenditure, whether in percentage or absolute terms, is much greater in the mid to late 1980s (1983-1989) than it was in the period after 2000, yet not only was there was no pronounced decline in record sales during the mid 1980s, there was a fairly moderate increase. Thus, the evidence from the VHS/DVD market does not support a claim that the recent growth in DVD expenditures can explain the fall in CD sales.

## V. An Econometric Analysis

I have argued that because of the popularity of file-sharing and the size of the change in industry sales, the approach above provides a great deal of information. Nevertheless, there are alternative approaches. One alternative approach is to examine a cross section of locations and see how file-sharing is related to the purchase of sound recordings. Although it is possible to use countries as the

unit of analysis (Zentner 2004, Peitz and Waelbroek 2004), comparing metropolitan areas in a single country such as the United States would seem to allow easier control of variations in terms of income, tastes, indigenous music industries, and social mores.

I was fortunate to have available a data set created by Eric Boorstin, an undergraduate at Princeton who had written a senior thesis in the spring of 2004.<sup>65</sup> Boorstin used Census data on the number of individuals in major metropolitan areas with and without Internet access and merged them with SoundScan data for the sales of CDs in each metropolitan area. A novelty in this data comes from the fact that it separates individuals into various age groups so that, in principal, the analyst can determine the impact of file-sharing for different age cohorts.

The amount of file-sharing within a city or cohort is not directly measurable. Internet use, therefore, is taken as a proxy for file-sharing. Certainly, file-sharing is limited to those with Internet access, but there are likely to be other differences between those with and without Internet access that might bias any estimates using this proxy. In particular, individuals with Internet access are likely to have greater wealth (and better education) than those without such access, which was the basis for the concern surrounding the debate on the so-called “digital divide” in Internet use. The Census data supports this view since cities with higher incomes also have a greater share of individuals in each age group who have Internet access, as Table 4 makes clear.

| Age Group | Share of individuals without Internet access | Share of individuals with Internet access |
|-----------|--|---|
| 10-24     | -0.3321                                      | 0.4063                                    |
| 25-44     | -0.2866                                      | 0.4875                                    |
| 45-64     | -0.3534                                      | 0.3797                                    |
| 65+       | -0.4483                                      | 0.1433                                    |

Yet there are no variables available to control for these socioeconomic differences between Internet users and non-Internet users. Thus we would expect the coefficient on the file-sharing proxy to be biased upward. The Special Census data cover the years 1998, 2000, and 2001. By ending in 2001, the data contain only two of the four recent years of declining sound recording sales.

---

<sup>65</sup> I want to thank Mr. Boorstin for graciously making his data available to me. Although I suggest that a specification error caused him to reach an inappropriate conclusion, I do believe that his paper was a tremendous accomplishment, particularly for a senior thesis. Except for some changes in specification and variable measurements I am largely following the methodology that he pioneered.

In table 5 I report regression results using a fixed-effects model, with CD sales per capita as the dependent variable. Independent variables include the average income per capita as well as the share of the population broken down by age group and Internet access. SoundScan data, being based only on retail outlets, are likely to understate somewhat the true impact of file-sharing since non-retail sales appear more susceptible to the impact of downloading as discussed above.

The impact of file-sharing on an age cohort is measured as the difference between the coefficients of the age cohort with Internet access and the coefficient for the cohort without Internet access. This difference is calculated in the column labeled “Internet Differential” and includes a test of significance for the hypothesis that the difference is zero.

