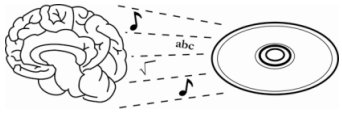


CAPRI

CENTER FOR THE ANALYSIS OF PROPERTY RIGHTS AND INNOVATION

Testing File-Sharing's Impact by Examining Record Sales in Cities

Stan J. Liebowitz

CAPRI Publication 05-02

Stan J. Liebowitz is on the faculty of the School of Management at the University of Texas at Dallas. Contact: liebowit@utdallas.edu.

Napster began a revolution in music listening. The very language used by music consumers has changed. Phrases such as "ripping and burning," "Ipods," and even "MP3" (which predates Napster) can be traced to behavior engendered by Napster. Although Napster was effectively shut down as an unauthorized file-sharing service within two years of its birth (and the fate of Grokster is unclear after the recent Supreme ruling against it), its progeny live on.

File-sharing is merely the most recent example in a long line of technologies that have lowered the cost of unauthorized copying by individuals. Start with the examples of photocopying, introduced in 1959 by Xerox, audio taping, when almost everyone had dual cassette recorders built to make copies, and video taping. It is true that videotaping a movie was more difficult due to copy protection built-in to most prerecorded movies. Floppy discs allowed computer software to be copied.

Each of previous copying technologies engendered cries of alarm from the copyright industries effected. But the harm was not forthcoming. An analysis by Liebowitz (1985) concluded that photocopying had a net beneficial impact upon the industry. After the Supreme Court's Betamax decision allowed home taping, a new market for prerecorded movies emerged which now provides the movie industry with revenues far in excess of box office revenues even though real box office revenues per capita have more than doubled since the decision.[1] To be fair to the industry, an alternative technology supported by the industry, video laserdiscs, did not allow potential pirating and in the absence of VCRs might have brought all the playback advantages of the VCR without the possibility of copying.

Early in its history the software industry expressed similar concerns about unauthorized copying and experimented with anti-piracy devices such as mechanical dongles or discs. The industry soon decided that, for applications aimed at business users at least, the anti-piracy devices did more harm than

[1] 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.

good, and the industry appears now to believe that organized piracy is more destructive and/or amenable to solutions than is personal copying.[2]

In a similar vein, the sound recording industry objected to home audio taping. Alan Greenspan, before he became chairman of the Federal Reserve, stated in testimony before Congress: "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."[3] Just as this testimony was being given, sales of sound recordings were beginning a decade-long advance, once again making the claims of concern by the copyright industry appear unwarranted.[4]

The growth in file-sharing, unlike previous copying technologies, has been accompanied by a large drop in sales. The industry's response has been to bring lawsuits against file-sharing services and thousands of individuals engaged in filesharing. These lawsuits have attracted a good deal of notoriety and discussion.

Each of the previous copy-enhancing technologies brought forth academic work 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.[5] The copying envisioned in this literature, and in this paper, is performed by individuals, as opposed to organized forms of piracy where a criminal entity makes thousands or millions of counterfeit copies which it tries to sell in the

[2] 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.

[3] This is from Greenspan's testimony on the Home Recording Act. Hearings before the Subcommittee on Patents, Copyrights and Trademarks, October 25, 1983.

[4] 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. See Liebowitz (2004).

[5] 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.

market. Hui and Png (2003) examine the impact of organized piracy on the music business and find it to be, not surprisingly, negative.

Supporters of file-sharing suggest that the music industry is 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 appears to have vitiated an increase that might otherwise have occurred.

I. What do we know about file-sharing?

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

Napster came into existence in 1999 with the stated 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.[6] Into the void stepped numerous other file-sharing programs, particularly those which, unlike Napster, were not based on a central server.

By most estimates, file-sharing is enormous, but those estimates are few in number and often at variance with one another. At the high end, there are claims that up to 60 million Americans have used

[6] A&M Records v. Napster, 239 F.3d 1004 (9th Cir. 2001).

peer-to-peer networks,[7] that perhaps as many as five billion music files were downloaded by Americans in a typical month in 2003 (18 files for every man, woman, and child in the country!),[8] and that perhaps 60% or more of all Internet bandwidth is taken up by file-sharing.[9] Webnoize 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.[10] 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 figures that the number of songs being downloaded on file-sharing networks was equivalent to the number of songs purchased in the legitimate market.[11] IDate claims that file-sharing was four times as large as legitimate sales, worldwide, in 2003.

These numbers should be taken with a grain of salt. The methodologies used by IDATE and Webnoize to determine most of these numbers are not reported. Nor does the verbiage surrounding these numbers enhance confidence. One organization which does make its methodology clear, NPD, claims that there were 140-380 million music files downloaded in the United States per month in 2004 and that these values were higher than their 2003 counterparts. The lower figure (for February 2004) is less than *one twentieth* the monthly estimates reported by IDATE for 2003 and less than one fifth the numbers reported by Webnoize for a period two years earlier.

---

[7] 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

[8] According to IDATE, there were either sixty five billion audio files downloaded in the US in 2003 or twelve billion in the world, depending on which of two seemingly inconsistent statements you wish to believe. See http://www.idate.fr/fic/news_telech/102/IDATE_News_282VA.pdf The breathless prose goes on to predict that by 2007 broadband users will download an average of 4300 audio files per year, which seems somewhat outlandish.

[9] 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."

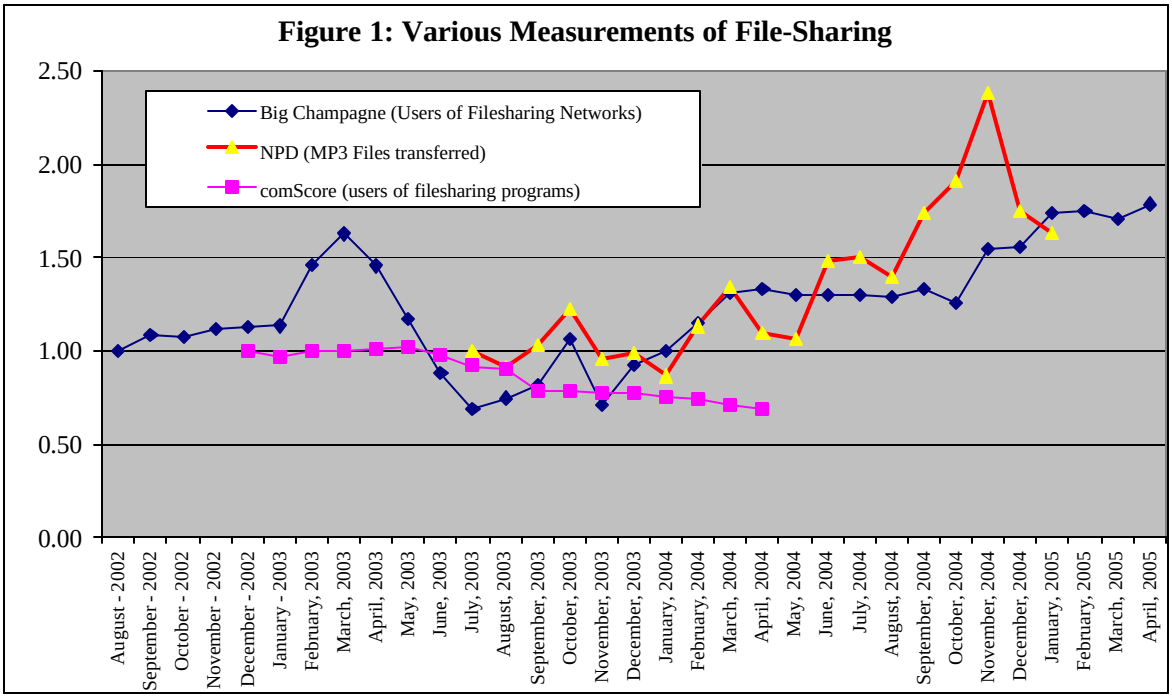
[10] For example, see VNUnet.com 07 Sep 2001 "Music downloads on the rise" by John Geralds, at http://www.infomaticsonline.co.uk/news/1125301.

