Testing File-Sharing's Impact by Examining Record Sales in Cities[1]

Stan J. Liebowitz

School of Management

University of Texas at Dallas

First draft: September 2005

Second draft: April 2006

Abstract:

Although previous forms of copying have been found to often have benign effects on copyright owners the rise in file-sharing has coincided with a steep decline in the sale of sound recordings. This paper attempts to empirically examine the extent, if any, to which file-sharing has caused the decline in record sales. Using a data set for 99 American cities containing information on Internet use, record sales, television viewing, radio listening, and other demographic variables, an econometric analysis is undertaken to examine the relationship between record sales and file-sharing, as proxied by Internet use, from 1998 to 2003. First, we find that the Internet itself reduces time spent on other entertainment activities, but only by a small amount. We then subtract this generic Internet impact from the overall Internet impact and conclude that the decline associated with that component of Internet use that we can attribute to file-sharing indicates that file-sharing has caused the entire decline in record sales and appears to have vitiated what otherwise would have been growth in the industry. Looking at sales in individual musical genres reinforces the primary conclusion since those genres that seem most likely to fall prey to file-sharing have the strongest measured negative impact from Internet usage.

[1] Thank Marc Fuhrmann, Annie Hung for research assistance. Margolis and Zentner and others.

Is the Internet going to be the scourge or the savior of copyright based industries? Managers in the increasingly important copyright based industries are feverishly trying to determine the best business strategies to deal with dislocations being brought about by the Internet. This paper focuses on a potential dark side of the Internet otherwise known as peer-to-peer file-sharing.

Napster introduced to the world the idea of organized "file-sharing" and after its introduction in the fall of 1999 experienced rocket-like growth in usage and visibility. This growth made it an inevitable target for prosecution as a copyright violator and it was effectively shut down by a preliminary injunction within two years of its birth.[1] Nevertheless, its progeny live on and the repercussions on music listening and the music industry have not yet run their course.

File-sharing, simply put, allows one computer on the Internet to search for and access files on the hard drives of other computers that have joined the file-sharing network. The end result of file-sharing, in spite of its innocuous sounding name, is that individuals who do not own and have not purchased a particular song, program, or movie, can nevertheless obtain it from unknown third parties. Copyright owners fear that file-sharing will diminish the paid markets for their works since the great majority of files that are shared are copyrighted.

Its novelty not withstanding, file-sharing is merely the most recent example in a long line of technologies that have lowered the cost of unauthorized copying by individuals. Some earlier copying technologies were photocopying, audio taping, video taping, and of course, computer discs which allowed computer software to be copied.

Although each of the previous copying technologies engendered cries of alarm from the copyright industries affected—the print media, the movie industry, the sound recording industry, the software industry—there was no sustained decline in sales to support the rhetoric coming from these industries. In some cases, such as the VCRs and photocopiers, a boom in the movie and journal publishing industries

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

coincided with and appears to have been caused by increases in the penetration of VCRs and photocopiers.

The failure of the predicted damage to materialize, to say nothing of matching the rhetoric of industry spokesmen, has led some to conclude that managers of copyright industries do not understand the impacts of copying. However, Liebowitz (2006) suggests that there were various *sui generis* aspects to these previous technologies that undermined the conclusions drawn from simplistic analogies of prior copying technologies and that file-sharing is different in important ways from these previous technologies.

Unlike these previous instances, the growth in file-sharing has been accompanied by a large drop in sales of the material being copied—sound recordings. The industry's response has been to bring lawsuits against file-sharing services (such as Napster and Grokster) and also against thousands of individuals engaged in file-sharing.

The type of copying considered here is performed by individuals, as contrasted to organized forms of piracy where a criminal entity makes thousands or millions of counterfeit copies. Hui and Png (2003) examine the impact of organized piracy on the music business and, not surprisingly, find it to be negative.

The goal of this paper is to determine the impact of file-sharing on record sales in the United States, the world's largest market for sound recordings. The conclusion is that file-sharing has caused the recent decline in record sales and appears to have vitiated a sales increase that would otherwise have occurred. The managers in these industries do not appear to be crying wolf.

The procedure used in this paper has certain advantages over other recent studies. Liebowitz (2006) examines the historical trend of record sales and alternative explanations that might explain the decline in record sales in the United States but provides no direct estimates of the impact of file-sharing. Zentner (2006) examines the impact of file-sharing on the proclivity to purchase sound recordings in the EU although that estimate does not translate directly into an estimate of the quantity of records purchased.

Rob and Waldfogel (2006) examine the impact of filesharing based on a survey of self-reported purchases of records and file-sharing activity by American college students at four campuses. Due to the limitations of their sample, Rob and Waldfogel caution against generalizing their results to the entire marketplace. Peitz and Waelbroeck (2004) and Zentner (2005) use Internet use and record sales statistics to examine sales in a cross section of countries in an attempt to determine the direction of impact of file-sharing, but do not measure the overall predicted impact relative to actual changes in sales, nor do they take account of the impact of the Internet itself on sales. Two working papers with similar methodologies but contrary conclusions, Blackburn (2004) and Oberholzer and Strumpf (2004), examine sales and downloads of individual records in the US. Each needs to overcome a very serious simultaneity problem because the most popular songs are both heavily downloaded and heavily purchased. Each also needs to deal with potential fallacy of composition since what may be true for individual records may not be true of the entire market.[2] Two other working papers using a similar methodology to each other, Hong (2004) and Michel (2004), use national data on self-reported purchases of CDs and Internet use. Except for the paper by Oberholzer and Strumpf, all of these papers find some degree of harm brought about by file-sharing.

In contrast with several other papers that use surveys of self-reported purchases, the empirical work in this paper is based on actual sales of record albums. Further, this paper uses American cities whose inhabitants all function in a single national marketplace with similar musical trends, similar advertising, similar retail outlets, and similar pricing. This paper demonstrates the difficulties in using Internet usage as a proxy for file-sharing in a fixed effects type model concluding that misspecification can be avoided if the final level of Internet use as the independent variable and the beginning period precedes the birth of file-sharing. This study is also the first to separately measure and then remove the impact of the Internet on general entertainment consumption.[3] This study is also the only one to use information on genres to provide a further test the impact of file-sharing. Finally, the empirical results from this study are easily

[2] See Liebowitz (2005) for a more detailed explanation of potential problems throughout this literature.
[3] I would like to thank an anonymous referee for suggesting this.

translated into determining an overall national impact of file-sharing. That is not to say the approach below is not without weaknesses—they will be addressed as they appear.

I. What do we know about file-sharing?

File-sharing currently encompasses sound recordings, films and television programs, computer software, various forms of pornography, and other products that can be digitized. Most estimates indicate that audio files have been and continue to be the most popular type of file being shared, making them the best source for assessing the impact of file-sharing.[4] As file-sharing behavior responds to changes in bandwidth and other improvements to hardware it is natural to expect that file-sharing's impact on music sales may become applicable to movies and other copyrighted works.

By most estimates, the amount of file-sharing is enormous. Liebowitz (2006) documents many of these estimates of overall size and trend, and finds that the high estimates indicate that file-sharing by Americans is perhaps two or three times as large (in terms of number of song files) as the legitimate US sound recording market and the lower estimates indicate that file-sharing is perhaps in the vicinity of one third of the legitimate market. Although the variance in these estimates is unsettling, it nevertheless seems clear the file-sharing is very large.

