Research Note: Testing File-Sharing's Impact on Music Album Sales in Cities

Stan J. Liebowitz
School of Management, University of Texas at Dallas, Richardson, TX 75083, liebowit@utdallas.edu

Using a data set including album sales, Internet penetration and various demographic measures for 99 American cities over the period 1998- 2003 this paper empirically examines the extent to which file-sharing has caused the US decline in sound recording sales over that period. Also examined is the impact of the Internet on entertainment activities so as to help cleanse the Internet penetration coefficient of that impact. The conclusion from this analysis is that file-sharing appears to have caused the entire decline in record sales and appears to have vitiated what otherwise would have been growth in the industry.

Key words: File Sharing, Peer-to-Peer, Copying, Internet, Sound Recordings, Television, Radio

History: This paper was first submitted September 2005.

I. Introduction

File-sharing is merely the most recent example in a long line of technologies (e.g., photocopying, audio and video taping) that have lowered the cost of unauthorized copying by individuals. Although each of the previous copying technologies engendered cries of alarm from the industries owning the endangered copyrights, there was no sustained decline in sales to support the rhetoric coming from these industries. Unlike prior copying technologies, however, the growth in file-sharing has been accompanied by a large drop in sales of sound recordings. Managers and policy-makers are still nervously trying to determine the best strategies to deal with file-sharing.

Organized file-sharing began with Napster in late 1999. Although Napster was effectively shut down within two years of its birth, its progeny live on and the repercussions on music listening and the music

industry have not yet run their course. By most estimates, the quantity of music files that are shared is enormous. Liebowitz (2006) documents many estimates of the extent of and trends in filesharing.

The unauthorized downloading of a copyrighted song can easily be seen as a substitute for the purchase of that song, leading, in such instances, to a negative financial impact on the sound recording industry. Nevertheless, when Napster first became popular, most downloaders would not have had in place the CD burners and MP3 players that would break the shackles of listening to downloaded music tethered to a computer. For these reasons, MP3 files were not, at the time of Napster, terribly good substitutes for music purchased on a CD, although over the next few years downloaded files became increasingly better substitutes for the purchase of prerecorded music.

There are other, more ambiguous, impacts of filesharing, however, as also discussed in Liebowitz (2006). Determining the impact of filesharing, therefore, requires an empirical exercise. Most recent research, such as Hong (2006), Rob and Waldfogel (2006), and Zentner (2006), has found that the impact of file-sharing is negative although the methodology used in these papers generally makes it difficult to determine clear estimates of the portion of the recent decline in sound recording sales that might be caused by file-sharing.[1] Oberholzer-Gee and Strumpf (2007) is, to my knowledge, the only paper to find a benign impact of file-sharing although Liebowitz (2007) suggests there are important problems with that paper.

The goal of this paper is to estimate the impact of file-sharing on sound recording (album) sales in the United States, the world's largest market for sound recordings. The procedure used in this paper allows the calculation of an overall national impact of file-sharing and in doing so attempts to adjust for the Internet's entertainment impact on the sound recording market. The conclusion is that file-sharing has caused the recent decline in album sales and appears to have vitiated an increase that otherwise would have occurred.

[1] The e-companion includes a much fuller description of the literature.

II. Econometric Specification and the File-sharing Variable

The empirical objective is to determine whether file-sharing impacts the sales of albums and if so, by how much. The unit of analysis is a "city" as defined by AC Nielsen which uses the term "designated metropolitan area" (DMA). The plan is to estimate a first differenced regression across *i* cities of the following form:

$$(1) \qquad \Delta RS_i = \beta \Delta FS_i + \gamma \Delta Z_i + \Delta u_i$$

where RS stands for record (album) sales, FS stands for file-sharing and Z is a vector of other explanatory variables. Our main interest is in the size of β.

Unfortunately, there is no direct measure of file-sharing. Instead, Internet penetration will be used as a proxy for file-sharing. There are several difficulties in using Internet penetration as a proxy for file-sharing. First, Internet penetration reflects the number of users, not their intensity or frequency of use, although it seems reasonable to expect the city-wide Internet usage and file-sharing to be strongly related to city-wide Internet penetration. Second, Internet penetration is likely to reflect all net-based forms of 'piracy', including transmitting songs by email or instant messaging, therefore measuring more than just the impact of anonymous file-sharing. Since these latter forms of sharing tend to require personal knowledge of the music donor music, however, they are will tend to replace old fashioned forms of piracy (e.g., cassette tapes) so that some of their impact would have already been built into the sales statistics.

Finally, the Internet can also be a form of entertainment competing with sound recordings for the attention of the entertainment consumer. This means that the Internet penetration coefficient will include both a file-sharing impact β as well as a potential impact from the Internet as a new and growing form of alternative entertainment. This last factor would lead to an overly large negative estimate of β. An attempt to separate out the entertainment impact is undertaken in Section V.