[11] 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.

Even these low numbers, however, indicate that file-sharing is large, perhaps in the vicinity of one third of the legitimate market.

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. Most estimates indicate that audio files still represent the vast majority of files being shared.[12]

Figure 1: Various Measurements of File-Sharing



An issued related to but different from the size of file-sharing is the trend in such activities. Figure 1 presents the trend, measured in different ways from three different sources. The comScore measure suffers from being limited to an analysis of a fixed set of programs and as such it undercounts the number

[12] IDATE claims the ratio of audio files to video files is 100:1. 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. The OECD (2004) reports (figure 5.7) a most implausible figure taken from Big Champagne that indicates the number of movie files transferred is half the number of audio files in 2003.

of file-sharers.[13] The two sources that are based on a more robust methodology, NPD and Big Champagne, both show that file-sharing has increased in the last few years, but not by a large order of magnitude. Estimates of the magnitude of file-sharing at the time of Napster can be found in Figure Q* below.

When Napster first came into existence, many downloaders would not have had in place the requisite CD burners that would allow listening to 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. For these reasons, MP3 files were not, at the time of Napster, terribly good substitutes for music purchased on a CD.

In the next few years, however, the MP3 audio files that were traded on file-sharing networks became much better substitutes for the music on prerecorded CDs.[14] CD burners became much more common and most new DVD/CD players added the ability to directly play CDs which contained MP3 data files.[15] The rise of MP3 players such as iPods only enhanced the ability to use MP3 files in place of purchased CDs. Thus, file-sharing now produces files that are, for most music listeners, very good substitutes for the purchased CDs.

Naturally, the current concern over the impacts of copying brought forth renewed academic interest. Recent papers include those by Blackburn (2004), Hong (2004), Liebowitz (2004), Michel (2004), Oberholzer and Strumpf (2004), Peitz and Waelbroeck (2004), Rob and Waldfogel (2004), and

[13] The two measures based on specific programs, ComScore and BGLM, tend to overstate the decline if file-sharers are migrating away from monitored programs toward programs that were not monitored, which in fact is what happened. In August of 2005 I was told, in a telephone conversation, that comScore no longer was no longer reporting these measurements because they did not feel they could keep up with the many new downloading programs that were in use.

[14] This ignores the possible dilution of expected quality due to 'spoof 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 spoof 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.

[15] 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.

Zentner (2003, 2006). All of these papers find some degree of harm brought about by file-sharing, except for the paper by Oberholzer and Strumpf.[16]

The empirical work in this paper is based on actual sales of record albums, unlike papers that use surveys of record buyers. It compares sales of records over time, both before and after organized file-sharing. It has data that run through 2003, which is more recent than many previous studies, and has data on the income, by age group and city, of Internet users versus those without Internet access. Finally, it uses cities as the unit of analysis as opposed to some papers which use record albums as the unit of analysis. This avoids the difficulty of having to find instruments in the hope of overcoming the serious simultaneity problems the beset such studies and also avoids potential problems having to do with a potential fallacy of composition.[17] That is not to say the approach below is not without weaknesses—they will be addressed as they appear. First, however, I turn to a brief description of the theory of file-sharing's impact.

II. The Theory of File-Sharing's Impact

In the last few decades we 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.

To start, the unauthorized downloading of a copyrighted file can easily be seen as a substitute for the purchase of that copyrighted work (song or CD). 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 sued or the guilt from not supporting one's favorite artists.

[16] 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).
[17] See Liebowitz (2005) for a more detailed explanation of these potential problems.

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 downloaded song in the same and as wide a variety of locations are can be done with an authorized original.

The 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. 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.

This is how copying was viewed until 1981. At that time several other possibilities were proposed. A different possible impact of file-sharing argues that users might merely use downloaded songs to sample from available music to help guide their purchases. Although this idea can be traced back several decades, and was originally referred to as the *exposure* effect (Liebowitz 1981, 1985) it is now most frequently referred to as the *sampling* effect, as it was in the Napster case.

Sampling is often misunderstood, however. If consumers can at low cost learn about record albums they will make superior choices. They, and society, will be better off, as long as the cost of providing this information is low enough. But record companies are not necessarily better off. Imagine, by way of analogy, a group of consumers who drink beer to get drunk but who have different capacities to hold their alcohol. Assume there is no extra cost for alcohol content and bottles have random amounts of alcohol in them. A high alcohol bottles contain sufficient alcohol to satisfy consumers with a high tolerance to alcohol. If bottles are not labeled as to alcohol content then by the luck of the draw some consumers will get enough alcohol without even finishing their bottles, while others will need to order many low alcohol bottles in an attempt to achieve their desired 'high'. If, on the other hand, the bottles are labeled, each consumer can purchase a bottle that matches his or her tastes. Society is clearly better off when the bottles

are labled. But the number of bottles purchased may decline (or increase) and the beer industry revenues may also decline.[18] Thus sampling might increase revenues, but it might also decrease revenues. It cannot be counted on to offset the substitution effect.

A third claim about the impact of file-sharing concerns the role of network effects. Models such as Takeyama (1994) and Conner and Rumelt (1991) 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.[19] These models make some sense for software. It is easy to imagine that illicit copies of a spreadsheet, say, in the hands of individuals, might increase demand for that spreadsheet. Firms might be the primary market for spreadsheets and firms will have a greater demand for the spreadsheet if their employees already know something about the product. This logic meshes with the fact that college students often receive large discounts on such software, presumably for similar reasons. In order for this argument to hold, however, it must be the case that a large portion of potential employees would be less skilled in using this software without the access to illicit copies.

But application of this idea to music seems a rather remote possibility. The problem is that everyone already knows how to listen to and enjoy music and they had access to a free source of music prior to file-sharing. It is possible that the likelihood of purchase, for some potential purchasers, may be influenced by how many others already listen to it via downloading. Presumably, such network effects are transmitted through conversations between friends. But is it reasonable to think that the album-purchasing portion of society will spend more time listening to music and therefore purchase more albums just because there are other individuals presumably listening to more music because they can get *prerecorded* music for free through file-sharing? People can already listen to as much free music as they want by

[18] This is discussed in more detail in Liebowitz *** The bottle labeling effectively lowers the price of alcohol, which is why society benefits. But if the demand for alcohol is inelastic, total revenue will fall. If beer is a constant cost industry the price of bottles remain the same and producers lose revenue.