| <b>Table 5: Age Groups Based on Share of Population; Dependent Variable = Per Capita Sales per city; n=297</b> |           |           |      |                       |              |
|--|-----------|-----------|------|-----------------------|--------------|
| Share of Population  | Coef.     | Std. Err. | t    | Internet Differential | Significance |
| 10-29 no Internet  | 2.863291  | 1.81591   | 1.58 |                       |              |
| 10-29 Internet   | 0.2094592 | 1.570257  | 0.13 | -2.6538318            | 0.1024       |
| 30-44 no Internet  | 1.751894  | 1.8017    | 0.97 |                       |              |
| 30-44 Internet   | 2.658693  | 2.121496  | 1.25 | 0.906799              | 0.606        |
| 45-64 no Internet  | 3.471715  | 1.528949  | 2.27 |                       |              |
| 45-64 Internet   | 5.396905  | 1.691543  | 3.19 | 1.92519               | 0.2006       |
| 65+ no Internet  | 4.859013  | 1.651844  | 2.94 |                       |              |
| 65+ Internet   | 5.962898  | 2.43968   | 2.44 | 1.103885              | 0.6637       |
| City Income  | 0.0000158 | 8.32E-06  | 1.9  | Number of Obs         | 297          |
| Constant   | 0.1774351 | 1.12407   | 0.16 | Adjusted R-squared    | 0.93         |

The results have a certain intuitive appeal, but also some problems. For youthful individuals, file-sharing as measured by Internet usage is deemed harmful to sales as indicated by the large negative coefficient (although only marginally significant) in the Internet differential column. The Internet differential coefficients for older age groups are positive and of smaller absolute value, although only the 45-64 age group comes close to achieving statistical significance.

The interpretation of the Internet differential is fairly straightforward. For the 10-29 age group the Internet differential of -2.65 implies that a complete shift from no Internet usage to full Internet usage by members of this group would lower sales of sound recordings by 2.65 units per person. The coefficients for older age groups might be taken to imply that file-sharing increases record sales but before we accept these values we need to exam the possible bias from income differentials.

As already explained, because these regressions do not account for income differentials by Internet access, we should expect that the Internet differential will be biased upward due to the higher

incomes of Internet users. This bias thus becomes a plausible explanation for the positive measured impact of file-sharing for the older age groups when there is no reason to expect a positive coefficient.

| Table 6: Percentage of Adult Internet Users answering yes to question: "Do you ever download music files onto your computer so you can play them at any time you want?" Source: PEW |             |             |        |        |
|---|-------------|-------------|--------|--------|
|   | July-Aug-00 | Aug-Sept-01 | Oct-02 | Nov-03 |
| Overall   | 22%         | 26%         | 32%    | 14%    |
| 18-29   | 37%         | 48%         | 54%    | 28%    |
| 30-49   | 19%         | 23%         | 29%    | 13%    |
| 50-64   | 8%          | 10%         | 15%    | 6%     |
| 65+   | 14%         | 10%         | 19%    | 7%     |

According to the PEW figures reported in Table 6, the share of those with Internet access admitting to downloading was approximately the same for the two groups of age 50-64 and age 65+, and the proportion of Internet users for these groups was considerably less than for the younger groups. Thus the file-sharing impact should be much smaller for the individuals over the age of 50 and we are likely to see a relatively stronger impact of the income bias for this group. For this reason, one might argue that the groups with the oldest individuals can be a benchmark of the upper limit for the income bias in the Internet differential. If so, it is possible that the income bias in the Internet differential variable might be as high as 1.5 CDs per person, if the two oldest age groups are averaged together.

We can combine the information in Table 5 with two other pieces of information to calculate an estimate of the overall impact of file-sharing on sound recording sales in 2001. The first required piece of information is the share of each age cohort in the purchase of sound recordings in 2001.<sup>66</sup> The second piece of information is the share of each age cohort using the Internet in 2001.<sup>67</sup> By multiplying these pieces of information together we can derive the overall impact of file-sharing (Internet Access) for each cohort and then sum the total to derive an overall impact.

These calculations are undertaken in Table 7 which performs two sets of calculations, one for the unadjusted Internet differentials and a second for the adjusted Internet differentials.