A different problem arising from the use of Internet penetration is that the amount of file-sharing is related to the product of the number of Internet users and the propensity of those Internet users to engage

in file-sharing. This is represented in equation (2) where FS stands for the quantity of file-sharing, IU stands for the number of Internet users and FP stands for average filesharing propensity.

$$(2) \quad FS \equiv IU \bullet FP$$

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

$$(3) \quad FS_{\text{it}} - FS_{it-1} = IU_{it} \bullet FP_{it} - IU_{it-1} \bullet FP_{it-1}$$

Since the only one of these rhs variables for which data exist is IU, it is not possible, in general, to measure the change in file-sharing without making some further assumptions about file-sharing propensity.

The strategy adopted in this paper is to go back to a period where file-sharing propensity was zero, which is taken to be anytime prior to Napster, which began in late 1999. If FP_{t-1} is equal to zero then (3) 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.

$$(4) \qquad \Delta FS_i = IU_{it} \bullet FP_{it} \ .$$

With this strategy the number of internet users can be taken as a proxy for file-sharing if file-sharing propensity (for Internet users) is assumed constant across cities at a moment in time, or if it is possible to control for the factors that might lead to different file-sharing propensities across cities. Under this assumption, equation (1) is transformed into (5), below.

$$(5) \qquad \Delta RS = \beta\, IU_t + \gamma \Delta Z + \Delta u \, .$$

This is not quite a standard fixed effects or first differenced regression since the first term on the rhs is not a normal first difference. Even though the first term is, by construction, a first difference of sorts, the construction forces the first period value to zero in the early period. This provides less information than a typical first difference where there is information contained in every first period observation, eliminating controls for fixed effects in the Internet variable. Yet, as noted, it would be incorrect to use a traditional

first difference on the Internet variable in such a regression because all Internet users first obtained the ability to engage in file-sharing during this interval, not just the new users.

It is necessary to choose a timeframe that will be consistent with the analysis above. As discussed in the next section, the most recent Census on Internet use occurred in October of 2003. This will provide the final year in the analysis. The first such survey, in December 1998, occurs when file-sharing is zero, and is the last survey conducted before Napster's appearance, so it becomes our starting date.[2]

The analysis below takes as given that the supply of records and the price of records are the same in all cities, which appears to be a reasonable assumption. First, the list price is the same throughout the country. Second, a majority of record sales occur in national retail chains (according to data on the RIAA website). Discussions with industry executives indicate that chains selling CDs have the same transaction price on a particular CD in all their stores throughout the country (except in the very smallest localities). Internet sites also provide uniform prices to customers throughout the country.

III. The Data

The approach in this paper is to examine the sales of sound recording albums in leading American cities over the period 1998-2003. Conducting the econometric investigation requires the merging of several data sets. First, the US Census, through its Current Population Surveys, conducted surveys on Internet and Computer use in 1998, 2000, 2001, and 2003.[3] These surveys provide information on Internet use, income, age, sex, race, gender and education for the inhabitants of Census metropolitan areas known as

[2] The e-companion provides regression results for subperiods. Consistent with expectations, the impact of Internet usage on albums sales is considerably larger in the later period than it is in the early period.

[3] The e-companion provides details on where to find and download this data and also detailed descriptions of each variable used.

MSAs. Second, Nielsen SoundScan sells data on album sales for geographic areas (including metropolitan and rural areas) which it refers to as DMAs.[4]

I purchased data for the 100 largest DMAs which contain approximately 83% of the total population of the country, although one (Paducah, Kentucky) was not matched to any MSA. Finally, Nielsen Media Research sells data on television viewing in DMAs and Arbitron collects data on radio listenership in its own set of metro areas, which are similar to MSAs. These last two data sources are used when examining time devoted to alternative forms of entertainment.

Combining these data sets is not a trivial task.[5] Nielsen DMAs (records and television) are larger and more populous than Census or Arbitron (radio) metro areas and Nielsen DMAs often contain several smaller MSAs or metro areas. MSAs (Metro areas) were matched to DMAs as closely as possible and a variable called "Coverage" was created to measure the portion of the DMA population contained in the metro areas included in a DMA.[6] When Coverage is low it is possible that the socioeconomic data from the MSAs included in a DMA will not properly reflect the true population characteristics of the DMA because the included MSAs leave out a large portion of the DMA population. Also, when MSA population is low the Census measurements are known to be less precise. The regression analysis below attempts to reduce the impact of such questionable observations.

Table 1 presents summary statistics for 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. Weighted averages are included for comparison purposes as well as averages for the 17% of the population too rural to be in the

[4] 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.

[5] A detailed description of the construction of the final data set from its components, including links for retrieving the data, is found in the e-companion.