[19] 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

listening to the radio (and the average American listens to radio more than three hours per day versus less than one hour per day listening to prerecorded music).[20] So it is unclear that file-sharing will increase the time spent listening and talking about music among that portion of the population that engages in file-sharing. The only empirical evidence, of which I am aware, that touches on this question indicates that when radio broadcasting began in earnest (presumably increasing interest in and time spent consuming music) record sales dropped precipitously (Liebowitz, 2004a) which doesn't support the network effect theory in the case of music.

The final impact of copying that might apply to file-sharing is indirect appropriability. This is a concept coined in Liebowitz (1985). 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 (Klein, Lerner, and Murphy).

The takeaway from this section is that there are at best very weak possibilities of file-sharing not having a negative impact of record sales. The fact that almost everyone knows a relative or acquaintance

[20] See table No. 1119. "Media Usage and Consumer Spending: 1999 to 2007" in the US Statistical Abstract.

(or friend of a child) who seems to have stopped purchasing CDs because of their free availability on file-sharing networks should not be discounted. Enough anecdotes make a census.

III. Estimating the Impact Using a Panel of American Cities

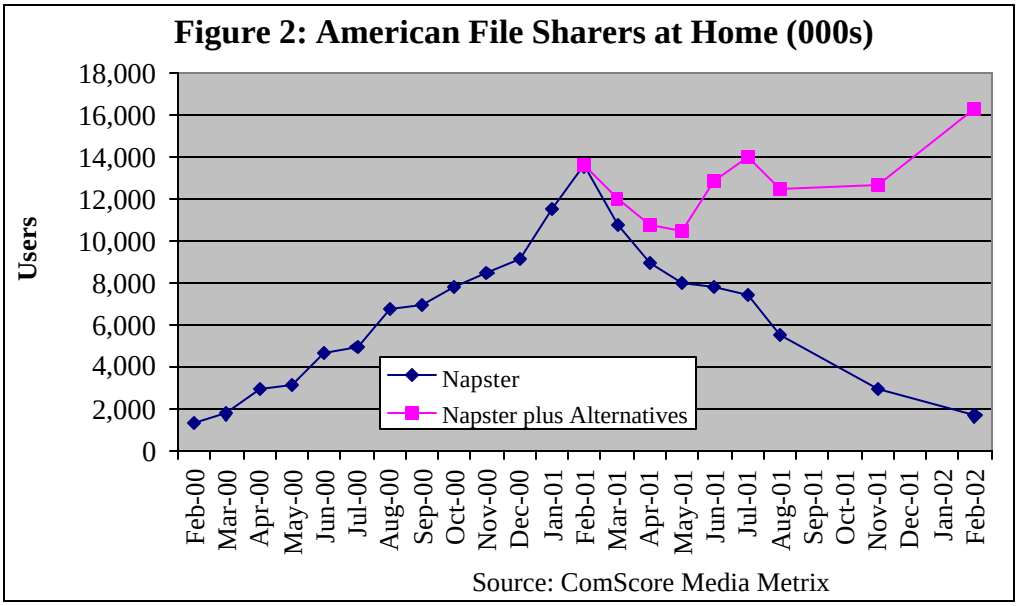
One approach to measure the impact of file-sharing is to examine a cross section of locations and see how file-sharing is related to changes in the purchase of sound recordings. Although it is possible to use countries as the unit of analysis (Zentner 2003, Peitz and Waelbroek 2004), comparing metropolitan areas in a single country such as the United States would seem to allow better control of macro economic variations such as GDP or currency fluctuations, as well as cross-sectional differences in non-Internet piracy and indigenous music industries.

The US Census, as part of it Current Population Survey undertaken for the Bureau of Labor Statistics, conducts surveys on Internet and Computer use. These surveys, conducted in 1998, 2000, 2001, and 2003, ask question of approximately 130,000 individuals. These question provide information on the respondents' geographic location, household income, age, Internet usage, as numerous other variables that seemed unrelated to the analysis.[21] The Census provides information on approximately 250 cities, known as MSAs. Nielsen SoundScan provides information on record sales for 100 areas it refers to as DMAs. The values from the Census MSAs were aggregated to match the Nielsen DMAs and then combined with record data obtained from Nielsen SoundScan to create the data set used in the analysis below.[22] This

[21] This sample is not large enough to provide great confidence about the statistics generated for the smaller cities. For example, the population was divided into four age groups, and further subdivided into whether they had Internet access or not. These age groups were rather wide (10-29, 30-44, 45-64, 65+) and yet for some cities some cells were reported to have zero individuals. Most of these problems were found for the smaller cities. Further, there are many questionable population changes in these data. For example, ten cities show population changes of greater than 20% between 2000 and 2003, with one as high as 64%.

[22] I want to thank Mr. Boorstin for graciously making his data available to me. Liebowitz (2005) used his data and altered Boorstin's regression variables and specifications (including an incorrect criticism of his use of year dummies). Early in 2005 an update to the data, based upon the 2003 survey, became available. In the process of extracting the 2003 data I recreated the data set from the prior years as well. The novelty of the earlier results was that the impact of Internet use on record sales differed by the age of the Internet user, being negative for young Internet users but not for older Internet users. Inconsistencies in the size and patterns of some of the coefficients were noted as a cause of concern. The inclusion of 2003 data should have enhanced the results since the impacts of file-sharing should have been considerably more pronounced. Nevertheless, the entire pattern of the earlier results disappeared when 2003 data were included.

basic methodology was first proposed by Eric Boorstin in a senior thesis at Princeton University in the spring of 2004.



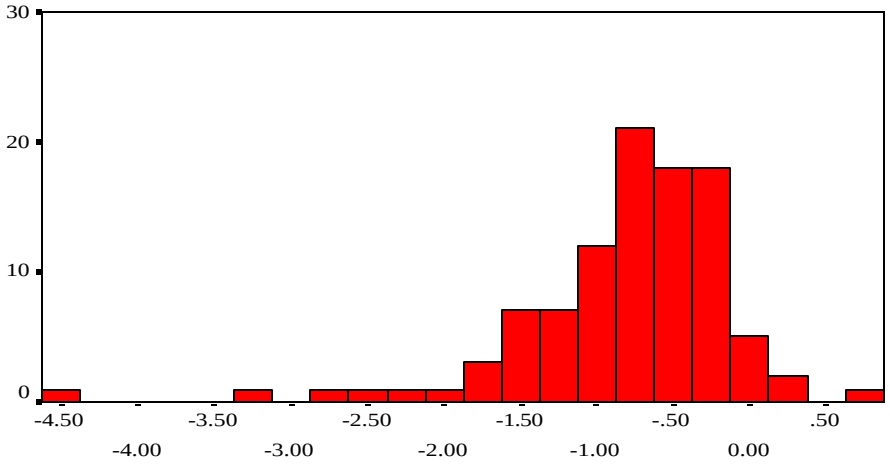
The timing of the Census surveys is important in determining the specification of the regression analysis. The first survey, in December of 1998 clearly preceded Napster and file-sharing. The second survey, in August of 2000, occurs just as file-sharing was gaining steam. The number of Napster users was estimated to be very small in the first half of 2000 but had become approximately seven times as large by the end of 2000, according to ComScore Media Metrix, as shown in Figure 2. According to these measurements, although Napster peaked in February of 2001, file-sharing continued to grow as users migrated to other file-sharing applications. It is not clear how steeply file-sharing was growing when the third survey occurred in September of 2001.[23] The most recent survey, in October of 2003, took place in

[23] Although not shown in Figure 2, the comScore estimate of the number of file-sharers had grown to more than forty million American file-sharers by January of 2003, with over thirty four million users attributed to one program—Kazaa. As discussed above, however, these measures are of questionable reliability, and the sharp change from the third quarter of 2002, where comScore lists Kazaa with slightly under ten million *home* users, to January of 2003 where comScore lists Kazaa with over thirty million users at *home, work and school*, might reflect a change in methodology as much as a change in the number of users and some additional locations, although that is uncertain. ComScore had estimated (for December of 2000) that the number of file-sharers at work to be only about twenty percent of the number of file-sharers at home. According to the US Statistical Abstract there are about fifteen

what appears to be a far more mature and stable file-sharing environment, although file-sharing was still growing, according to our best estimates.