---

<sup>66</sup> The share of purchase information comes from the RIAA and is based on polls taken during 2001 by Peter Hart Research. More information is available at the RIAA website [www.riaa.com](http://www.riaa.com)

<sup>67</sup> The share of Internet usage is from PEW although PEW only reports figures for the 18-29 age group which I apply to the entire 10-29 group. Similarly, they provide figures for the age groupings of 30-49 which I attribute to the 30-44 group and PEW provides data on the over 50 group which I attribute to those over 45.

| Table 7: Impact on Sales from File Sharing (Internet Usage) |                                 |                                     |                                   |                                |                                 |                               |
|---|---------------------------------|-------------------------------------|-----------------------------------|--------------------------------|---------------------------------|-------------------------------|
| (a)<br>Age Group  | (b)<br>Share of Music Purchases | (c)<br>% using Internet on Aug 2001 | (d)<br>Unadjusted Internet Effect | (e = b*c*d)<br>Impact on Sales | (f)<br>Adjusted Internet Effect | (g =b*c*f)<br>Impact on Sales |
| 10-29   | 45%                             | 75%                                 | -2.65                             | -0.886                         | -4.15                           | -1.388175                     |
| 30-44   | 31%                             | 68%                                 | 0.91                              | 0.191                          | -0.59                           | -0.1235696                    |
| 45+   | 24%                             | 39%                                 | 1.5                               | 0.140                          | 0                               | 0                             |
| <b>Average Impact Weighted by Share of Purchases</b>        |                                 |                                     |                                   | <b>-0.556</b>                  |                                 | <b>-1.512</b>                 |

We can walk through the analysis, beginning with the 10-29 year old cohort. According to Pew, approximately 75% of individuals in the 10-29 age group had Internet access in 2001. The Internet differential coefficient is an estimate of the impact of Internet access on per capita CDs sales if all individuals in the cohort used the Internet. Thus, the product of these variables provides an estimate of the impact on sales for the members of the cohort who have Internet access. That would be a decline in CDs of approximately 2 units per person in the cohort. This value is then multiplied by the 45% share of the sound recording market made up of consumers between 10 and 29 years of age to arrive at an overall impact of file-sharing on the market due to this cohort, which would be a decline of .886 units.<sup>68</sup>

The overall market impact is merely the sum of the three age cohorts, seen to be -.556 for the unadjusted Internet differentials. This decline in sound recordings is just slightly less than the per capita decline in record sales that had occurred as of 2001 (.68 from 1998 and .88 from 1999).

The last two columns redo this calculation on the assumption that the 45+ cohort is the baseline measuring the income bias in the Internet coefficient. With this adjustment the Internet differential changes to -4.15 for the 10-29 age group for example, implying that Internet usage would lower yearly sales to this group by slightly more than 4 CDs per capita. The overall impact for all age groups would be a loss of 1.5 CDs per capita, which is considerably larger than the actual per capita decline that had occurred by the year 2001. Even though the decline caused by file-sharing could in fact be larger than the measured decline in CD sales, since sales might have increased from their 1999 baseline if not for file-sharing, it is very unlikely that sales would otherwise have increased by 1.2 units per person in the two year period since there is no other two year period with such a large increase. Thus this estimate is obviously too high. It appears that using the 45+ age group as a baseline estimate of the income bias for individuals using the Internet overstates the true bias.

<sup>68</sup> The statistics on sound recording purchases include singles as well as albums, possibly overstating the impact of teenagers, but singles are currently a very small part of the market.

Nevertheless, these results are consistent with the conclusion that most, all, or more than the entire decline in record sales is due to file-sharing.

What about the inconsistencies in the results to which I alluded above?

The main problem has to do with the larger coefficients for older age groups. It would appear, for example, that the larger the portion of the population consisting of those older than 65, the larger the sales of CDs in a city. Yet we know from column (c) of Table 7 that young adults are more intensive purchasers of CDs than are older individuals, although since this is based on surveys it might be less accurate now than it was before file-sharing.