The pattern of file-sharing's growth, to the extent that it is known, can be quickly summarized: there were 1.3 million US users in February of 2000 rising to 13.5 million at Napster's peak in February of 2001 according to ComScore MediaMetrix; users migrated to other programs after Napster's shutdown and growth seems to have continued, based upon surveys such as those from the Pew Internet & American Life Project; that the lawsuits against file-sharers by the Recording Industry Association of America (RIAA) announced and then begun in mid 2003 are most likely responsible for a reduction in file-sharing in 2003 according to measurement from comScore, Big Champagne, Pew and from academic

[4] IDATE claims the ratio of audio files to video (films) 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.

analyses such as Bhattacharjee et al. (2006); and that the upward trend in filesharing then appears to have begun again in 2004 and has continued into 2005.

When Napster first became popular, 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.

Over 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. CD burners became common, DVD/CD players added the ability to directly play MP3 files, and MP3 players such as iPods became much more widely used. Thus the amount of file-sharing, and the ability of shared files to substitute for purchases should have grown, independent of any increase in the number of individuals engaged in file-sharing.

II. The Theory of File-Sharing's Impact

The impact of unauthorized copying on sales of authorized copies is not necessarily the simple negative outcome it was once thought to be. The relative strength of potential competing forces needs to be taken into account. The possible impacts of file-sharing have been elucidated in detail elsewhere (Liebowitz 2005) and are listed here in a very condensed manner.

The unauthorized downloading of a copyrighted song can easily be seen as a substitute for the purchase of that copyrighted work. This substitution can only work to reduce the effective demand in the market and can only harm the financial position of the sound recording producers. It is hard to imagine that this substitution effect does not play an important role for some reasonable subset of the file-sharing population.

A different possible consequence of file-sharing, frequently referred to as the *sampling* effect, might be for downloaders to merely use the downloaded songs to help guide their later music purchases. The economic impact of sampling on sellers is often misunderstood, however. If consumers sample music to learn more about potential purchases they will make superior choices but they will not necessarily purchase more albums. To briefly illustrate this point, assume that music listeners have a constraint on the daily amount of time during which they can listen to music. Further, assume that the time consumers spend listening to a particular song or albums is a function of how much they enjoy it. Since sampling allows consumers to purchase music that is on average more enjoyable it will increase the time they spend listening to each purchased item. If so, consumers fill up their allotted music listening time with fewer songs or albums. This is not to say that sampling cannot increase sales, merely that it cannot be counted on to offset the substitution effect.

A third factor concerns network effects. Models such as Conner and Rumelt (1991) and Takeyama (1994) demonstrate that under certain conditions unauthorized users of an intellectual product might benefit sellers by creating positive network effects of value to the purchasers of legitimate copies. These models make some sense for computer programs but application of this idea to music seems a rather remote possibility.[5] File-sharing is unlikely to increase network effects since everyone already listens to music and everyone already has free access to an unlimited amount of music on radio. Nor does time spent listening to radio seem to have a positive impact on record sales, contrary to the network effect theory, as discussed later in the paper.

The final impact of copying that conceivably applies to file-sharing is called indirect appropriability. The basic idea (Liebowitz, 1985) is that copying increases the demand for originals because buyers of the originals appropriate some of the value of the copies through payments in money, in kind, or perhaps in good will from those granted permission and access to make copies. In order for

[5] For example, if employees become familiar with a spreadsheet by using a unauthorized copy at home, then their employers, who purchase legitimate copies, might place higher values upon purchasing spreadsheets since training costs are lower

indirect appropriability to work, however, some form of appropriation is required and none seems possible when users are anonymous, copies are unlimited in number, and direct payment is nonexistent.

Overall then, the implication of this section is that the effects that might offset substitution in other forms of copying are likely to be absent for file-sharing of music on the Internet.

III. Data and Econometric Issues

This econometric investigation requires the merging of several data sets. First, the US Census as part of its Current Population Survey (CPS) undertaken for the Bureau of Labor Statistics conducts surveys on Internet and Computer use. These surveys, conducted in December 1998, August 2000, September 2001, and October 2003, provide information on the penetration of home internet use, the type of internet connection, household family income, Metropolitan Statistical Area (MSA) of residence, the age, sex, race and education of respondent, as well as a host of other variables that were not used in the analysis. This information is based on responses from approximately 130,000 individuals. The size of the sample in small MSAs is sometimes insufficient to provide accurate estimates for various demographic data. Second, Nielsen SoundScan sells data on album sales, by genre and by year, for the largest 100 metropolitan areas, which it refers to as Designated Market Areas (DMAs) of which there are 210 covering the entire country. Third, Nielsen Media Research sells data on television viewing in DMAs (DMAs are based on television markets). Finally, Arbitron collects data on radio listenership in its own set of metro areas.

Combining these data sets is not a trivial task. Every county in the US is allocated to a Nielsen DMA. The 100 largest DMAs include approximately 83% of the total population. Unlike the Nielsen DMAs, the 241 named Census MSAs, when summed, do not cover every household in the US. Nielsen DMAs generally have larger populations than similarly named Census MSAs or Arbitron Metro areas.

With the help of DMA maps, MSAs and Arbitron metros can be allocated into DMAs. This was done to create the data set used in the analysis below. This basic methodology of combining Census data

with Nielsen SoundScan data was first proposed by Eric Boorstin in a 2004 senior thesis at Princeton University. [6]

Because DMAs are larger than MSAs, matching them together often required adding several MSAs together to approximate the DMA. Further, aggregating Census MSAs will occasionally match only a small portion of the DMA population, particularly for the DMAs with smaller populations. For that reason a variable called "Coverage" was created to measure the portion of the DMA population covered by the MSA aggregation. When Coverage falls to a low level it is possible that the census variables will not properly reflect the actual population characteristics in the DMA. In the analysis that follows the influence of observations where the Coverage is small will often be restricted in order to reduce potentially misleading measurements.

Although the data from Nielsen SoundScan include 100 DMAs, one DMA could not be matched with any census MSAs and was dropped from the analysis. Further, missing data for other variables removed some DMAs in some regressions.

The SoundScan data for record sales are primarily based on information from electronic scanners in retail outlets. The data include not only physical CDs sold locally but also CDs sold on the Internet (based on the zip code where the CD is delivered) and also digital downloads (based on the zip code of the credit card). As a factual matter, digital downloads played no role in the analysis since they were a trivial component of the market even as late as 2003 (they made up only 2.4% of the market in 2004, the first year they were officially counted by the RIAA).

Table 1 presents summary statistics for most variables used in the analysis for the ending year 2003 and the change from 1998 until 2003, allowing the reader to infer the 1998 statistics if desired. Some weighted averages are included for comparison purposes. The population of the MSAs matched to the DMAs cover about 86% of the population of the included DMAs, which themselves only cover 83% of

[6] I want to thank Mr. Boorstin for graciously making his data available to me although the current data set was created independently.

the population in all DMAs, so that in total our sample covers about 72% of the US population. Because of their more rural characteristics, these left out individuals are poorer, have fewer broadband connections, lower Internet penetration rates, and lower income as shown in the rightmost column of Table 1. Nevertheless, some of our included DMAs have lower broadband penetration, lower Internet penetration, and lower incomes than the average left-out population.