The SoundScan data allow us to examine several questions. One simple question that might arise is whether the national decline in record sales that has occurred reflects a common experience or instead masks wide geographic differences. The histogram in Figure X makes clear that, for the period 2000 until 2003, the decline was broad-based geographically. This figure represents the distribution in the change in record sales per capita. (To put the numbers on the horizontal axis in perspective, note from Table 1 below that the average per capita purchase of albums per year is approximately four). Only 3 out of 100 cities experience an increase in sales per capita. Thus the decline is broad based.

Figure X: Distribution of Album Sales Change 2000-2003



The econometric approach taken here, since the units of analysis are various cities in the US, is to assume that the supply of records, the quality of records, and the price of records is the same in all cities, which appear to be fairly reasonable assumptions. We thus need to focus on demand shifters, and particularly on factors related to file-sharing. This is illustrated in equation (1) which represents the

million college students of whom approximately three million live in dormitories. It doesn't seem possible that the addition of these two groups could provide as much of an increase as indicated in the comScore numbers. The 2000 figures are found in "Napster Service Continues Rapid Growth, Reports Jupiter Media Metrix," Feb. 12, 2001, on the web at: http://www.comscore.com/press/displaycontent.asp?press=235&suffix=htm The 2001 figures are found http://www.comscore.com/press/release.asp?press=40

relationship between *per capita* record sales in a city, R, and traditional demand shifters, such as income, tastes, product quality (MusicQ), demographics, price of substitutes (and complements), the quantity and quality of file-sharing represented by FSQN and FSQL respectively.

$$(1) \quad R = F\left(FSQN, FSQL, Income, Demographics, P_{substitutes} P_{complements}, MusicQ, Tastes\right)$$

Ideally, we would want to have measurements of both the quantity of file-sharing and also its quality, by which I mean the ability of file-sharing to produce close substitutes for records. We have already noted that a downloaded MP3 file in 1999 provided a far poorer substitute for a CD purchase than a file downloaded in 2003. As CD writing devices became more plentiful and audio devices were developed which had the capacity to play MP3 files, downloaded files became much better substitutes for a CD purchase. It would be useful to have data on the penetration rate of CD burners or MP3 playing devices in individual cities. Unfortunately, we do not. Thus, we are limited to the quantity of file-sharing, which is proxied by Internet usage.

Measuring the quantity of file-sharing is not straightforward. The amount of file-sharing is related to both the number of Internet users and the share of Internet users engaged in file-sharing.

$$(2) \quad FS = Internet\,Users \bullet Filesharing \quad propensity$$

The number of internet users can only be taken as a proxy for file-sharing if file-sharing propensity is assumed constant or randomly distributed across cities. File-sharing propensity is related in part to behavioral differences across population but also to the availability of file-sharing programs and the quality of the shared files. The availability of file-sharing is the same for all cities although the quality of MP3 files may differ if the ownership of CD writers and MP3 players differs. Since we have no information on these latter factors we assume that the quality of MP3 files is the same for all cities. In the analysis below, we assume that any behavioral differences across cities is random.

Since file-sharing propensity increased after 1999, it would be incorrect to use the change in Internet users between any post-1999 years as a proxy for file-sharing. By way of numerical example,

assume that in year1, prior to the emergence of file-sharing, Internet penetration was at 50% in city A and 10% in city B, and each city had a ten percentage point increase in Internet penetration through year2 so that the respective penetration rates became 60% and 20%. Assume also that during year2, file-sharing had been fully embraced by those Internet users interested in such activity. The impact of file-sharing would be expected to be larger in city A than in city B since city A has a much larger base of Internet users. In fact, everything else the same, we would expect the impact of file-sharing to be approximately three times as large in city A since its Internet penetration is three times as large in year2. Yet, if we ran a traditional first differences or fixed effects model using the change in internet usage as an independent variable explaining the change in record sales, the coefficient on the file-sharing proxy would not properly reflect the impact of file-sharing since each city would have a proxy value for file-sharing of ten percentage points when in fact the file-sharing in city A is likely to be much larger than the file-sharing in city B.

Fortunately, we do not need to make such comparisons. Prior to 1999 file-sharing was zero, not because there were no Internet users, but because the file-sharing propensity was zero. The first year of data, 1998, occurs when file-sharing is zero. As file-sharing became a very popular activity in later years, the level of file-sharing in a city should mirror the *level* of Internet penetration as long as the file-sharing propensity does not differ too greatly per city. We include some demographic variables in an attempt to control for any such differences in propensity.

For this reason, we will use the level of Internet penetration in 2003 as the proxy for file-sharing, and not the change the change in Internet penetration (which is equivalent to setting Internet penetration at zero in 1998).

One concern might be that the number of Internet users depends on the behavioral aspect of file-sharing propensity. In other words, households get Internet service in order to engage in file-sharing. If this were so, cities with higher file-sharing propensities would also have greater Internet penetration rates and the relationship between FS and Internet uses would be non linear with those cities with large shares

of Internet use having populations engaging in more than their share of file-sharing. File-sharing would increase at an increasing rate as Internet use increased. There are reasons to believe that this is unlikely, however. First, as shown in Table 1 below, the August 2000 Internet penetration rate was 78.1% of 2003 penetration rate used in the regressions. Thus the great majority of Internet signups occurred before file-sharing achieved much popularity (the August 2000 values in Figure 1 indicate that file-sharing was still quite small). Second, one might expect that youthful individuals would be most likely to sign up for Internet access in order to participate in file-sharing. Yet, in a regression explaining the change in Internet penetration from 1998 to 2003 where the share of those between ages 10 and 29 was included as an independent variable, the coefficient was very small (.004) and very insignificant (although income was positive and statistically significant). Thus I conclude that this potential problem is not serious.

Because we have panel data that mixes both time series and cross section observations, either first differences or a fixed effects model can be used to advantage. Since just two years of data are used, 1998 and 2003, first differencing and fixed effects provide identical results and first differencing will be the technique chosen. The resulting estimate of the impact of file-sharing using first differencing should control for city factors that are time invariant.