[6] Separate Coverage ratios were calculated for Census MSAs and Arbitron metro areas. The coverage ratios were usually quite close and the one used in the regressions was the lower of the two.

top 100 DMAs. The population of the MSAs matched to the DMAs covers about 83% of the population of the included DMAs (coincidentally almost the same number as the top 100 DMAs share of the total population). Consistent with industry claims, album sales per capita fell. This decline, from 1998, does not reflect the full extent of the decline from the peak value since album sales were still growing in 1998 and 1999. The share of the population with Internet access can be seen to basically double from 1998 through 2003.

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	0.248	0.065	0.120	0.420	0.259	0.15
Coverage Ratio	99	0.728	0.220	0.203	0.999	0.828	
Dialup Share	99	0.366	0.064	0.170	0.500	0.362	0.40
Music Radio Listening (hours per day)	96	2.335	0.193	1.866	2.815	2.293	
Radio Listening (hours per day)	96	2.723	0.174	2.373	3.233	2.77	
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	-0.4933	
Average Income Change	99	8,523	7,087	-6,660	26,901	9783.48	
Dialup Share Change	99	0.063	0.079	-0.200	0.250	0.048	
Music Radio Listening Change	95	-0.323	0.123	-0.623	-0.036	-0.282	
Population Change (%)	99	6.71%	6.57%	-6.00%	35.00%	7.81%	
Radio Listening Change	96	-0.294	0.104	-0.600	-0.050	-0.272	
Share of Internet Users Change	99	0.310	0.058	0.120	0.466	0.307	
Share of Males Change	99	0.001	0.035	-0.137	0.143	0.000	
Share of Population 12-29 Change	99	0.001	0.045	-0.110	0.140	0.000	
Share of Population Hispanic Change	99	0.023	0.038	-0.070	0.210	0.024	
Share 55+ Change	99	0.011	0.047	-0.120	0.191	0.012	
Share College Change	99	0.018	0.040	-0.114	0.208	0.022	
Share Black Change	93	-0.004	0.043	-0.131	0.104	-0.003	
TV Viewing Change	99	0.060	0.355	-0.552	1.224	0.066	

IV. Estimation

Table 2 presents the results from our attempt to implement the regressions representing equation 5.[7] All variables are in first differences except for Internet penetration as discussed above (regressions for individual years are in the e-companion). The dependent variable is the change in annual albums sold per capita.

The first column shows regression results for the entire data set, including DMAs with poor data coverage. This regression specification, referred to as the 'naïve regression,' is shown only for completeness and will not be part of the analysis since the inclusion of DMAs with poor coverage is highly questionable. Due to the possible data problems when coverage or population low, the second column shows a regression weighted by the combination of population and coverage.[8] The next two columns present unweighted regressions limited to observations with Coverage ratios above .6 or .75 where the cutoffs were chosen to keep the number of observations as high as possible while removing poorly measured observations.[9] The final two columns present the regressions limited to observations with those coverage ratios but now also weighted by average MSA population within the DMA.

[7] One observation, Pueblo/Colorado Springs, was removed. It had a disproportionate impact on several coefficients (including making the Internet variable more strongly negative) and reduced the fit considerably. Removal also brought the results into closer alignment with Stata's robust regression routine (RREG). Results for the entire sample are in the e-companion.

[8] The weight is the product of Coverage squared and the square root of population which tended to give equal weight to both factors.

[9] Limiting observations to those with coverage >.6 (>.75) reduces the summed population of the remaining DMAs by 11% (20%).

Table 2: 1998-2003 Differences in Album Sales per Capita						
			OLS		Population Weighted	
	Naïve OLS	Weighted by Cov&Pop	Cov>.6	Cov>.75	Cov>.6	Cov>.75
2003 Internet Penetration	-0.551	-2.072	-2.201	-2.885	-2.562	-2.456
	(0.60)	(2.19)**	(1.75)*	(2.57)**	(2.39)**	(2.12)**
Income Change (000)	0.005	0.021	0.025	0.031	0.027	0.035
	(0.61)	(2.29)**	(1.77)*	(1.92)*	(2.31)**	(2.50)**
Population Change (00,000)	-0.010	-0.012	-0.012	-0.003	-0.017	-0.022
	(0.53)	(0.78)	(0.65)	(0.11)	(1.41)	(1.60)
Change in Music Radio	-0.325	-0.845	-1.109	-1.716	-0.436	-0.466
	(0.41)	(1.49)	(1.61)	(2.10)**	(0.82)	(0.67)
Share of Young Change	4.654	4.673	8.024	6.877	4.515	4.332
	(1.66)	(1.55)	(2.04)**	(1.50)	(1.35)	(1.15)
Share of Males Change	-1.235	-2.882	-5.449	-5.497	-2.895	-1.869
	(0.62)	(1.25)	(1.87)*	(1.55)	(1.13)	(0.71)
Change in Share of 55+	1.514	1.402	5.247	1.617	1.640	-0.443
	(0.73)	(0.70)	(1.65)	(0.48)	(0.59)	(0.20)
Change in College Grads	2.595	3.372	4.847	7.244	2.899	3.326
	(1.29)	(1.21)	(1.55)	(1.52)	(0.97)	(0.94)
Change in Hispanics	0.756	0.077	-0.153	-1.562	0.953	-0.022
	(0.45)	(0.05)	(0.07)	(0.86)	(0.49)	(0.01)
Change in Blacks	-3.018	-1.225	-0.645	-1.449	-0.297	-1.118
	(1.80)*	(0.89)	(0.41)	(0.80)	(0.19)	(0.66)
Constant	-0.404	0.286	0.201	0.435	0.661	0.602
	(0.60)	(0.48)	(0.28)	(0.69)	(1.01)	(0.91)
Observations	89	89	60	40	60	40
R-squared	0.18	0.24	0.4	0.52	0.24	0.38
Pueblo removed; Cov=Coverage, Pop=population; t-statistics based on robust standard errors in parentheses; * significant @ 10%; ** significant @ 5%; *** significant @ 1%						