Table 1: Some Summary Statistics

2003	Obs	Mean	St. Dev.	Min	Max	weighted average	Left out Population
Album Sales per Capita	99	2.32	0.440	1.499	3.879	2.44	
Average Income	99	47,966	8,986	20,380	75,895	50,540	38,496
Broadband Share	99	0.25	0.065	0.120	0.420	0.259	0.15
Coverage Ratio	99	0.73	0.220	0.203	0.999	0.828	
Dialup Share	99	0.37	0.064	0.170	0.500	0.36	0.40
Music Radio Listening (hours per day)	96	2.34	0.193	1.866	2.815	2.29	
Population	99	2,350,517	2,727,490	630,774	19,400,000		79,147,155
Radio Listening (hours per day)	96	2.723	0.174	2.373	3.233	2.78	
Share of Internet Users	99	0.61	0.071	0.440	0.740	0.62	0.55
Share of Males	99	0.48	0.023	0.400	0.520	0.48	0.49
Share of Population 12-29	99	0.30	0.044	0.200	0.410	0.31	0.25
Share of Population Hispanic	99	0.09	0.111	0.000	0.530	0.13	0.11
Share of Population 55+	99	0.23	0.054	0.130	0.410	0.22	0.14
Share of Population College Degrees	99	0.20	0.051	0.087	0.345	0.22	0.14
Share of Population Black	95	0.13	0.104	0.007	0.502	0.14	0.09
TV Viewing (hours per day)	99	4.44	0.652	3.264	6.096	4.76	
98-03							
Album Sales per Capita Change	99	-0.58	0.695	-3.484	1.049		
Average Income Change	99	8,523	7,087	-6,660	26,901		
Coverage Ratio Averaged over Years	99	0.73	0.216	0.200	0.990		
Dialup Share Change	99	0.06	0.079	-0.200	0.250		
Music Radio Listening Change	95	-0.31	0.140	-0.651	0.043		
Population Change (%)	99	0.07	0.07	-0.06	0.35	mean	weighted
Radio Listening Change	96	-0.28	0.121	-0.600	0.026	absolute	absolute
Share of Internet Users Change	99	0.31	0.058	0.120	0.466	value	value
Share of Males Change	99	0.001	0.035	-0.137	0.143	0.024	0.017
Share of Population 12-29 Change	99	0.001	0.045	-0.110	0.140	0.036	0.025
TV Viewing Change	99	0.06	0.355	-0.552	1.224		
Share 55+ Change	99	0.011	0.047	-0.120	0.191	0.035	2.75%
Share College Change	99	0.018	0.040	-0.114	0.208	0.032	2.89%
Share Black Change	93	-0.004	0.043	-0.131	0.104	0.032	2.24%

Consistent with industry claims, album sales per capita fell. This decline does not reflect the full extent of the decline in the last few years since the peak of record sales occurred in 1999 or 2000, depending on which data source is used. Liebowitz (2006) reports on the pattern of album sales in the US

using RIAA data and concludes that per capita sales of albums have dropped over 30% from their peak in 1999. The SoundScan data used in this paper show the peak year to be 2000. The RIAA data indicate a larger decline than the SoundScan data in part because RIAA data include outlets particularly hard hit in recent years—record clubs and direct sales—whereas SoundScan data do not. Nevertheless, the two data sets generally move together.

The share of Internet users can be seen to basically double from 1998 through 2003, reaching a (unweighted) level of just over 61% in 2003. Note that the change in dialup was small for most cities and negative in some, as users switched from dialup to broadband. It seems likely that many broadband users were the original dialup users, not the new Internet customers. Note as well that television viewing increased while radio listenership declined.

A. Does Internet Use Impact Time Spent on Entertainment?

The key interest in this paper is the effect of filesharing on record sales. Since we do not have a direct measure of file-sharing, a measure of Internet penetration—the share of Internet users in a city—will be used instead. It is assumed that cities with higher Internet penetrations also have greater Internet usage (the terms will be used interchangeably) and more file-sharing.

This leads to a direct question: If Internet penetration is the chosen proxy for file-sharing, could the results that be contaminated by the Internet having other possible impacts on album sales independent of file-sharing? For example, perhaps using the Internet takes time away from many recreational activities including listening to sound recordings, so that a decline in album sales would be expected when Internet use increases even without any independent impact from file-sharing. The analysis in this section attempts to gauge whether there is such an impact, and if so, how large it might be.

The approach taken here is to examine the relationship between Internet penetration and the usage of television and radio, the two most time consuming entertainment activities consuming over seven

hours of the average person's day. The working assumption is that if the Internet is a substitute for entertainment it would reduce radio and television usage.

There is a small survey-based literature that examines the impact of Internet usage on television usage. The Digital Future Report (2004) finds that one third of Internet users say they watch less television because of the Internet (virtually none report watching more). An analysis of longitudinal survey results by Nie et al. (2005) concludes that Internet use reduces overall TV viewing and the authors ague that cross section survey analysis is insufficient to find the correct result, consistent with our results below.

These claims of reduced television viewing are somewhat difficult to reconcile with historical television ratings. According to Nielsen Media, hours of television viewing per capita increased fairly smoothly during the period from 1995 to 2005 and this was true for every age category (e.g., 9.5% for ages 2-11, 14.1% for 12-17, 8% for 18-34, 16.8% for 35-54 and 12.4% for 55+). Nevertheless, it is possible that the Internet could have had a negative impact on viewing that was overcome by other factors tending to increase television viewing (such as a greater number of available cable networks). With the number of US Internet users growing from about 20 million in 1995 to about 180 million in 2004, however, it seems fairly clear that the reported increase in television viewing could not have occurred if Internet usage had had a *large* negative impact on television viewing habits.

Television requires an attention level that precludes serious Internet activity whereas music listening can easily coexist in the background. It seems reasonable, therefore, to expect that the Internet's impact on television would be larger than its impact on radio (or prerecorded albums).

In Table 2 we run regressions where televisions usage and radio usage (measured in hours per day) are the dependent variables. In order to weaken the impact of observations where the Census population does not seem to match the DMA population and to weaken the impact of cities that are less well

measured in the Census, we weight each observation by a combination of the Coverage ratio and population.[7]

The first four regressions are for the individual years 1998 and 2003 for each medium. Television usage is generally thought to be related to various demographic variables. It is well known that time spent viewing television increases with age and that minorities watch considerably more television, particularly black viewers (Trac Media Services, 2001). We include, therefore, two measures of minority statues (black and Hispanic) and two measures of the age distribution (12-29 and 55+). We also include average income and a measure of educational attainment (college graduate).

The yearly regression results are consistent with prior evidence. The coefficients for age groups and minority status both strongly support expectations about television. Interestingly, these results hold for radio as well. It also appears to be the case that cities with a larger share of males have greater television usage (but not radio) and although there is some indication of a negative relationship between share of population with college degrees and television and radio use, it is not strong. Listenership and viewership are greater in larger cities, most likely because of the increased set of broadcast choices available. The effect is not strong as an increase of ten million individuals would lead to an increase in viewing of only 30 minutes. Income seems to have no clear effect. There is little noticeable relationship between Internet penetration and radio or television use although the relationship is more negative in 2003 than in 1998.

The last two columns provide the main test of the relationship between Internet penetration and media usage—running the regressions in first differences. The extra information added by this temporal element should provide for a more powerful test as long as there is sufficient variation in the independent variables. Since the Internet was growing rapidly and unevenly over this period we need not worry about its variation. The variables in these two regressions are all measured in first differences from 1998 until

[7] The specific weighting variable is the product of the squared Coverage term and the square root of population. This particular format was chosen to equalize in the impact (measured in terms of correlations) of both factors on the weighting variable. These results are representative of separate regressions run on subsets of the data.

2003. Because broadband penetration was effectively zero in 1998 the 2003 level is treated as a first difference. An examination failed to reveal any overly influential observations.