The data provide income levels for each of the cities in our universe along with population, Internet usage, and income level broken down by age groups and Internet usage. There are data on 99 actual metropolitan areas and one hodgepodge called "DMA0" which is a (very large) summation of otherwise left out smaller localities. DMA0 had a very large population (approximately 75 million) and also the very lowest value of record sales per capita. The low value of per capita record sales for the catchall city might be due to rural users having a smaller selection of retailers and therefore joining record clubs or using mail order to a greater extent than users in other areas. Since record club and mail order sales are

not reflected in the SoundScan data, and record club sales represented almost a quarter of sales in the mid 1990s, DMA0 was removed from the analyses.[24]

Table 1: Summary Statistics						
	Variable	Obs	Mean	Min	Max	Std. Dev
Internet Users	all	380	0.50327	0.140142	0.13727	0.756569
Share of Population	1998	87	0.305275	0.065925	0.13727	0.490884
	2000	96	0.482127	0.078016	0.253412	0.66077
	2001	96	0.587033	0.07291	0.370887	0.724663
	2003	99	0.617621	0.075392	0.387262	0.756569
Records Sold	all	380	3.893513	1.773622	1.39735	13.82495
Per Capita	1998	87	3.883457	1.781789	1.764094	13.82495
	2000	96	4.200311	1.821451	1.870703	12.13022
	2001	96	4.068389	1.830705	1.609474	11.94825
	2003	99	3.430553	1.602438	1.39735	11.85128
Average Household	all	378	44323.72	7487.719	17455.99	68627.48
Income ($1998)	1998	87	44273.58	7566.949	23310.65	64652.3
	2000	96	44988.73	7170.08	27680.69	64577.1
	2001	96	45192.89	6952.666	26426.81	64118.93
	2003	99	42880.1	8090.607	17455.99	68627.48
share of ind 10-29	all	380	0.279279	0.034526	0.146226	0.363574
	1998	87	0.27748	0.036726	0.161318	0.363574
	2000	96	0.277918	0.032111	0.170434	0.356085
	2001	96	0.28055	0.034687	0.146226	0.356409
	2003	99	0.280272	0.035074	0.192253	0.356331
Income of	all	378	2.168755	0.526834	1.097394	4.873449
Internet Users	1998	87	1.968387	0.427799	1.097394	4.306938
Relative to	2000	96	2.026705	0.506117	1.142653	4.323478
Others	2001	96	2.18927	0.408503	1.522925	3.691067
	2003	99	2.462687	0.594942	1.395208	4.873449

Table 1 presents summary statistics for the remaining 99 cities. There are less than 99 observations listed per year because empty cells for any of eight age-group/Internet-access combinations were taken to indicate imprecision in the data and led to the elimination of the city (see footnote 21). The variables are generally self explanatory. Only the last two variables need any explanation. Since the data provide information on the ages of individuals in a city, it is possible to construct variables measuring the share of the population belonging to any particular age profile. In this case I focus on individuals between the ages of 10 and 29 inclusive, since they are thought to be the most heavily engaged in filesharing. Also, it is

[24] All the results were also run with DMA0 included. Its inclusion tended to increase somewhat the measured impact of file-sharing in the regression weighted by population and to have little impact otherwise.

possible to separately construct average income estimates for Internet users versus non users. If the sale of record albums is related to income and if file-sharing is differentially engaged in by different incomes groups, then the relative income of Internet users to non users will be related to record sales even holding file-sharing constant. I constructed a ratio of incomes of individuals with and without Internet access to address this possibility.

The mean values of these variables, although unweighted, largely reflect national averages, with the exception of record sales per capita and income.[25]

The share of Internet users basically doubled from 1998 through 2003, reaching a level of just over 61% in 2003. Record sales rose from 1998 until 2000 and then fell. The decline has a maximum value of approximately .77 albums per capita from 2000 until 2003, but is only .45 albums from 1998 until 2003. As expected, demographic variables such as the share of young individuals in a city, or real income, changed very little over this short interval.[26]

Internet users had incomes more than double that of non-users and there was a relatively large increase in the relative income of Internet users to non users over this period. Of course, this change is almost certainly caused more by the shifting of relatively high income individuals from the non-Internet to the Internet group than by the fortunes of the populations within each group.

[25] All of these averages give equal weight to each city and therefore are somewhat different than would be found in national (weighted) averages. The differences were very small except for the case of records sold per capita which are approximately .7 units lower per year in the weighted values, and income, where the weighted values are approximately $2500 higher.

[26] The decline in family income, which is impacted by changes in family size, is overstated relative to the weighted average, although even the weighted average indicates a somewhat larger decline than is normally reported in terms of *median* family income. Household Income is very broadly defined in these surveys (it includes income from almost any source) and for this reason may change in a manner that does not reflect changes in GDP very well.

Table 2: Dependent Variable is Albums Sold per Capita					
	Pooled	1998	2000	2001	2003
% of Internet Users	1.118	7.259	9.290	5.665	8.210
t-stat (robust)	(1.60)	(1.35)	(3.18)***	(1.80)*	(2.50)**
Household Income	1.66E-05	1.40E-05	-3.77E-05	-1.83E-05	-4.56E-05
t-stat (robust)	(1.33)	(0.41)	(1.04)	(0.78)	(2.36)**
Internet Income Advantage	-0.297	-0.009	0.085	-0.802	0.304
t-stat (robust)	(1.54)	(0.02)	(0.24)	(1.64)	(1.04)
Population	-1.85E-07	-1.88E-07	-1.66E-07	-1.72E-07	-1.52E-07
t-stat (robust)	(5.39)***	(2.39)**	(2.48)**	(2.61)**	(2.74)***
Share of 10-29 yr olds	-8.390	-7.280	-9.215	-10.555	-4.610
t-stat (robust)	(2.73)***	(1.40)	(1.15)	(1.77)*	(1.12)
Constant	5.966	3.490	4.145	6.642	1.181
t-stat (robust)	(5.48)***	(2.14)**	(1.67)*	(2.71)***	(0.52)
# of Obs	378	87	96	96	99
Adj R-Sq	0.13	0.2	0.24	0.2	0.19

Table 2 presents results from a pooled regression as well the results from running regressions for each individual year. Consistent results from these regressions, if we were to believe the results, are that a larger number of records sold per capita is related to higher shares of Internet users in a city, larger cities have lower per capita sales of records, and cities where 10 to 29 years olds make up a larger share of the population have lower record sales per capita. Higher household income is more likely to lower per capita sales of albums and the income differentials of Internet and non-Internet users do not have any consistent results. None of these results are particularly intuitive and it is probably more accurate to say that they seem counterintuitive. Although Internet users might be younger and wealthier than average and thus have a proclivity to listen to (and purchase) music, one might expect, at least during the later years when file-sharing was prevalent, that there would no longer be a positive relationship between record sales and Internet use. One might also expect that prior to file-sharing, having a large number of 10-29 year olds in a city would be positively related to record sales since young people are generally more interested in music, but that is not supported in these results. One possible explanation for the negative relationship between city size and record sales is the impact of the greater choice of radio stations found in large cities which would be the case if radio were a substitute for record sales, a conclusion found in Liebowitz (2004a). The negative coefficient on income found in 2003 would seem to defy economic logic.

Of course, models such as these are likely to provide unreliable results. There are many differences across cities that are not taken into account in these regressions, possibly leading to an omitted variables bias. Since other approaches allow the analyst to control for time-invariant differences across cities, we now turn in that direction.