The regressions presented in Table 2 indicate that the coefficient on our variable of interest, the share of Internet users in 2003, is negative in all instances, and with statistical significance.[10] This tells us that cities with the largest share of Internet users experience the largest declines in per capita record sales. The

[10] Although 2003 broadband penetration and dialup penetration are both in the data (correlation -.41), when both variables were included in the regression the coefficients were always very similar (see the e-companion for the regressions). The overall Internet coefficients were always between the separate dialup and broadband coefficients. Although somewhat surprising, dialup speeds are capable of allowing the operator to download vast quantities of songs in the background, so the extra speed may not be as important for filesharing music as for other Internet activities.

typical coefficient, approximately -2.43 (excluding the naïve regression), implies that an Internet usage of 50% would reduce record sales by 1.21 units per person per year, which is quite large relative to the 2000 sales level of 2.86 albums per capita. This is discussed in more detail in Section VI.

Table 2 also indicates that income is positively related to record sales, as would be expected. The increases in the audience for music radio appear to be negatively associated with record sales and the share of young people and college graduates appear to be positively related to album sales while the share of males is negatively related, although these findings do not generally have statistical significance.

Similar sets of regressions were run for different musical genres. Those genres with the most intense file-sharing activity had the largest negative relationship between Internet usage and sales declines, appearing to further confirm these findings. Nevertheless, there is some uncertainty that the Internet variable in those regressions was picking up file-sharing activity as opposed to an entertainment aspect of the Internet. The results are reported and discussed in detail in the e-companion.

V. Does Internet Use Impact Time Spent on Entertainment?

Since there is no direct measure of file-sharing, the share of Internet users in a city has been used instead. As discussed in above, the file-sharing result could be contaminated if the Internet entertainment that diverts users from listening to sound recordings. Since it is not possible to directly measure the entertainment-diversion impact of Internet activity on the purchases of sound recordings, a less direct approach must be used. The key is to find other similar activities from which the Internet might divert users away and use these as proxies for the impact on record sales. Two activities quickly come to mind—listening to radio and watching television—which are the two most time consuming entertainment activities and together consume over seven hours of the average person's day.[11]

[11] See the average values in Table 1 which closely match information in Table 1102 in the 2002 US Statistical Abstract.

Since listening to music (radio or sound recordings) is largely a background activity (say while driving a car) it is not clear that Internet usage necessarily precludes spending time listening to music. On the other hand sitting in front of the computer would be expected to preclude activities that require foreground attention, such as watching television. Television usage, therefore, seems likely to be more strongly impacted by Internet usage, making the entertainment-diversion impact of the Internet on television something of an upper bound for the Internet's entertainment-diversion impact on sound recording sales.

There is a small survey-based literature that examines the impact of Internet usage on television usage (I am not aware of any such analyses for radio). 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) which would seem to put an upper limit of 33% on the diversion of television time by Internet use. An analysis by Nie et al. (2005) on a non-representative sample finds that Internet use reduced overall TV viewing by 9 minutes per day for all respondents and 30 minutes for the heavier Internet users, leading to reductions of 7.4% and 24.8% respectively for their sample. These measurements represent the overall impact of the Internet and as such should be larger than the estimated marginal impact of the Internet on these activities from 1998 through 2003.

Table 3 contains first difference regressions where changes in television usage (measured by Nielsen and based on DMAs) and radio usage (measured in hours per day as per Arbitron) are the dependent variables. The included demographic variables are the same as before (based on MSAs from the US Census). As before, the regression removes or underweights observations where the MSA populations that are included in a DMA leave out a large share of the DMA population or where MSA populations are small. Unlike before (see footnote 10), separate dialup and broadband Internet variables are used. Dialup is in first differences since the special conditions in equations 2-4 do not hold here. Because broadband penetration was essentially zero in 1998, however, the 2003 broadband penetration level is essentially a

first difference without the information usually contained within a first difference.[12] Also, as before, regression results for individual years are available in the e-companion.