Table 2: Regressions weighted by Coverage Ratio and Population							
Dependent Variable	TV Usage (hours)		Radio Usage (hours)			TV	Radio
Independent Variable	1998	2003	1998	2003		98-03	98-03
Broadband		-1.313		-0.641		-2.025	-0.771
		(0.83)		(1.44)		(2.14)**	(3.01)***
Dialup	1.141	-1.161	-0.008	0.095		-1.021	-0.034
	(1.13)	(0.75)	(0.02)	(0.24)		(1.35)	(0.15)
Income (000)	-0.010	0.001	0.005	0.000		0.010	-0.002
	(0.91)	(0.15)	(1.28)	(0.06)		(1.76)*	(0.97)
Population (000,000)	0.005	0.004	0.002	0.002		-0.015092	0.010
	(4.68)***	(3.73)***	(4.08)***	(6.30)***		(2.61)**	(4.20)***
55 and over	2.121	3.723	0.553	0.840		0.973025	-0.119
	(1.15)	(1.47)	(1.09)	(1.20)	Change in Variables	(0.77)	(0.40)
Share with College Degree	-2.641	-0.260	-0.734	-0.364		1.8681196	-0.244
	(1.90)*	(0.13)	(1.56)	(0.71)		(1.26)	(0.58)
Share Males	3.387	5.708	-0.429	1.267		-1.477526	0.271
	(1.53)	(1.60)	(0.57)	(1.22)		(0.80)	(0.67)
Hispanic Share	1.038	0.953	0.507	0.468		-0.899611	0.384
	(2.15)**	(1.92)*	(2.28)**	(2.57)**		(0.63)	(1.31)
Black Share	3.252	3.585	0.465	0.590		0.875224	-0.217
	(6.25)***	(4.77)***	(2.45)**	(3.08)***		(0.78)	(0.70)
Share Young	-0.213	-1.259	-0.686	-1.129		-0.177066	-0.236
	(0.11)	(0.49)	(0.93)	(1.88)*		(0.11)	(0.59)
Constant	2.391	1.482	3.134	2.308		0.589	-0.089
	(1.61)	(0.57)	(6.15)***	(3.08)***		(2.04)**	(1.35)
Observations	95	95	92	92		93	90
R-squared	0.55	0.47	0.44	0.63		0.15	0.32
T statistics (robust) in parentheses		* significant at 10%; ** significant at 5%; *** significant at 1%					

For both television and radio there is a significant negative relationship between changes in broadband and changes in audience usage, with a weak negative impact of dialup on television viewing and no impact of dialup on radio. The larger impact of broadband may be due to broadband's speed advantage over dialup leading to greater intensity of Internet use, or it may be due to the more intense Internet users being the first to adopt broadband, or some of both. There is little consistency among the demographic variables, which may reflect the reasonable supposition that important demographic characteristics of cities do not change much over five year intervals.

The size of the coefficients on Internet penetration, irrespective of its statistical significance, implies a reduction in television viewing and radio listening due to the change in Internet usage during 1998-2003 that is fairly small. Table 3 provides the intermediate steps to calculate the impact of the Internet, for both dialup and broadband, on time spent with television or radio. Using the coefficients from the regressions, the net impact of the Internet is to lower television viewing by about 13% and radio by 7%, with broadband penetration being responsible for the great majority of the decline in television and the entire decline in radio. The complete impact of the Internet would be larger than these measurements since the Internet was already well established in 1998 and the current measurements only measure the marginal impact since then. Our mini-examination indicates fertile ground for a more detailed study.

Table 3: Impact of Internet on TV and Radio Audience

	TV		Radio	
	broadband	dialup	broadband	dialup
1. Coefficient	-2.02	-1.02	-0.77	-0.03
2. 1998-2003 Change	0.25	0.06	0.25	0.06
3. Expected Impact [1*2]	-0.50	-0.06	-0.19	0.00
4. Average TV or Radio Usage	4.47	4.47	2.86	2.86
5. % Decline caused by Internet [3/4]	-11.22%	-1.44%	-6.67%	-0.07%

To put this result in perspective, it means that increased Internet penetration will lead to a measured decline in record sales independent of any impact of file-sharing. If the impact on sound recordings is similar to that of radio (or television) then the changes in Internet penetration over this time period, independent of file-sharing, might cause a decline in record sales of 7% (13%). If we go out to the bounds of the 90% confidence interval for the Broadband coefficient, the maximum decline would be 10% for radio and 20% for television. This is for the period 1998-2003. Since file-sharing didn't technically begin until the second half of 1999 and didn't really garner large numbers until 2000, we will reduce these values when calculating its impact during the file-sharing era.

B. Specification of the File-sharing Variable

There are several issues involved with trying to use Internet penetration as a proxy for file-sharing. Internet penetration reflects the number of users, not their intensity or frequency of use. Inclusion of a variable measuring broadband should help ameliorate this problem, as should the inclusion of several variables related to file-sharing intensity. First, however, we focus on the proper specification of the file-sharing variable for our use in a first differenced equation.

The amount of file-sharing should be related to the product of the number of Internet users and the propensity of Internet users to engage in file-sharing, represented in equation (1) where FS stands for the quantity of file-sharing, IU stands for the number of Internet users and FP stands for average filesharing propensity.

$$(1) \qquad FS \equiv IU \bullet FP$$

The change in file-sharing from period t-1 to period t can be represented as:

$$(2) \qquad FS_t - FS_{t-1} = IU_t \bullet FP_t - IU_{t-1} \bullet FP_{t-1}$$

Equation 2 can be rewritten (subtracting and adding FP_t ? IU_{t-1} to the rhs) as:

$$(3) \qquad \Delta FS = \Delta IU \bullet FP_t + \Delta FP \bullet IU_{t-1}$$

Since the only variable that we have measurements for is IU (and the change in IU), we cannot, in general, measure the change in file-sharing without making some further assumptions about file-sharing propensity, about which we have only limited and often contradictory claims. File-sharing propensity is related to the number of people who engage in file-sharing and also the frequency and intensity with which they engage in file-sharing. With changes in technology and the filing of industry lawsuits, file-sharing propensity has certainly not remained constant. Therefore it would generally be inappropriate to use first differences or fixed effects where the (change in) Internet usage was taken to be a proxy for (the change in) file-sharing since this would ignore the entire second term on the rhs of (3).

By way of numerical example, assume that in period t-1, at the emergence of file-sharing, Internet penetration was at 50% in city A and 10% in city B. Now let each city have a ten percentage point increase in Internet penetration through period t, so that the respective penetration rates became 60% and 20%. Assume also, over this period, that Internet users become twice as likely to engage in downloading music files on file-sharing networks (FP doubles). The impact of file-sharing would be larger in city A than in city B since city A has a much larger initial base of Internet users who are impacted by the increase in FP. Running a traditional first differences or fixed effects model would provide misleading results since both of these cities would have a proxy value for file-sharing of ten percentage points when in fact the file-sharing increase in city A is much larger than the file-sharing increase in city B.

The strategy adopted in this paper is to go back to a period where file-sharing propensity was zero, which is anytime prior to Napster and thus doesn't require the first period to be earlier than late 1999. If FP_{t-1} is equal to zero then (2) transforms into equation (4) below, which indicates that the change in file-sharing is equal to the product of second period Internet use and second period file-sharing propensity.

$$(4) \qquad \Delta FS = IU_t \bullet FP_t$$

With this strategy the number of internet users can be taken as a proxy for file-sharing if file-sharing propensity is assumed constant across cities at a moment in time, or if we can control for the factors that might lead to different file-sharing propensities, or if the factors that lead to different propensities across cities are random and merely create noise.