Table 3 presents results from using a first differenced dependent variable. The dependent variable is the change in albums sold per capita. The table includes regression results over the full 1998-2003 interval. The models in Table 3 differ from one another by the inclusion of a different sample of cities (either the entire sample, or the 75 largest, or 50 largest), by whether the results are weighted by city population, and by the inclusion of different independent variables.[27] The first six columns use first differenced independent variables (except for the measure of Internet usage) whereas the last six columns allow several variables to take the form of their 2003 level as opposed to first differences.

Table 3: Dependent Variable--Change in Albums per Capita, 1998-2003

	entire	weighte	<75	<75 W	<50	<50 W	entire	weighte	<75	<75 W	<50	<50 W
2003 rate of	(3.39)	-2.220	-3.297	-2.309	-1.257	-1.241	-3.501	-2.400	-3.692	-2.577	-1.871	-1.927
Internet Use	2.96)***	(2.29)**	3.19)***	(2.17)**	(1.10)	(1.06)	3.32)***	(2.56)**	3.71)***	(2.54)**	(1.73)*	(1.53)
Income Change	8E-06	2E-05	2E-06	2E-05	8E-06	2E-05	5E-06	2E-05	2E-06	2E-05	2E-05	3E-05
	(0.71)	(1.64)	(0.16)	(1.44)	(0.63)	(1.69)*	(0.46)	(1.58)	(0.15)	(1.47)	(1.48)	(2.06)**
Change in Internet	-0.053	-0.041	-0.088	-0.030	0.121	0.034						
Income Advantage	(0.47)	(0.44)	(0.79)	(0.26)	(0.84)	(0.25)						
Population	-3E-07	-2E-07	-3E-07	-2E-07	-3E-07	-1E-07						
Change	(1.64)	(1.79)*	(1.98)*	(1.86)*	(2.10)**	(1.45)						
Change in Share	1.877	0.347	0.148	-0.262	-2.882	-3.075						
of 10-29 yr olds	(0.88)	(0.18)	(0.06)	(0.11)	(1.17)	(1.12)						
2003 share							0.252	-2.278	-3.365	-3.414	-7.666	-5.814
of 10-29 yr olds							(0.12)	(1.37)	(1.82)*	(1.93)*	2.86)***	(2.28)**
2003 population							1E-08	-2E-09	-1E-09	-3E-09	5E-09	-2E-09
							(0.76)	(0.25)	(0.08)	(0.38)	(0.37)	(0.16)
Constant	1.608	0.962	1.639	1.026	0.246	0.297	1.520	1.635	2.738	2.085	2.762	2.330
	(2.28)**	(1.59)	(2.57)**	(1.55)	(0.35)	(0.41)	(1.74)*	(2.18)**	3.46)***	(2.56)**	(2.39)**	(1.97)*
Observations	87	87	69	69	48	48	87	87	69	69	39	39
R-squared	0.18	0.16	0.22	0.18	0.18	0.2	0.15	0.14	0.22	0.18	0.27	0.25

t statistics based on robust standard errors in parentheses
* significant at 10%; ** significant at 5%; *** significant at 1%

[27] To pick the largest cities the cities were ranked by their Nielsen DMA code which closely corresponds to city size. Do not forget that twelve cities had questionable data and although most of these were small cities, one of these cities was the 47th largest.

The coefficient on our variable of interest, the share of Internet users in 2003, is negative in all instances, and usually with statistical significance (except for some regressions excluding the fifty smallest cities).[28] This is a striking difference from the pooled and single year regressions and indicates the dividends from using procedures that control for time invariant city characteristics.

The regressions weighted by population provide the results most appropriate for discussing the overall impact on the entire country since larger cities have a larger impact on overall record sales. They have the additional virtue of giving more weight to the larger cities which, due to the nature of the sampling, are likely to experience less severe measurement errors. The size of the coefficients imply, in all instances, a decline in record sales due to filesharing (given actual Internet usage rates) somewhat greater than the actual decline that has occurred. A coefficient of magnitude -2.1 (the average of the six weighted regressions), for example, implies that the 2003 Internet penetration rate of 60% would lower record sales by about 1.27 units per capita. This is slightly more than twice the actual decline (0.57 units) from the peak in 2000. It implies that record sales would have been 1.27 units higher per capita in 2003 than they actually were, or put another way, that were it not for file-sharing the sales of albums would have been approximately 3.55 per capita in 2003 instead of the actual (weighted) value of 2.28. That would have been an increase of .86 albums from the 1998 level of 2.69 units per capita, or an increase of .17 albums per year (implying a growth rate of 5.7% per year) which is somewhat greater than the average growth rate during 1973-1998 of 4.46% (in the RIAA data). However, it is common to find this level of implied five-year cumulative growth percentage for other intervals of five or even fewer years, putting this result clearly within the realm of possibility, based on historical evidence.[29]

[28] The coefficient begins to rise again, however, as the samples continues to get smaller (30 or less). If the data are divided into three groups based on city size, the coefficient on Internet use is greater for the middle group than for the largest or smallest cities. The coefficients for the middle group range from 4.2 to 5.7, which are somewhat larger than those reported in Table 3, and the coefficients for the largest and smallest groups are within the lower ranges found in Table 3 although somewhat less precisely measured (except for the weighted regression based on the 22 smallest cities which is essentially zero).

[29] In the time span 1973-1999 there were 6 five-years periods, 2 four-year periods and 2 two-year periods that had this level of cumulative growth (and 1 two-year period that came within a whisker).

Diagnostics were run to examine possible impacts of influential observations. First, regressions were run using the "robust regression" routine (rreg) in Stata, which iteratively weights cases to decrease the impact of cases with large residuals after first lowering the weight for cases with large leverage values (Cook's D). The resulting regression coefficients on the Internet variable were very similar to those reported in Table 3.[30] As an additional precaution, DFbetas for the Internet coefficient were also examined. Although there were a few observations that might be considered worthy of examination, given that the value of the DfBeta was greater than the usual cutoff of 2/sqrt(n), there were no extreme values.[31]

The other results from the pooled and individual year regressions, many of which seemed quite problematical, are, fortunately, gone in Table 3. The change in income no longer has a significant negative impact found in some of the individual year regressions, although the positive coefficient in the first differenced regressions is not statistically significant. Neither the change in population, the change in the share of young people, nor the change in the income of Internet users versus non-Internet users has consistent or statistically significant impacts.

Because the levels of population and the share of young people were often significant in the individual year and pooled regressions, I reran the first differenced dependent variable with the 2003 value of these variables (note levels, not differences) as independent variables, the results of which are found in the six rightmost columns of Table 3. The share of individuals aged 10-29 is generally negative and sometimes significant, as it was in the pooled and individual year regressions. In this case such a coefficient implies that cities with a larger share of teenagers and young adults experienced larger declines in record sales, a result that is consistent with the idea that file-sharing, which by all accounts is disproportionately engaged in by young people, is harmful to record sales. The coefficient on population level is essentially zero, indicating that although population might have been related to the level of record

[30] The coefficient on the Internet use variable from the robust regression (with t-statistic in parenthesis) was -3.05 (3.20) for the full sample, -2.62 (2.62) for the 75 largest cities, and -.93 (.76) for the 50 largest cities.

[31] The results for the full sample and for the 75 largest cities, based on removing observations having mildly suspicious DfBetas, were essentially unchanged. For just the largest 50 cities, however, removal of such observations (5) make the coefficients and t-statistics essentially zero. It is important to remember, however, that removal of influential variables may eliminate useful information.

sales in a city, it is not related to the changes in record sales. The percentage income advantage of Internet users is not significant in any of these specifications.