One possible explanation comes from the fact that the coefficients for each age group hold constant the share of the other age groups. If you increase the share of those over 65 it is possible that the ratio of grandparents to grandchildren rises. This might increase overall CD sales because grandparents give their teenage grandchildren money to purchase CDs. When you increase the share of individuals between the ages of 25-45 this implies that there are fewer children in the typical family or fewer families. That might free up additional resources to allow young adults to purchase CDs. This effect could be smaller than the grandparent effect. Alternatively, it might be that younger individuals, even those without Internet access, have mechanisms for indirectly receiving downloaded files through less sophisticated transmission mechanisms, such as the transfer of a physical CD. In this case the large scale pirating of music files would overflow into the non-Internet groups, and this would presumably occur largely between individuals in the same age categories. This would tend to reduce the overall coefficient for younger users (although it seems unreasonable that it would be smaller than for older age groups) and also would work to reduce the Internet differential coefficient in the above regressions. Or, the share of CDs purchased by young individuals might no longer be accurate.

This is not to say that any of these explanations is correct, but merely that one can construct possibilities where the pattern of regression coefficients does not necessarily contradict common sense.

## **VI. Other Evidence**

There are several other pieces of evidence that are consistent with the hypothesis that file-sharing has negatively impacted the sales of CDs. I report on two here, but acknowledge that the evidence is not sufficiently strong to place much confidence in these results.

The first has to do with the impact of file-sharing on different genres of music. In earlier work I reported on the great yearly variability in the sales of CDs by genre and suggested that the variability was too high to allow much confidence in the results from such analysis. Nevertheless, I feel compelled to mention these results because they are of a piece with the other data reported here, and because of contrary claims that have been made about these data. Oberholzer and Strumpf state in a brief before the Supreme Court that “musical genres which are not heavily downloaded [country and catalog] on file sharing networks experienced the same reduction in sales as other genres.”<sup>69</sup>

|      | Total   | Alternative | Classical | Country | Metal  | Jazz   | R&B     | Rap     |
|------|---------|-------------|-----------|---------|--------|--------|---------|---------|
| 1994 | 614,669 | 82,164      | 27,003    | 75,976  | 38,739 | 16,546 | 80,819  | 40,995  |
| 1995 | 616,172 | 94,004      | 23,836    | 76,095  | 31,101 | 14,797 | 80,718  | 41,537  |
| 1996 | 616,642 | 105,175     | 21,456    | 66,883  | 26,409 | 21,794 | 74,035  | 56,343  |
| 1997 | 651,978 | 106,690     | 19,148    | 70,702  | 28,983 | 20,042 | 141,613 | 61,709  |
| 1998 | 728,268 | 116,489     | 16,948    | 74,043  | 30,086 | 18,123 | 166,379 | 83,641  |
| 1999 | 754,835 | 120,952     | 17,311    | 69,300  | 82,698 | 19,557 | 175,339 | 87,663  |
| 2000 | 785,138 | 131,138     | 16,403    | 67,115  | 89,924 | 18,416 | 197,141 | 105,515 |
| 2001 | 762,781 | 131,594     | 15,846    | 67,241  | 88,158 | 19,514 | 195,498 | 89,279  |
| 2002 | 680,960 | 125,752     | 14,776    | 75,362  | 74,677 | 19,901 | 160,183 | 83,346  |
| 2003 | 656,293 | 128,344     | 17,727    | 70,944  | 74,629 | 22,366 | 149,972 | 75,854  |

Here are the raw SoundScan numbers for various genres. My concern over their reliability can be seen by looking at the ‘Metal’ category in 1999 or ‘R&B’ in 1997. There are enormous jumps that dwarf normal market changes and should make us wary about the reliability of testing hypotheses using such data. Nevertheless the claim that ‘Country’ has fallen as much as overall sales is not true. In 2003 Country sales were actually higher than was the case in 1999 or 2000.<sup>70</sup>

More to the point, Oberholzer and Strumpf (2004) report data that, if reflective of the entire market, allow somewhat greater precision in determining which categories are most heavily downloaded. Using numbers from their Table 4 and Table 8 it is possible to construct a ratio of downloads to sales for the albums in their sample, by genre. Those ratios are reported below. The results indicate that in a relative sense ‘hard’ rock is downloaded the most and jazz the least. The category ‘catalogue’ is quite high, contrary to the claims of Oberholzer and Strumpf. I am surprised that rap is as low as they find and this makes me suspicious of the generality of their downloading statistics.