Table 3: First Difference Regressions 1998-2003						
Dependent Varia	Change in TV Viewing Time			Change in Radio Listening Time		
First Differences	Cov & Pop Weighted	Cov > .60	Cov > .75	Cov & Pop Weighted	Cov > .60	Cov > .75
Broadband	-2.044	-2.682	-1.287	-0.720	-0.576	-0.998
	(2.03)**	(1.79)*	(0.76)	(2.86)***	(1.60)	(2.17)**
Dialup	-1.026	-1.304	-0.786	0.034	0.106	-0.207
	(1.31)	(1.05)	(0.58)	(0.15)	(0.37)	(0.55)
Income (000)	0.010	0.013	0.017	-0.002	-0.003	0.000
	(1.67)*	(1.49)	(1.47)	(1.50)	(1.52)	(0.16)
Population (00,000)	-0.0152	-0.0184	-0.0154	0.0094	0.0100	0.0104
	(2.38)**	(1.76)*	(1.35)	(3.83)***	(2.73)***	(2.51)**
55 and over	0.667	1.638	0.836	-0.244	-0.593	-0.049
	(0.49)	(0.96)	(0.38)	(0.78)	(1.52)	(0.09)
Share with College Degree	1.456	0.960	0.738	-0.356	-0.619	-0.246
	(0.91)	(0.49)	(0.28)	(0.84)	(1.16)	(0.33)
Share Males	-1.480	-2.457	-0.960	0.192	0.488	0.080
	(0.69)	(1.01)	(0.24)	(0.45)	(0.92)	(0.13)
Hispanic Share	-1.274	-0.570	-0.799	0.260	0.076	-0.028
	(0.76)	(0.26)	(0.24)	(0.89)	(0.20)	(0.06)
Black Share	0.948	1.111	1.510	-0.201	-0.478	-0.341
	(0.76)	(0.70)	(0.52)	(0.74)	(1.23)	(0.73)
Share Young	-0.141	0.420	0.115	-0.383	-0.613	-0.104
	(0.07)	(0.18)	(0.03)	(0.90)	(1.25)	(0.14)
Constant	0.637	0.805	0.350	-0.090	-0.114	-0.033
	(2.04)**	(1.78)*	(0.75)	(1.37)	(1.21)	(0.27)
Observations	93	62	41	90	61	40
R-squared	0.13	0.15	0.11	0.36	0.31	0.38
T-statistics from Robust Standard Errors in parentheses; Cov = Coverage; Pop = Population; * significant at 10%; ** 5%; *** 1%						

The first three regressions in Table 3 are for television. The key variables, broadband and dialup, are both negative with broadband exerting a stronger impact. The results are only occasionally statistically significant. The last three columns of Table 3 provide the results for radio. Although broadband appears to exert a negative impact on radio usage, dial-up does not. [13]

[12] The FCC reported that fewer than 1 million broadband users existed in the US in 1998. http://www.fcc.gov/Bureaus/Common_Carrier/Orders/2000/fcc00290.pdf

[13] If a single unitary Internet variable is used, in first differences, the television impact is about half of the one in Table 3 and the radio impact disappears.

The coefficients in Table 3 imply a reduction in television viewing and radio listening due to the change in Internet usage during 1998-2003 that is fairly small. Table 4 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 average coefficients from the regressions, the net impact of the Internet is to lower television viewing by about 12.5% and radio by 6.6% over the period 1998-2003. Confidence intervals around these estimates would be quite wide. These results, given their shorter timeframe, seem to fit in quite well with the previous literature.

Table 4: Impact of Internet on TV and Radio Audience 1998-2003

	TV		Radio	
	broadband	dialup	broadband	dialup
1. Coefficient	-2.004	-1.039	-0.765	-0.022
2. 1998-2003 Internet Change	0.259	0.063	0.259	0.063
3. Predicted Impact [1*2]	-0.519	-0.065	-0.198	-0.001
4. 1998 TV or Radio Usage (hrs/day)	4.694	4.694	3.041	3.041
5. % Decline caused by Internet [3/4]	-11.07%	-1.39%	-6.52%	-0.05%

We conclude that increased Internet penetration appears to reduce television and radio usage. If the entertainment diversion impact of the Internet on radio (television) is a reasonable proxy for its impact on sound recordings then the changes in Internet penetration over this time period, independent of file-sharing, would be predicted to cause a decline in record sales of 6.6% (12.5%). Nevertheless, it is possible that the Internet's impact on younger people might be greater than for the average person and thus might have a larger impact on sound recording sales than indicated since young people have a somewhat greater propensity to purchase sound recordings.[14]

[14] According to statistics on the RIAA's web page (for 2001), the share of purchased music from 15-24 year olds was 25.2% whereas it was 21.2% for those aged 25-34, 20.5% for those aged 35-44, 8.5% for those aged 10-14 and 23.7% for those consumers aged 45 and up.