There is a well-known danger in extrapolating these point estimates as I have done, and there is always the chance that the coefficients themselves are somewhat inaccurate. Nevertheless, these results are quite unambiguous in their implication—regressions based on variations across cities are consistent with the hypothesis that file-sharing harms record sales. The size of this impact is generally greater than the decline in CD sales that has occurred. Although the usual caveats from basing conclusions on any single set of regression results clearly apply, the results from these regressions do not appear to contain the seeds of a refutation of that hypothesis.

IV. Examining Genres

In principle, the use of data containing information on sales by musical genres would seem capable of providing important additional clues about the impact of file-sharing. This is in large part because of the very different populations that are attracted to different genres of music. Classical music and jazz, for example, would be expected to appeal to individuals less likely to engage in file-sharing, whereas hard rock and rap would be expected to appeal to younger individuals more prone to engage in file-sharing.[32] Country might be expected to be somewhere in the middle, although this is a rather seat-of-the-pants estimate.

The use of these data is somewhat problematic, however, because albums are often classified in more than one genre. Given that albums can belong to multiple genres, and given that the process of album dual-classification appears somewhat arbitrary, as is the choice of genre itself, yearly fluctuation may not fully represent changes in market conditions so much as changes in judgment calls about

[32] Oberholzer and Strumpf (2004) report data that, if reflective of the entire market, would allow somewhat greater precision in determining which categories are most heavily downloaded. Using numbers from their Tables 4 and Table 8 it is possible to construct a ratio of downloads to sales for the albums in their sample, by genre. The results indicate that in a relative sense 'hard' rock is downloaded the most and jazz the least. I am surprised that "hard" is so much higher than other categories and that "rap" is as low as they find, making me suspicious of the generality of their downloading statistics. Nevertheless, here are the ratios: Hard 2.30; Alternative 0.82; R&B 0.52; Rap 0.38; Country 0.30; Jazz 0.17. [Latin .29]

classifications. If classification schemes change, the numbers may not be comparable from year to year. Two examples of this are found in Table 4, which lists seven genres reported by SoundScan. The extreme changes in the Hard category beginning in 1999 and the R&B category in 1997 were due to alterations in genre definition.[33] This reduces the confidence that we can have in results based upon analyses of these data. Nevertheless, with this caveat in place, we can continue our examination of the genre data.

Table 4: Album Sales (000s)								
	Total	Alternative	Classical	Country	Hard	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
% change 00-03	-16.41%	-2.13%	8.07%	5.71%	-17.01%	21.45%	-23.93%	-28.11%

Table 4 allows a cursory examination of total sales changes by genre. Three genres increased in absolute terms from 2000 until 2003—Classical, Country, and Jazz.[34] These results are generally consistent with a view that file-sharing has a negative impact on sales since classical and jazz would seem to be the least genres least susceptible to file-sharing and each increased during a period of decline, whereas those genres most likely to be impacted by file-sharing decreased. The actual increase in country music is a bit surprising, and as we will see below, inconsistent with the results from the genre regressions.

[33] Recent correspondence with SoundScan officials has confirmed what was originally just a hunch. Rap albums were eligible to be flagged as R&B beginning in 1997 and Alternative albums were eligible to be flagged as Hard beginning in 1999. The SoundScan representative claimed that there were no other changes during this period.

[34] Oberholzer and Strumpf state on page 12 of a brief before the Supreme Court that "musical genres which are not heavily downloaded on file sharing networks experienced the same reduction in sales as other genres." More specifically, they make this claim for two categories of music (Catalog and Country). Catalog represents sales of albums more than 18 months old and is not really a musical genre. The claim that Country has fallen as much as overall sales is clearly not correct. See Brief of Felix Oberholzer-Gee and Koleman Strumpf as Amici Curiae in support of Respondents, MGM v. Grokster, No. 04-480.

Table 5: Ratio of sales to radio listenership					
	classical	country	jazz/smooth jazz	urban/rap	metal/alternative
1998	7.73	6.04	4.68	7.91	14.97
1999	8.07	5.82	5.46	8.84	31.20
2000	6.98	6.10	5.27	10.80	35.39
2001	8.41	6.36	5.54	8.54	36.91
2002	8.59	7.41	5.50	7.67	32.09
2003	10.55	6.80	6.11	6.31	34.70
% change 2000-2003	51.19%	11.47%	15.96%	-41.60%	-1.95%

Table 5 uses radio listenership as a control for record purchases. If a genre is becoming less popular, the decline in popularity should show up in both radio listenership and record sales. If, on the other hand, a genre remains popular but file-sharing decreases sales of albums, the genre's share of radio listenership should hold up unless file-sharing decreases the time spent listening to radio, which seems somewhat unlikely.[35] Because Arbitron radio station genres match up only imperfectly with SoundScan album genres, Table 5 has fewer genres than Table 4. The results are largely similar to those drawn from looking at Table 4, but problems with genre definitions between the two mediums weaken somewhat any extra confidence this table might provide.[36]

We now repeat our econometric analysis for each genre of music. Regressions measuring the impact of file-sharing were rerun using genre-based sales per capita as the dependent variable. Regressions were run using three procedures: unweighted OLS, weighting observations by population, and the robust regression procedure in Stata that was discussed above.

For the sake of brevity, only the results from the weighted regressions are shown and those are found in Table 6. The results for the individual genres generally follow the same pattern of coefficients that was found for the entire population of records. Internet use has a negative impact on sales although of

[35] There are several reasons to believe this is the case. First, one third of radio listening takes place in automobiles and this is unlikely to have been affected by file-sharing. Second, file-sharing is a replacement for the purchase of music. By lowering the price of prerecorded music radio might be thought to become more valuable as a means of finding new music to download. This is tempered, somewhat, by the finding in Liebowitz (2004a) that radio is more a substitute for prerecorded music than it is a complement.

[36] The match-ups are far from perfect. The radio category that might be related to the record category 'jazz' is called New AC (Adult Contemporary) /Smooth Jazz which contains NAC and Jazz as components. The share of the Jazz component declines over time to zero in the charts. Does this category mesh with the record category? Perhaps not. Further, the radio category of R&B is a subcategory of 'remaining formats' and also has disappeared to zero, whereas the record category of R&B is very large, so these are not the same and not included in the table.

varying significance for different genres, change in income is positive but generally not statistically significant, population change has no clear effect even though the results are statistically significant for some genres, changes in the income differentials for Internet users has no clear impact, and changes in the share of young people have virtually no impact on record sales in any genre. Examination of the ordinary and weighted regression results indicated that some results were quite sensitive to outliers, far more so than the regressions based on the entire set of genres, making the results from the 'robust' technique of particular interest.[37]