File-sharing propensity is related to behavioral differences across populations, the availability of file-sharing programs, prices of originals versus the ease of downloading and the quality of the shared files. The availability of file-sharing programs would seem to be identical for the Internet users in all cities. The quality of the downloaded file depends on its ability to substitute for a CD as well as its pure sonic quality. Since we have no information on factors impacting the quality, such as the ownership of CD writers and MP3 players (although income might be related to the ownership of these devices) we can

do no better than to assume that the ability to substitute MP3 files for purchased CDs is the same for Internet users in different cities.

There are some measurable demographic dimensions to file-sharing, however. The Pew Internet Life Project has found that young people are far more likely to engage in file-sharing than are older individuals and that men are somewhat more likely to engage in file-sharing than are women. The differential rates are reported in Table 4.

Table 4: Group File-sharing Use Source: Pew Internet and American Life

	Jul-00	Dec 00	Feb 01	Sept 01	Dec 01	Oct 02	May 03	June 03	Nov 03	Avg
Men to Women	130%	141%	172%	146%	154%	134%	134%	152%	169%	147.9%
Young (18-29) to Older (30+)	349%	376%	378%	359%	378%	296%	320%	344%	371%	352.5%

In the analysis below, we try to directly control for both of these demographic differences across cities—ratios of males and share of young. Our variable measuring the share of young individuals in an area includes those aged 12-17 in addition to those 18-29 because this younger group is likely engage in such behavior and because the Pew survey was limited to those over 18. Whether there will be sufficient change in these demographic variables over time to play a role in the regressions is unclear.[8] The type of Internet connection is also a control for file-sharing propensity. The greater download speed of broadband might enhance the lure of file-sharing although the ability of file-sharing programs to download songs unattended by the operator limits the cost and inconvenience of having a slow connection.

The analysis below takes as given that the supply of records and the price of records is the same in all cities, which appear to be reasonable assumptions. First, the list price is the same everywhere. Second, most big box retailers are national chains as are some record stores. Over the 1998-2003 period big box retailers increased their market share of CD sales to 56% while the share of record stores declined to 33% (according to the RIAA website). Discussions with the leading distributor of sound recordings indicated that individual chains selling CDs have the same transaction price on a particular CD in all their stores

[8] The surprisingly large change in the share of males and young people (found in Table 1) for certain DMAs is possibly due to imprecision in the measurements in small cities as previously alluded to. For example, the average absolute percentage change in these two variables from 1998 to 2003 was three times as large for the 25 smallest DMAs relative to the 25 largest DMAs.

throughout the country (except in the very smallest localities). Internet sales (5% of the market) are also national. Consumers in any city, therefore, had access to identically priced CDs in national chains or over the Internet. Under these circumstances it seems unlikely that prices would differ systematically in our different DMAs. Even if the mix of retailing were different in some set of cities, unless these differences changed over our time frame, our use of first differences should control for these retailing/price anomalies.

Finally, a choice of timeframe that will be consistent with the analysis in this section must be made. There are four data sets in the special CPS Internet survey. The most recent census, from October of 2003 will provide the final year in the analysis. Finding the starting point is just slightly less simple. The first special survey, in December 1998, occurs when file-sharing is zero, thus making it a perfect candidate for the starting date. The second survey, in August of 2000, occurs approximately one year after Napster began and at a time when the number of Napster's users was slightly greater than half of Napster's eventual peak, according to comScore MediaMetrix. Thus it would appear to be incorrect to assume that file-sharing propensity was effectively zero in 2000. The use of the broadband variable would also become much more questionable with 2000 as a starting point, since about 12% of all Internet users were using broadband at the time of the 2000 survey. Therefore, 1998 is chosen.

Because we have panel data that combines both time series and cross section observations, either first differences or a fixed effects model can be used and they will provide identical results.

IV. Estimation

Table 5 presents results from individual year regressions on the per capita sales of sound recordings. These regressions were performed analogously to those in Table 2, meaning that they are weighted by a combination of population size and Coverage. They are run first with a partial set of demographic variables and then with a more complete set. Although we will depend on first differences for our ultimate regressions there are some clear patterns in these pure cross section results. Cities with higher Internet access rates also have higher record sales, although this result is slightly more pronounced

in 1998 than it is in 2003. The positive coefficient might reflect the possibility that cities with a more media savvy population have both higher Interest in sound recordings and the Internet. This would be consistent with a common finding in surveys that file-sharers are often those who also purchased records at an above average level. We also find that cities with higher income have higher record sales, although this result disappears when more demographic variables (especially college degree) are included. Cities with large populations also have higher sales per capita in 1998, although this is not found in 2003.

Table 5: Individual Year Regressions on Record Sales per Capita				
	1998		2003	
Dialup	3.205	2.479	1.489	2.207
	(3.39)***	(2.50)**	(1.70)*	(2.38)**
Broadband			2.15	2.79
			(1.68)*	(2.43)**
Average Income (000)	0.018	0.001	0.013	0.002
	(1.78)*	(0.07)	(2.08)**	(0.39)
Population (000,000)	0.0027	0.0021	0.0010	-0.0009
	(3.13)***	(2.41)**	(0.82)	(0.93)
Radio Music (hours per day)	-0.387	-0.215	-0.564	-0.771
	(1.57)	(0.77)	(1.70)*	(2.43)**
Share of 12-29	-0.281	-2.782	-1.952	-4.611
	(0.27)	(1.80)*	(1.97)*	(2.75)***
Share of Males	-0.398	-0.024	-1.612	-3.761
	(0.22)	(0.01)	(0.66)	(1.58)
55 and over		-3.492		-2.640
		(2.50)**		(1.95)*
College Degree		3.041		1.731
		(2.04)**		(1.26)
Hispanic		-0.161		1.331
		(0.33)		(3.26)***
Black		-0.253		1.204
		(0.48)		(1.90)*
Constant	2.356	3.647	3.340	5.712
	(1.79)*	(2.30)**	(1.72)*	(2.51)**
Observations	95	91	96	92
R-squared	0.47	0.55	0.42	0.57
Robust t statistics in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%				

Table 5 includes a variable on the amount of time spent listening to music-based station on the radio, which is to be distinguished from the earlier radio variable which measured time spent listening to

music or talk. There (section III) we were interested in the impact of Internet use on radio listening and there was no reason to believe it would be different for music or talk. Here we are examining record sales, and it is natural to presuppose that music radio would be a more important influence on record sales than would talk radio. The coefficients indicate that time spent listening to music radio is associated with lower sale of records per capita, with a more pronounced impact in 2003. This is an interesting result, implying as it does that radio is more of a substitute than a complement, and warrants further study.

Table 5 includes the same set of demographic variables as before. Population size is positively related to record sales in 1998 but not in 2003, perhaps because big box retailers or the Internet has helped to equalize consumers' purchase opportunities. Surprisingly (because young people are supposed to be the more intense record purchasers), having more young people has a negative impact on record sales, but with a much greater magnitude in 2003. The share of males has no relationship to record sales in 1998 but a negative one in 2003, although statistical significance is not quite reached. Since age is strongly related to file-sharing activity whereas gender is more weakly related to such activity the change over time of these coefficients is generally consistent with the file-sharing hypothesis.

The additional set of demographics indicates that cities with more old individuals have lower sales of albums although this effect weakens somewhat over time. Cities with a large share of individuals with college degrees have greater record sales although this effect also weakens over time. The minority variables show no relationship to record sales in 1998 but become positive in 2003.