This has been a somewhat cursory, but as far as I can tell the first, market-based analysis of the impact of the Internet on the usage of the two leading forms of entertainment. This is an area begging for further study.

VI. Determining File-Sharing's Impact

The key interest of this paper is in the impact of file-sharing on record sales. Since an overall impact of Internet use on record sales has been calculated, as well as an entertainment impact of the Internet on record sales, it is possible to estimate the file-sharing impact of the Internet by subtracting the latter from the former.

Table 5 performs the calculations for determining the impact of filesharing. It uses three scenarios: a Mid, High and Low estimate of the impact. The Internet coefficients in Table 2 ranged from -2.07 to -2.89 which are rounded to the nearest tenth when choosing the High and Low scenarios in the Table 5. Row 1 contains the rounded Internet regression coefficients and their midpoints. The product of the Internet coefficient and the Internet penetration (weighted value is 62.1%) implies for the Mid (Low, High) case that the overall impact of the Internet has been to decrease record sales by about 1.55 (1.30, 1.80) units per capita, which is found in row 3 of Table 5. A quick glance at Table 1 reveals that these are very large numbers, relative to actual sales.

Table 5: Internet's File-Sharing and Entertainment Impacts	Mid	Low	High
1. Typical Table 3 Coefficient (Internet's Total Impact)	-2.50	-2.1	-2.9
2. Weighted Average Internet Use 2003	0.621	0.621	0.621
3. Product of Regression Coefficient and 2003 Internet Use: Measures Impact of Internet Use on Album Sales [1*2]	-1.55	-1.30	-1.80
4. Decline in Sales due to Entertainment Diversion of Internet [mid 12.5%, low 25%, high 5%; all multiplied by row 8]	-0.37	-0.73	-0.15
5. Net: Decline in 2003 Album Sales Due to File-Sharing [3-4]	-1.19	-0.57	-1.65
6. 2003 Weighted Average Album Sales per Capita	2.44	2.44	2.44
7. Predicted 2003 Album Sales without File-sharing [6-5]	3.63	3.02	4.10
8. 1998 weighted Average Album Sales per capita	2.93	2.93	2.93
9. Yearly Growth Rate to get from 1998 to Predicted 2003	3.63%	0.47%	5.75%

Next, the entertainment diversion impact of the Internet on record sales is calculated. The estimated decline in media usage caused by the Internet during 1998-2003 was 6.6% for radio use and 12.5% for

television. Although I have suggested that radio is a more appropriate proxy than is television, the analysis uses a rounded television number (12.5% as the Mid value) to be conservative. Since these estimates of the entertainment diversion impact of the Internet have wide confidence intervals, two other values are chosen—5% as the value for the High file-sharing impact and 25% as the Low file-sharing impact.

These diversion values, when multiplied by the 1998 album sales, lead to a decline in album sales based on Internet entertainment diversion equal to .37 (.73, .15) units per capita as found in row 4. After deducting the entertainment impact of the Internet from the total Internet impact what remains is a net decline attributed to file-sharing of 1.19 (.57, 1.65) albums per capita, found in row 5.

Row 7 takes the actual (population) weighted 2003 sales per capita found in row 6 and adds the records presumably lost to file-sharing to arrive at a hypothetical sans-file-sharing estimate of 3.63 (3.02, 4.10) records per capita in 2003, which are all well above the 2003 actual sales of 2.44 albums. Row 9 reports the growth rate needed to get from the 1998 level of album sales (row 8) to the 2003 predicted values (row 7). Each of these estimates leads to the conclusion that file-sharing is responsible for a reduction in sales that is larger than the sales decline that occurred and that file-sharing aborted what otherwise would have been a growth in sales. Record sales would have needed to grow at 3.63% (.47%, 5.75%) per year to arrive the projected value(s).

How does this predicted growth compare to historical growth rates? The average yearly growth rate during 1973-1998 was 4.46% and the cumulative growth rate was 2.75% (based upon RIAA data). Indeed, it is quite common to find a 3.63% five-year cumulative growth percentage in historical intervals of five or even fewer years, and even a growth rate of 5.75% is well within the historical norm.[15]

[15] In the time span 1973-1999 there were 12 (out of 22) five-year periods, 11 four-year periods, 7 three-year periods, 5 two-year periods, and 1 one-year period with cumulative growth greater than 3.61% over 5 years. Even for the largest estimate (5.75%) there were 5 five-year periods, 2 four-year, and 2 three-year periods.

The main results are that file-sharing harms record sales and the size of this impact appears to be greater than the decline in CD sales that has occurred. The level of confidence is lower for this latter claim, however.

VII. Conclusions

We have examined album sales in 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 has caused a large decline in record sales. In order to estimate the impact on sound recordings caused by file-sharing it was necessary to remove any generic entertainment impact of the Internet on record sales from the econometric estimates which would have included and such generic impacts. These generic entertainment impacts appeared to be relatively small, in the vicinity of 5-10%.