Table 6: Weighted Regressions							
	Alternative	Rap	R&B	Metal	Classical	Country	Jazz
2003 intuse	-0.38661	-0.25707	-0.44368	-0.1804	-0.08277	-0.04837	-0.00895
	(2.78)***	(1.77)*	(1.63)	(1.27)	(1.02)	(0.41)	(0.22)
income change	4.31E-06	1.01E-06	2.66E-06	1.5E-06	1.57E-06	6.08E-07	4.05E-07
	(2.44)**	(0.60)	(0.82)	(0.93)	(2.09)**	(0.54)	(0.93)
chg income adv int users	0.013911	-0.02334	-0.03552	-0.0048	-0.0073	-0.00463	-0.00202
	(0.77)	(1.53)	(1.23)	(0.32)	(0.69)	(0.35)	(0.48)
pop change	-5.2E-08	-2E-09	-1.7E-08	5E-09	-4.7E-08	3E-09	-1.2E-08
	(3.40)***	(0.13)	(0.52)	(0.37)	(3.06)***	(0.31)	(3.40)***
chg in shr of yng in pop	0.336256	0.077334	0.124209	-0.1459	0.100789	-0.12335	0.016645
	(1.16)	(0.17)	(0.14)	(0.58)	(0.71)	(0.59)	(0.24)
Constant	0.271043	0.114693	0.184032	0.05808	0.069568	-0.00265	0.023138
	(3.11)***	(1.19)	(1.02)	(0.66)	(1.41)	(0.04)	(0.95)
Observations	87	87	87	87	87	87	87
R-squared	0.22	0.08	0.07	0.05	0.32	0.02	0.16
Robust t statistics in parentheses							
* significant at 10%; ** significant at 5%; *** significant at 1%							

Our main interest is in the relative size of the file-sharing coefficients for different genres. Table 7 provides those results. In Table 7 the regression coefficients for the Internet use were multiplied by the 2003 Internet penetration rate to generate a predicted impact of file-sharing on sales per capita for that genre. To put these numbers in perspective I then further scaled them by the per capita sales *for that genre* in 2000. First I will discuss the magnitudes of these values and then I will discuss the pattern.

[37] For example, in the unweighted regression the removal of a single city lowers the coefficient by more than 40%.

Table 7: Predicted Download Impact Relative to 2000 Sales			
Genre	robust	weighted	unweighted
Alternative	*-0.434*	***-0.485***	**-0.755**
Country	-0.435	-0.119	**-0.588**
Rap	-0.215	*-0.401*	*-0.500*
R&B	-0.217	-0.370	-0.445
Hard	-0.333	-0.359	**-0.925**
Classical	0.132	-0.830	-1.971
Jazz	0.007	-0.080	0.192
Statistical Significance of Coeficient used: Italics = significant at 10%; **Bold = significant at 5%**; ***Bold Italic*** = significant at 1%			

The numbers in the table give the predicted change in sales due to file-sharing as a percentage of 2000 sales. Thus, for example, the number from the robust regression for the musical category 'Alternative' would imply a decline in sales due to filesharing in 2003 that was 43.4% of the sales in 2000. As a matter of comparison, using this methodology for the entire market based on the average weighted regression coefficient in Table 3 provides a value of -44%. As pointed out in the discussion surrounding Table 3, this number is not particularly large relative to historical growth, and implies that without file-sharing record sales would have grown at a high but unexceptional rate of 5.7% from 1998 until 2003.

Nevertheless, some of the numbers in Table 7 seem very large and a few appear to be plainly implausible. Most of the extreme results are based upon the unweighted regression coefficients, represented in the rightmost column of the Table. The decline in classical recording is unreasonably large, yet it is so imprecisely measured that the very large coefficient is not statistically significant. The declines in Hard and Alternative implied by the unweighted regression also stretch somewhat our sense of reasonableness.

Examining the pattern of coefficients provides some background for putting these results in perspective. Two genres have very different results depending on the sample and the weighting—Classical and Country. The coefficients for the other genres have far more similar patterns between the different regressions. Classical music has the largest (in absolute terms) negative value in two columns,

but a very small *positive* value in the robust regression column. The coefficient for Country music is small in the weighted regression but considerably larger in the other two. The fact that the robust regression technique provides such different answers for these two genres is indicative of possible problems with outliers.

Upon inspection we discover that the unweighted results for classical music are heavily influenced by two (small) cities,[38] without which the results virtually to go zero. For this reason, I believe the results from the robust regression are more likely to provide useful information. The coefficient for the Country Music regressions, on the other hand, do not seem unduly influenced by a small number of cities, nor does restricting the sample to remove the 20%, 30%, or 40% of smallest cities change the results very much for the unweighted results. My conclusion is that the robust regression results provide the best information for both of these categories, leaving us to conclude that classical music was not impacted by file-sharing but that country music was.

How do these results comport with our hypotheses? Clearly, the fit is less than perfect but the general pattern conforms to the predictions. Jazz and Classical are at the bottom, as far as impact. Rap, R&B, Hard and Alternative are all fairly high. The one genre that doesn't fit expectations is country, which has a file-sharing impact as large as any other genre and are inconsistent with the general increase in country music in the file-sharing era. Although these results are far from overwhelming, on balance they too support the conclusion that file-sharing is harmful to record sales.

V. Conclusions

File-sharing is the newest and most publicized technology lending itself to unauthorized copying. Nevertheless, the results of this analysis indicate that the impact of file-sharing has been to bring significant harm to the recording industry. The decline due to file-sharing appears to be larger than the measured decline—the regression results indicate that file-sharing not only reduced sales but also vitiated an increase that otherwise would have occurred. File-sharing threatens or promises, depending on one's

[38] DMAs 92 and 74, which are Burlington, Vermont and Portland, Maine.

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

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. When this can happen without limit among strangers it would be amazing if there were not a powerful substitution effect.

As must always be the case, there are still many questions unanswered and unaddressed. And file-sharing is still a relatively recent phenomenon. For this reason I would caution the reader to draw too strong a conclusion from this analysis. It is always dangerous to make strong statements about markets that are in transition, because the future may always prove you wrong.

Blackburn, David "Does File Sharing Affect Record Sales?" Working paper, April, 2004, Harvard University (dissertation).

Boorstin, Eric "Music Sales in the Age of File Sharing" Senior Thesis, Princeton University, April 2004.

Conner K. R, and R. P. Rumelt Software Piracy - An Analysis Of Protection Strategies, Management Science 37 (2): 125-139 Feb 1991

Hirshleifer, Jack "Suppression of Inventions," *Journal of Political Economy*, March/April 1971, 79, 382—83.

Hong, Seung-Hyun "The Effect of Napster on Recorded Music Sales: Evidence from the Consumer Expenditure Survey" SIEPR Policy paper No. 03-018 January 2004

*Kai-Lung Hui and Ivan Png (2003) "Piracy and the Legitimate Demand for Recorded Music", Contributions to Economic Analysis & Policy: Vol. 2: No. 1, Article 11. http://www.bepress.com/bejeap/contributions/vol2/iss1/art11

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

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. "Copying and Indirect Appropriability: Photocopying of Journals," *Journal of Political Economy*, 93-5 October 1985, Pp. 945-957.

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.

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). http://www.oecd.org/dataoecd/55/57/32927686.pdf

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

*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

*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*, Summer 2005**

Vogel, Harold Entertainment Industry Economics Cambridge University Press, 6th edition.

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

Zentner, Alejandro, "File Sharing and International Sales of Copyrighted Music. An Empirical Analysis with a Panel of Countries", Unpublished manuscript. University of Chicago 2003.

Zentner, Alejandro "Measuring the Effect of Music Downloads on Music Purchases" *Journal of Law and Economics, forthcoming.*