We are now ready for the main regression results. Because they are so important to the analysis the presentation will be expanded to allow various other permutations for dealing with the population and coverage issues. Besides the particular weighting scheme that has been used up to this point, Table 6 also includes regressions with hard cutoffs on the coverage ratio, regressions with pure population weighting

and some regressions using Stata's built-in routine to reduce the impact of outliers (rreg).[9] These will be explained as we move along.

First, a comment on the Internet variable used in Table 6. Although broadband penetration and dialup penetration both exist in the data, and although they have a correlation with each other of -.46, the coefficients and standard errors for each variable were virtually identical in every regressions. In order to keep the analysis as simple as possible the regressions presented in Table 6 will use an overall Internet penetration variable, whose coefficient was always between the already very close separate dialup and broadband coefficients

Table 6 presents results from using a first differenced dependent and independent variables, except for Internet penetration, consistent with the discussion in section III.B of the paper. The dependent variable is the change in albums sold per capita.

The first column shows a regression weighted by the combination of population and coverage. The next two columns present unweighted regressions limited to observations with Coverage ratios above .6 or .75. The next two columns present regression weighted by population and stratified by coverage ratios. The last two columns use the Stata rreg routine for reducing the impact of influential observations on regressions that are otherwise the same as those in columns 2 and 3.

The coefficient on our variable of interest, the share of Internet users in 2003, is negative in all instances, and usually with statistical significance. Cities with the largest increases in Internet use experience the largest declines in record sales. This is a striking difference from either set of single year regressions and indicates the dividends from using procedures that allow for changes over time. Nevertheless, there is substantial variation in the value of the Internet coefficient. The RREG coefficients indicate that the Internet coefficients in columns 2 and 3 are increased by influential observations, a topic we will return to shortly.

[9] This routine first eliminates observations with levels of Cook's D that are above a particular threshold, then it iteratively lowers the weight for observations with large absolute residuals until a convergence threshold is reached.

Table 6: 1998-2003 Differences in Record Sales per Capita							
	Weighted by	OLS		Pop weighted		RREG	
	Cov:Pop	Cov>.6	Cov>.75	Cov>.6	Cov>.75	Cov>.6	Cov>.75
Internet Penetration	-2.530	-2.571	-4.422	-2.398	-3.120	-1.196	-2.039
	(2.18)**	(1.67)	(2.28)**	(2.45)**	(2.71)***	(1.70)*	(2.32)**
Income Change (000)	0.009	0.000	0.017	0.011	0.025	0.004	0.019
	(1.10)	(0.02)	(1.51)	(1.24)	(2.23)**	(0.53)	(1.90)*
Population Change (000,000)	-0.013	-0.008	-0.008	-0.013	-0.019	-0.028	-0.029
	(1.04)	(0.47)	(0.36)	(1.07)	(1.52)	(1.85)*	(1.71)*
Change in Music Radio	-1.024	-1.539	-1.114	-0.981	-0.594	-0.197	-0.335
	(1.88)*	(2.34)**	(1.47)	(1.93)*	(1.27)	(0.58)	(0.72)
Share of Yng Change	1.882	3.861	2.361	1.206	1.262	-2.056	0.786
	(0.80)	(1.36)	(0.80)	(0.48)	(0.49)	(1.54)	(0.46)
Share of Males Change	-1.297	-1.487	-0.861	-0.824	-0.575	1.700	-0.360
	(0.72)	(0.59)	(0.40)	(0.37)	(0.33)	(1.02)	(0.16)
Change in Share of 55+	-0.355	1.769	-2.376	-0.191	-2.708	-2.564	-1.374
	(0.18)	(0.66)	(1.01)	(0.08)	(1.61)	(1.91)*	(0.79)
Change in College Grads	1.148	1.714	-0.385	0.274	-0.963	-0.714	0.432
	(0.53)	(0.63)	(0.12)	(0.11)	(0.40)	(0.45)	(0.19)
Change in Hispanics	0.584	-0.353	-0.958	0.504	-0.173	0.346	-0.272
	(0.33)	(0.17)	(0.45)	(0.26)	(0.09)	(0.25)	(0.16)
Change in Blacks	-0.990	-1.805	1.450	-0.618	1.043	1.030	0.782
	(0.63)	(1.07)	(0.83)	(0.37)	(0.60)	(1.02)	(0.52)
Constant	0.653	0.561	1.817	0.592	1.153	0.273	0.632
	(1.00)	(0.63)	(2.03)**	(0.96)	(1.84)*	(0.64)	(1.22)
Observations	89	71	54	71	54	70	54
R-squared	0.15	0.22	0.25	0.17	0.28	0.18	0.23
Robust t statistics in parentheses (not for rreg); * significant at 10%; ** significant at 5%; *** significant at 1%							

Table 6 also indicates that income increases are positively related to increases in record sales, but the precision only allows spotty statistical significance. Population increases seem to be negatively related to changes in per capita record sales although it is not clear what mechanism might be at work here. Increases in the audience for music radio are negatively associated with record sales, consistent with the finding in the yearly regressions that cities with greater music radio listenership have lower record sales. It would appear that radio is mainly a substitute for record sales and not a complement, although this result appears to be impacted by influential observations. There is little consistency among the coefficients for the other demographic variables.

Our main interest is in determining the overall relationship between file-sharing and record sales. Although there appears to be considerable variation in these coefficients, the Internet coefficients in the RREG regressions are lower largely due to their eliminating a single observation. With that observation removed from all regressions the coefficients in Table 6 would range from -1.4 to -2.7.[10] The regressions weighted by population provide the results most appropriate for our purpose since larger cities have a larger impact on overall record sales and these regressions have the additional virtue of giving more weight to the larger cities which, due to the nature of the sampling, are likely to experience the smallest measurement errors. These population weighted coefficients average -2.75 with all observations and -2.3 with the one observation removed. In picking a single value the mean of these coefficients will be rounded to -2.5.

Table 7 performs the calculations to determine the overall impact of filesharing. Row 1 contains the Internet use coefficient which has been rounded to -2.5. The product of the Internet coefficient and the Internet penetration (weighted value is 62%) implies that the total impact of the Internet would have been to have lowered record sales by about 1.55 units per capita from 1998 to 2003, which is found in row 3 of Table 7.

Table 7: Impact of File-Sharing	
1. Average Coefficient from Population Weighted Regressions	-2.50
2. Weighted Average Internet Use 2003	0.62
3. Product of Regression Coefficient and 2003 Internet Use: Measures Impact of Internet Use on Album Sales [1*2]	-1.55
4. Decline Due to 9% Internet Entertainment Impact [8*9%]	-0.26
5. Net: Decline in 2003 due to File-Sharing [3-4]	-1.29
6. 2003 weighted average Album Sales per Capita	2.44
7. Predicted 2003 sales without file-sharing [6-5]	3.74
8. 1999 weighted average album sales per capita	2.90
9. Yearly growth rate to get from 1999 to predicted 2003	5.23%

Next we must disentangle the general impact of the Internet on entertainment use estimated earlier. The estimated decline in media usage caused by the Internet during 1998-2003 was 7% for radio use and

[10] The observation is Colorado Springs-Pueblo, which is the second smallest city. Nevertheless, there is no obvious defect with the observation.

13% for television. Since file-sharing was in place for only four of those six years (treating 1999 as a non-filesharing year), we will limit the analysis of file-sharing's impact to the four years after 1999. Therefore this requires scaling down the generic Internet entertainment impact by 33%. Although one can argue that radio is a more appropriate proxy than is television we will use the larger television number (13%), to be conservative, scaled down to 9%. This leads to a decline of .26 units per capita found on line 4, calculated as the 9% decline from the 1999 value (row 8). Subtracting the generic entertainment decline from the overall decline leaves us with a net decline attributed to file-sharing of 1.29 sound recording albums per capita, found in row 5.