The estimate of the reduction in sales due to file-sharing appears to be larger than the actual measured decline in record sales. The results 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.

These results are, of course, subject to all the usual limitations found in empirical work of this kind. In addition, there are the specific concerns about the particular data and methodology chosen: the use of three imperfectly linked data sets, the inability to directly measure file-sharing, the non-traditional nature of the fixed-effects model that ensued from the use of Internet penetration as a proxy for filesharing, and the rough nature of the estimate of the entertainment-diversion impact of the Internet. Nevertheless, I have striven to overcome these difficulties, although readers will need to judge my success for themselves.

The findings in this paper appear to confirm the worst nightmares of the recording industry. They also indicate that the Internet is capable of harming business models, not by providing a superior replacement but instead by damaging an otherwise viable older technology. Whether other markets, such as movies and computer software, will be damaged in the file-sharing vortex, is yet to be seen.

Acknowledgements

The author thanks the Center for the Analysis of Property Rights and Innovation for financial support, Stephen Margolis and Alejandro Zentner for advice, Marc Fuhrmann for research assistance, and two anonymous referees and the AE for their insights and patience.

References

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

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

Hong, S. H. 2006. The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster. Working paper, University of Illinois.

Liebowitz, S. J. 2006. File-Sharing: Creative Destruction or Just Plain Destruction? *Journal of Law and Economics* **49**(1) 1-28.

Liebowitz, S. J. 2007. How Reliable is the Oberholzer-Gee and Strumpf paper on File-Sharing? Working paper, Center for the Analysis of Property Rights and Innovation.

Nie, N. H., A. Simpser, I. Stepanikova, L. Zheng. 2005. 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.

Rob, R., J. Waldfogel 2006. 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**(1) 29-62.

Zentner, A. 2006 Measuring the Effect of Music Downloads on Music Purchases. *Journal of Law and Economics* **49**(1) 63-90.

E-Companion:

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

There are 8 components to this e-companion. The topics covered are:

1. Full Literature Review
2. Examining Genres
3. Regressions from periods 1998-2001 and 2000-2003
4. Regressions with both dialup and broadband
5. Individual Year Regressions on Record Sales
6. Individual Year Regressions on Television and Radio Usage
7. Regressions with Pueblo Colorado
8. Construction of the Data Set

1. Full Literature Review

There have been several examinations of the impact of file-sharing, all with their own advantages and disadvantages. Liebowitz (2005) discusses at some length the potential problems with most of the empirical work on this topic.

Liebowitz (2004) examines alternative explanations that might explain the decline in record sales in the United States (such as increases in DVDs, videogames, librarying, music quality, and so forth) and although he rejects all these claims, he provides no direct estimates of the impact of file-sharing. Rob and Waldfogel (2006) find a large negative impact of filesharing based on a survey of self-reported purchases of records and file-sharing activity by American college students at four campuses. Because their results are based on a very narrow sample the authors warn against generalizing their results to the entire marketplace. The self-reported nature of the data is also a cause for concern. Zentner (2006) examines the impact of file-sharing on the proclivity to purchase sound recordings in the EU based on survey data although his estimate does not translate directly into an estimate of the overall reduction in records purchased.

Peitz and Waelbroeck (2004) and Zentner (2005) examine 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. The cross section of countries in these studies compare populations that do not face similar album pricing, do not speak the same language and do not necessarily listen to the same music.

Two papers with similar methodologies to each other, Blackburn (2006) and Oberholzer-Gee and Strumpf (2007), examine sales and downloads of individual songs in the US but reach contrary conclusions with the former finding a large negative impact on overall sales and the latter finding no impact. Each of these two papers needs to overcome a very serious simultaneity problem because the most popular songs are both heavily downloaded and heavily purchased and each uses a different form of instrumental variable to try to overcome this problem. Another problem of unknown magnitude is due to a potential fallacy of composition since file-sharing may have a different impact on individual songs that are downloaded vis-à-vis its impact on the entire market. If file-sharing popularity increases the attention paid to a song, the way radio play does, say, then that song may have an increase in sales because its share of the overall market increases even while the overall market might shrink due to competition from file-sharing. Liebowitz (2007) also questions the reliability of additional evidence put forward by Oberholzer-Gee and Strumpf which the latter authors suggest supports their main conclusion.

Two other working papers based on the Consumer Expenditure Survey, Hong (2006) and Michel (2004), use national data on self-reported purchases of CDs (as opposed to market statistics) and Internet (or computer) use. Hong finds a relatively small impact (less than half the overall decline) and Michel finds a relatively large impact (over half). By using Internet use or computer use as a proxy for file-sharing these studies suffer from the same type of problems as the current study. The data analyzed in Hong end in 2001, which is relatively early in the history of file-sharing and may not reflect its eventual impact. Although Michel's data go to 2003 he uses computer ownership, not Internet access as a proxy, which is further removed from file sharing than the Internet access variable. Further, it is likely that self-reported CD purchases are less accurate than actual market statistics on sales of CDs (they are 60% below the market statistics).