<sup>69</sup> Page 12 of their Brief of Felix Oberholzer-Gee and Koleman Strumpf as Amici Curiae in support of Respondents, *MGM v. Grokster*, No. 04-480. More specifically, they claim that two categories of music (Catalog and Country) are not heavily downloaded yet sales have fared poorly. I do not have data on catalog, which is a term used to describe old music.

<sup>70</sup> Note that SoundScan data show the peak year in sales to be 2000 whereas the RIAA numbers show it to be 1999.

| genre       | ratio downloads/sales |
|-------------|-----------------------|
| hard        | 2.303396429           |
| catalogue   | 1.251758653           |
| alternative | 0.815644682           |
| Soundtrack  | 0.607200861           |
| new         | 0.524594136           |
| R&B         | 0.516292044           |
| Rap         | 0.376237875           |
| Current     | 0.350524322           |
| Country     | 0.296733035           |
| latin       | 0.285830303           |
| jazz        | 0.172805005           |

In order to properly compare sales by genre with downloading intensity we would need to control for listening habits. In the following table I provide this control by dividing sales by the share of radio listenership.<sup>71</sup>

|      | classical | country | jazz/smooth jazz | urban/rap | metal/alternative | total    |
|------|-----------|---------|------------------|-----------|-------------------|----------|
| 1998 | 7.73      | 6.04    | 4.68             | 7.91      | 3.07              | 564.5488 |
| 1999 | 8.07      | 5.82    | 5.46             | 8.84      | 12.67             | 589.7148 |
| 2000 | 6.98      | 6.10    | 5.27             | 10.80     | 14.40             | 634.7114 |
| 2001 | 8.41      | 6.36    | 5.54             | 8.54      | 14.81             | 627.8033 |
| 2002 | 8.59      | 7.41    | 5.50             | 7.67      | 11.96             | 564.1756 |
| 2003 | 10.55     | 6.80    | 6.11             | 6.31      | 12.76             | 546.9108 |

The first three columns contain genres that have low downloading ratios.<sup>72</sup> These categories have all experienced increases in sales relative to radio audience since 1999/2000. The last column represents the category that O&S report is most heavily downloaded. It shows a decline in sales relative to audience. I include urban/rap since I would have expected it to be heavily downloaded although O&S do not find it to be so. It also experienced a decline.

Finally, there appears to have been a change in the decay rates of records staying on the charts (Bhattacharjee et. al.). Since 2000, they appear to drop off the charts much more frequently. Since it takes some time for the files to appear in quantity on file-sharing networks, this is consistent with file-sharing causing damage. There is ongoing work to more precisely measure this impact.

## VII. Conclusions

File-sharing is the newest and most publicized technology lending itself to unauthorized copying. The digital element of copying, and the fact that it is undertaken through computers—devices capable

<sup>71</sup> Radio genre categories do not match perfectly with the SoundScan categories. In this table I report the radio/categories.

<sup>72</sup> There are no numbers for classical in Oberholzer and Strumpf but I think we can be confident that classical music is not heavily downloaded both because of the age of its listeners and because of the length of the passages.

of measuring the activity we wish to analyze—seemed as if it would provide economists with a laboratory to finally allow us to analyze and understand the impacts of copying with a degree of precision that was previously missing. Unfortunately, the promise of quality data to use in this laboratory of copying has been, to date, more a mirage than a reality. There are plenty of numbers, but so far the internal consistency of these numbers has been so low that we are incapable of stating to within a factor of ten even such a modest fact as the number of files being downloaded.

Nevertheless, there is strong evidence that the impact of file-sharing has been to bring significant harm to the recording industry. The basic evidence in the United States over the last few years—the birth of file-sharing and the subsequent decline in CD sales—makes for an extremely compelling and simple explanation in spite of the protestations to the contrary from the large and vocal group of individuals supportive of file-sharing. The recent reversal in the decline in CD sales matches a reversal in the activity of file-sharing, providing additional support for this conclusion.