Row 6 takes the actual weighted 2003 sales per capita in our 99 cities and adds the 1.29 records presumably lost to file-sharing to arrive at a sans file-sharing estimate of 3.74 records per capita in 2003, which is well above the actual sales of 2.44. The last row in Table 7 calculates the growth rate needed to get from the 1999 values to the 2003 predicted values. Once again, we use 1999 and not 1998 because the last year of sales largely unaffected by file-sharing is 1999. Record sales would have needed to grow at 5.23% per year to arrive that the projected value.

This is slightly above the average growth rate during 1973-1998 of 4.46% and considerably greater than the cumulative growth rate of 2.75% (both of these numbers are based upon RIAA data). Nevertheless, it is quite common (about 25% of the time) to find this level of implied four-year cumulative growth percentage for historical intervals of four or even fewer years, putting this result well within the norm of historical sound recording trends.[11]

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 appears to be greater than the decline in CD sales that has occurred. Although the usual caveats from basing conclusions on any single analysis clearly apply, the direction of the results from these regressions

[11] In the time span 1973-1999 there were 6 (out of 22) four-year periods, 2 three-year periods, and 1 two-year period that had this level of cumulative growth.

are consistent with the overall impact found in other studies even if the more detailed calculations and conclusions about the size of the impact are novel to this paper.

V. 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. If regressions over our 99 cities indicate that the Internet has differential impacts on record sales by genre in accordance with our beliefs about file-sharing, we will have an additional independent test.

Different populations are likely to be attracted to different genres of music. Classical music and jazz, for example, would be expected to appeal to older individuals less likely to be engaged in file-sharing, whereas hard rock and rap would be expected to appeal to younger individuals more prone to engage in file-sharing.[12] Our expectations for country music are less clear. [13]

The use of these data is somewhat problematic, however, because albums are often classified in more than one genre. Since albums can belong to multiple genres and the process of album classification and dual-classification appears somewhat arbitrary, yearly fluctuation may not properly represent changes in market conditions so much as changes in judgment calls about classifications. Two examples of this are found in Table 8, which lists seven genres reported by SoundScan. The extreme changes in the Hard Rock

[12] Oberholzer and Strumpf (2005) report data that, if reflective of the entire market, would allow somewhat greater precision in determining which categories are most heavily downloaded. Unfortunately their data on genres do not include the top selling albums and so they are of questionable value. Nevertheless, using numbers from their Tables 1 and Table 3 it is possible to construct a ratio of downloads to sales for the albums in their sample, by genre. The results: Hard 47; Alternative 100; R&B 77; Rap 29; Country 21; Jazz 4.

[13] 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.

(Metal) category beginning in 1999 and the R&B category in 1997 were due to alterations in genre definition.[14] Nevertheless, with this caveat in place, we can continue our examination of the genre data.

Table 8: Album Sales (000s)							
	Alternative	Classical	Country	Hard Rock	Jazz	R&B	Rap
1994	82,164	27,003	75,976	38,739	16,546	80,819	40,995
1995	94,004	23,836	76,095	31,101	14,797	80,718	41,537
1996	105,175	21,456	66,883	26,409	21,794	74,035	56,343
1997	106,690	19,148	70,702	28,983	20,042	141,613	61,709
1998	116,489	16,948	74,043	30,086	18,123	166,379	83,641
1999	120,952	17,311	69,300	82,698	19,557	175,339	87,663
2000	131,138	16,403	67,115	89,924	18,416	197,141	105,515
2001	131,594	15,846	67,241	88,158	19,514	195,498	89,279
2002	125,752	14,776	75,362	74,677	19,901	160,183	83,346
2003	128,344	17,727	70,944	74,629	22,366	149,972	75,854
% change 00-03	-2.13%	8.07%	5.71%	-17.01%	21.45%	-23.93%	-28.11%

Table 8 allows a cursory examination of total sales changes by genre. Three genres increased in absolute terms from 2000 (the peak in SoundScan data) until 2003—Classical, Country, and Jazz. These results are generally consistent with a view that file-sharing has a negative impact on sales since the genres least susceptible to file-sharing, classical and jazz, each increased during a period of decline. But these are only raw numbers.

We now repeat our econometric analysis for each genre of music for the period 1998-2003. Regressions measuring the impact of file-sharing were run using genre-based sales per capita as the dependent variable. Six regressions were run for each genre, based on two categories of Coverage (>.6, >.75) and regression type (regular OLS, OLS weighted by population, and the Stata RREG procedure to underweight outliers) used previously. Only the youth and male demographic variables were included in the reported results since they have the strongest rationale and because the demographic variables were not important in the full-market regressions (adding the other demographic variables or a regression weighted by coverage/population didn't have a material impact on these results).

[14] Correspondence with SoundScan officials has confirmed that Rap albums were first eligible to be flagged as R&B beginning in 1997 and Alternative albums were first eligible to be flagged as Hard (Rock) beginning in 1999. The SoundScan representative claimed that there were no other changes during this period.

Table 9: Estimated Internet Impact on Per Capita Sales 1999-2003					Actual Sales	
	Average	Max	Min	Comments	00-03	98-03
Classical	-6.46%	-35.09%	21.36%	3+	5.07%	-0.40%
Country	-7.74%	-20.18%	0.92%	1 +	2.71%	-9.19%
Jazz	-14.51%	-21.31%	-4.14%		18.45%	18.41%
Hard Rock	-22.02%	-32.32%	-13.52%	1 @ 10%	-20.01%	-13.76%
Alternative	-32.36%	-50.72%	-8.19%	1@1%, 2@5%	-5.13%	1.11%
R&B	-40.12%	-57.56%	-14.55%	3 @ 10%	-26.93%	-14.86%
Rap	-45.26%	-62.97%	-19.95%	3 @ 10%	-31.11%	-14.31%
x @ y% means there were x regressions where coefficient was significant at the y% level; n+ means n regressions had a positive coefficient						

Our main interest is in the relative size of the file-sharing coefficients for different genres. For the sake of brevity, only a summary of the results is shown in Table 9. The numbers in Table 9 include the largest, smallest and average of the six Internet coefficients for each genre. The raw coefficients are then normalized to make them more easily comparable. The regression coefficients (for the Internet use variable) were multiplied by the 2003 Internet penetration rate to generate a predicted impact of the Internet on sales per capita for that genre. Then these numbers were further scaled by the sales *for that genre* in 1999 (the last year mainly untainted by file-sharing). The resulting numbers, found in the table, give the estimated impact of file-sharing as a percentage of 1999 sales for each genre. For example, the number in Table 9 for the 'average' coefficient in the musical category 'Alternative' would imply, at the 2003 Internet penetration rate, that sales were diminished by 32.36% of the 1999 sales. The actual changes in sales (per capita) are shown in the last two columns, for comparison purposes. The numbers in the column for 00-03, which represents the maximum decline from the peak year of SoundScan measured sales, are slightly different than those in Table 8 because Table 9 measures sales *per capita* and population was growing about 1% per year.

Examination of the pattern of coefficients reveals that these results are consistent with our prior findings. First, remember that there is likely to be some impact on sales due merely to the fact that the Internet decreases the time spent on other entertainment activities. Classical and Jazz, the two genres that seem least likely to be impacted by file-sharing, would appear to have a file-sharing impact close to zero if the overall entertainment impact of the Internet for these genres is anywhere near the average of all

genres. Although both of these genres have at least one instance where the coefficient implies a fairly large negative impact, the coefficients are imprecisely measured.