This conclusion is not likely to have been a surprise to most anyone, prior to this topic becoming so highly politicized. The basic intuition of most economists is not much different than that which occurs to members of the general population: when given the choice of free copies versus purchased originals, a significant number of individuals who might have purchased originals will chose to substitute the free copy. It would be amazing if there were not a strong substitution effect.

Although there are conditions which might work to mitigate or even overturn this theoretical expectation, those conditions are unlikely to occur in the case of file-sharing. Although the concept of ‘sampling’ has been mentioned as a possible mitigating factor, theory does not appear to support this surmise. A broad analysis of the various theoretical factors at work supports a view that file-sharing is likely to cause damage to the owners of copyright materials that are so shared.

It is also important to keep an open mind for direct and potentially powerful tests that might not require highly sophisticated econometric tools. I have suggested that such a test has been provided in the case of file-sharing. This does not preclude more traditional tests, and the regression analysis performed in this paper comes to the same conclusion about the deleterious impact of file-sharing on sound recording sales. Nevertheless, if forced to choose between the methodologies, in this case, I would suggest that simple linkage between sound recordings and file-sharing is the more compelling evidence.

As must always be the case, there are still many questions unanswered and unaddressed. Understanding how particular markets function has always been a very difficult undertaking. Economists look at markets from the outside. We generally do not know the thousands of institutional details known by those within the industry. We generally cannot run controlled experiments. What economists bring to the table is a broad understanding of economic phenomenon and our toolkit of economic and statistical techniques. Our tools have power, but they generally are not powerful enough to allow the elusion of humility when making prescriptive or proscriptive policy choices. It is always dangerous to make strong statements about markets that are in transition because the future may always prove you wrong.

In closing I can only repeat what is generally understood but all too frequently ignored, which is that before drawing inferences from data, whether file-sharing or any other economic activity, we need to carefully examine the data that we are provided. Economists are often merely consumers of data but particularly in markets that are new, or where data vendors have not yet created and demonstrated the value of their methodologies, we need to heed the *caveat emptor* warning before we accept the numbers that we are all so eager to put to use.

- Becker, Lee and Klaus Schoenbach editors Audience Responses to Media Diversification, LEA, Hillsdale, New Jersey 1989. Olderaan Frank and Nick Jankowski “The Netherlands: The Cable Replaces the Antenna” and Bouillin-Dartevelle R “Belgium: Language Division Internationalized.”
- Bhattacharjee Sudip, Gopal Ram D., Lertwachara Kaveepan, and Marsden, James R. “Impact of Legal Threats on Individual Behavior: An Analysis of Music Industry Actions and Online Music Sharing Activity” University of Connecticut working paper, 2004.
- Blackburn, David “Does File Sharing Affect Record Sales?” Working paper, April, 2004.
- Boldrin M, Levine DK “The case against intellectual monopoly” *International Economic Review* 45 (2): 327-350 May 2004
- Boorstin, Eric “Music Sales in the Age of File Sharing” Senior Thesis, Princeton University, April 2004.
- Gopal, Ram D., Sudip Bhattacharjee, and G. Lawrence Sanders “Do Artists Benefit From Online Music Sharing?” *Journal of Business*, forthcoming.
- Harrison, Ann “Are lawsuits intimidating students?” Network World Peer-to-Peer Newsletter, 09/25/03
- Hirshleifer, Jack "Suppression of Inventions," *Journal of Political Economy*, March/April 1971, 79, 382—83.
- T. Karagiannis, A. Broido, N. Brownlee, kc claffy, and M. Faloutsos “Is P2P dying or just hiding?” Technical Report, 2004.
- Klein, Benjamin, Lerner, Andres V., Murphy Kevin M. “Intellectual Property: Do We Need It? The Economics of Copyright ‘Fair Use’ in a Networked World,” *American Economic Review*, 206-208, May 2002
- Lessig, Lawrence *Free Culture*, The Penguin Press, New York 2004.
- Liebowitz, Stan J. “Pitfalls in Measuring the Impact of File-sharing.” CESifo Economic Studies, July 2005. <http://ssrn.com/abstract=583484>
- Liebowitz, Stan J. “Will MP3 downloads Annihilate the Record Industry? The Evidence so Far” *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth*, V. 15, 2004, pp. 229-260. <http://ssrn.com/abstract=414162>
- Liebowitz, Stan J. “The Elusive Symbiosis: The Impact of Radio on the Record Industry” *Review of Economic Research on Copyright Issues* Vol. 1, 2004a, pp.20-45. <http://ssrn.com/abstract=520022>
- Liebowitz, Stan J. *Rethinking the Network Economy*, Amacom, New York, 2002.

- Liebowitz, Stan J. "Copying and Indirect Appropriability: Photocopying of Journals," *Journal of Political Economy*, 93-5 October 1985, Pp. 945-957.
- Liebowitz, Stan J. "The Impacts of Cable Retransmission on Television Broadcasters," *Canadian Journal of Economics*, 15-3, August 1982, Pp. 503-524.
- Liebowitz, Stan J., "The Impact of Reprography on the Copyright System". Copyright Revision Studies, Consumer and Corporate Affairs, ISBN 0-622-11396-9, Canada 1981, <http://ssrn.com/abstract=250082>
- Lyman, P. and H.R. Varian (2003), "How much Information 2003", [www.sims.berkeley.edu/research/projects/how-much-info-2003](http://www.sims.berkeley.edu/research/projects/how-much-info-2003).
- Michel, Norbert J. "The Impact of the Digital Age on the Music Industry: A Theoretical and Empirical analysis" Working paper 2004.
- Oberholzer, Felix & Koleman Strumpf "The Effect of File Sharing on Record Sales: An Empirical Analysis" working paper, 2004.
- OECD Information Technology Outlook 2004 "Peer To Peer Networks In OECD Countries," Pre-release of Section from Chapter 5 (2004).
- Peitz, Martin and Patrick Waelbroeck "An Economist's Guide to Digital Music" Working Paper, May 2004.
- Peitz, Martin and Patrick Waelbroeck "The Effect of Internet Piracy on CD Sales: Cross-Section Evidence" Cesifo Working Paper No. 1122 Category 9: Industrial Organisation January 2004
- Rainie, Lee, Mary Madden, Dan Hess and Graham Mudd "RE: The impact of recording industry suits against music file swappers" Pew Internet Project and Comscore Media Metrix Data Memo. January 2004
- Rob, Rafael and Joel Waldfogel, "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students", Draft, April 5, 2004
- Siwek, Stephen E., "Copyright Industries in the U.S. Economy: The 2002 Report" 2002, Economists Inc.
- Takeyama, Lisa N. 'The Welfare Implications of Unauthorized Reproduction of Intellectual Property in the Presence of Demand Network Externalities,' *Journal of Industrial Economics*, 42, 1994, pp. 155-166.
- U.S. Congress, Office of Technology Assessment, Copyright and Home Copying: Technology Challenges the Law, OTA-CIT-422 (Washington, DC: U.S. Government Printing Office, October 1989).
- Varian, Hal "Copying and Copyright" *Journal of Economic Perspectives*, forthcoming.
- Vogel, Harold Entertainment Industry Economics Cambridge University Press, 6th edition.

Watt, Richard *Copyright and Economic Theory: Friends or Foes?* Edward Elgar: Cheltenham, 2000.

Watt, Richard “The Past and the Future of the Economics of Copyright” *Review of Economic Research on Copyright Issues*, 2004, vol. 1 (1), pp. 1-11.

Weimann, Gabriel “Cable Comes to the Holy Land: The Impact of Cable TV on Israeli Viewers,” *Journal of Broadcasting & Electronic Media*, 40, 1996, pp. 243-257.

Zentner, Alejandro “Measuring the Effect of Music Downloads on Music Purchases” working paper, April 2004