The four categories that would be presumed most likely to be impacted by file-sharing are Hard Rock (Metal), Alternative, Rap, and R&B (many Rap albums are cross-listed with R&B). The coefficients for these categories are always negative. There are a fair number of statistically significant coefficients and many other coefficients are fairly close to borderline significance for this group. The average adjusted coefficients are all considerably larger than the expected entertainment impact of the Internet, indicating a fairly large negative impact of file-sharing. Finally, the average coefficient for Country music implies that country music has fans who walk the line and do not engage much in file-sharing, a not implausible finding.

The numbers in the last two columns show that the estimated Internet impact better matches (correlation >81%) actual declines from the years when filesharing was becoming popular (00-03) than it matches (<47%) the sales change over the actual period used (98-03), showing the power of this first differenced approach to target file-sharing behavior. These data also imply that Alternative and Jazz appear to have had idiosyncratic shifts in demand unrelated to Internet use.

In general, the results from the genre regressions support a conclusion that file-sharing is harmful to record sales. The magnitudes of the genre declines are in line with the overall-market regressions. The genre regressions predict larger losses than actually occurred, which corresponds to the vitiation of sales gains found for the full-market regressions. It is true that the coefficients on Internet use for the genre regressions are not measured with the same precision as was the case with the entire market, lowering our confidence in them, but the story that they tell is consistent with everything else we have seen.

VI. Conclusions

Has the large decline in record sales that coincided with the advent of file-sharing been caused by file-sharing? We have examined album sales in the leading American cities over a five year period of

time that begins just prior to the genesis of file-sharing and our results indicate that file-sharing indeed has caused the decline.

In order to come to this conclusion we needed to first calculate a generic impact of Internet use on entertainment products caused by the fact that for many individuals the Internet is a substitute form of entertainment. Examining the impact of Internet penetration on usage of the two most popular forms of entertainment, television viewing and radio listening, led to a conclusion that the Internet did indeed reduce consumption of these activities, although by a relatively small 7 to 13 percent. This finding is, to my knowledge, the first estimate of the Internet's impact on these activities based upon nationwide statistics.

The penultimate stage of the analysis then required finding the relationship between Internet penetration and changes in record sales across 99 large market areas. The results indicated a negative impact after controlling for other factors that might impact file-sharing.

The final stage of the analysis required removing the generic impact of the Internet from the specific overall impact, to arrive at an estimate of the file-sharing component. The reduction in sales due to file-sharing appears to be larger than the actual measured decline in record sales—the regression results indicate that file-sharing not only reduced sales but also vitiated an increase that otherwise would have occurred. The regressions imply that except for filesharing, there would have been an increase in record sales from 1998 through 2003 that was quite close to the historical industry average.

As an additional piece of evidence, this analysis was repeated for individual genres of music. The results were consistent with basic intuition about different genres' affinity for file-sharing. This would rule out alterative explanations of the sales decline that didn't point to the same set of culprits as those involved with file-sharing.

One possible limitation of the methodology used here should be mentioned. This paper has taken Internet use as a proxy for file-sharing. If it were the case that Internet users, for some hypothetical reason

other than file-sharing, lost their taste for music, and particularly for the categories of music that we currently think are likely to be related to filesharing, the methodology used in this paper would inappropriately assign the resulting decline in CD sales to file-sharing. Of course such a hypothetical would require that some other factor takes hold at the same time as file-sharing, that it have the same impact as file-sharing is hypothesized to have, that it impact genres the same way file-sharing would be expected to and yet that this factor remain invisible. This seems a remote possibility. It is difficult to think of any potential candidates, particularly since Internet users listen to music while on the Internet. It has been casually suggested that video games or DVDs might fit the bill. When these alternatives have been more carefully examined (Liebowitz 2004b, 2006), however, there was no evidence that they were undergoing trend changes like those occurring in file-sharing and record sales. We might also note that if Internet users have lost their taste for music the recent explosion in the purchase of expensive iPods would be very difficult to explain. So this hypothetical limitation to the methodology remains merely hypothetical.

The findings in this paper confirm the worst nightmares of the RIAA members. They also indicate that the Internet might have a larger dark side than has been previously supposed and than its legions of supporters might wish to admit. This would be the damaging of business models, not by a superior replacement but by a parasitic use of an otherwise productive technology. These problems may become far worse if other markets, such as movies and computer software, were to be sucked up in the file-sharing vortex. Of course, the Internet itself is not to blame. But it is useful for analysts to keep an open mind about the ramifications of some Internet activities. The impacts of the Internet need not be entirely benign.

Bhattacharjee Sudip, Gopal Ram D., Lertwachara Kaveepan, and. Marsden, James R. "Impact of Legal Threats on Music Sharing Activity: An Analysis of Music Industry Legal Actions" Journal of Law and Economics, 49: April 2006

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

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

IDATE. 2003. "Taking Advantage of Peer-to-Peer: What Is at Stake for the Content Industry?" IDATE News No. 282. IDATE, Montpellier, October 28. http://www.idate.fr/an/_qdn/an-03/IF282/index_a.htm.

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

Kraut Robert, Sara Kiesler, Bonka Boneva, Irina Shklovski "Examining the Impact of Internet Use on TV Viewing: Details Make a Difference", in Computers, Phones, and the Internet, edited by Robert Kraut, Malcolm Brynin and Sara Kiesler, Oxford University Press, 2006.

Liebowitz, Stan J. "File-Sharing: Creative Destruction or Just Plain Destruction? *Journal of Law and Economics*, April 2006.

Liebowitz, Stan J. "Pitfalls in Measuring the Impact of File-sharing." *CESifo Economic Studies*, July 2005. http://ssrn.com/abstract=583484

Liebowitz, Stan J. "The Elusive Symbiosis: The Impact of Radio on the Record Industry" *Review of Economic Research on Copyright Issues* Vol. 1, 2004, 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.

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.

Nie, Norman H., Alberto Simpser, Irena Stepanikova, and Lu Zheng, "Ten Years After The Birth Of The Internet, How Do Americans Use The Internet In Their Daily Lives?" Stanford Center for the Quantitative Study of Society, November, 2005.

Oberholzer, Felix & Koleman Strumpf "The Effect of File Sharing on Record Sales: An Empirical Analysis" working paper, June 2005.

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, M. and P. Waelbroeck (2004), "The Effect of Internet Piracy on Music Sales: Cross-Section Evidence", Review of Economic Research on Copyright Issues, 1(2); 71-79.

Pew Internet Life Project**** Pew Internet & American Life Project. (2003). Usage over Time, from the World Wide Web: http://www.pewinternet.org/trends/UsageOverTime.xls

Rob, Rafael and Joel Waldfogel, "Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students", Journal of Law and Economics, 49: April 2006.

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.

The Digital Future Report, USC Annenberg School, Center for the Digital Future, September 2004.

TRAC Media Services "A Report on Ethnic Viewing", February 2001, Funded by the Corporation for Public Broadcasting. Available at http://stations.cpb.org/system/reports/ethnicviewingreport.pdf.

Zentner Alejandro (2005) "File Sharing and International Sales of Copyrighted Music: An Empirical Analysis with a Panel of Countries", Topics in Economic Analysis & Policy: Vol. 5: No. 1, Article 21. http://www.bepress.com/bejeap/topics/vol5/iss1/art21

Zentner, Alejandro "Measuring the Effect of Music Downloads on Music Purchases" *Journal of Law and Economics,* April 2006.