The empirical work in the current paper has certain advantages over some other studies although it also has the weaknesses mentioned in the main text. One advantage is that the data in this study are based on actual sales of record albums in American cities whose inhabitants all function in a single national marketplace with similar sound recording pricing across areas. Another advantage is the examination of

different genres of music so as to provide a separate test for the impact of file-sharing although that material is confined to the e-companion. The study also has data through 2003. It is also the first study of those using Internet use to estimate filesharing which attempts to separate out a generic entertainment-diversionary impact of Internet use. Finally, it is fairly straightforward to translate the empirical results from this study into an overall national impact of file-sharing and to compare that level to the actual sales decline.

References

Blackburn, D. 2006. The Heterogeneous Effects of Copying: The Case of Recorded Music. Working paper. National Economic Research Associates.

Hong, S. H. 2006. The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster. Working paper, University of Illinois.

Liebowitz, S. J. 2005. Pitfalls in Measuring the Impact of File-sharing. *CESifo Economic Studies* **51**(2/3) 435-473.

Liebowitz, S. J. 2004. Will MP3 downloads Annihilate the Record Industry? The Evidence so Far. *Advances in the Study of Entrepreneurship, Innovation, and Economic Growth* **15** 229-260.

Liebowitz, S. J. 2007. How Reliable is the Oberholzer-Gee and Strumpf paper on File-Sharing? Working paper, Center for the Analysis of Property Rights and Innovation.

Michel, N. J. 2005. The Impact of Digital File Sharing on the Music Industry: An Empirical Analysis. Working paper.

Oberholzer-Gee, F., K. Strumpf. 2007. The Effect of File Sharing on Record Sales: An Empirical Analysis. *Journal of Political Economy* **115**(1) 1-42.

Peitz, M., 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.

Rob, R., J. Waldfogel. 2006. 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**(1) 29-62.

Zentner, A. 2006 Measuring the Effect of Music Downloads on Music Purchases. *Journal of Law and Economics* **49**(1) 63-90.

Zentner A. 2005. File Sharing and International Sales of Copyrighted Music: An Empirical Analysis with a Panel of Countries. *Topics in Economic Analysis & Policy* **5**(1). http://www.bepress.com/bejeap/topics/vol5/iss1/art21

2. Examining Genres

In principle, the use of data containing information on sales by musical genres would seem capable of providing important clues about the impact of file-sharing. If regressions across cities indicated, for example, that the Internet has differential impacts on record sales by genre in accordance with the degree of file-sharing in various genres, that would have an important additional independent test of the file-sharing hypothesis.

Different types of music lovers 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 other music such as Alternative and especially Rap would be expected to appeal to younger users (how many people over 40 have any appreciation for Rap?) more prone to engage in file-sharing.[1] Our expectations for country music are less clear. The measurement firm NPD has collected data on peer-to-peer downloads by genre which largely confirm these expectations, and which are presented below.

[1] Oberholzer-Gee and Strumpf (2007) report data that, if reflective of the entire market, would provide a measure of which categories are most heavily downloaded. Unfortunately their descriptive statistics on genres do not include the top selling albums which are usually classified as belong to 'Top Current.' Genre data that is missing the top selling albums in each genre are of very 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 remaining albums in their sample, by genre. The results: Alternative 100; R&B 77; Hard 47; Rap 29; Country 21; Jazz 4.

The use of genre data requires care because genre definitions can change and albums are often classified in more than one genre. Yearly fluctuations, therefore, may not properly represent changes in market conditions so much as changes in classifications. Two examples of this are found in Table 6, which lists album sales in seven genres reported by SoundScan. The extreme changes in the Hard Rock (Metal) category beginning in 1999 and the R&B category in 1997 were due to alterations in genre definitions.[2] For this reason the 1998 'hard rock' value is replaced with the 1999 value, making its result below slightly non-comparable with the others.

Table 6: 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
2004	135,317	19,098	82,041	76,887	19,156	165,364	81,558

Another difficulty with using genre regressions is related to the potential entertainment impact of the Internet. If young people are more likely to treat the Internet as entertainment, and if they also gravitate toward particular genres of music, then genre regressions may pick up nothing more than the entertainment substitution of the Internet and not file-sharing. The genres of music affected will be the same genres most heavily shared since young individuals have a disproportionate impact on file-sharing activities. We have already seen estimates that the Internet reduces entertainment activities, but by relatively small amounts. The question is whether young people might be much more heavily impacted than average users and thus reduce their listening and purchasing of particular genres of music.

[2] 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.

There are two sources of information that would indicate that this is unlikely to be the case. The first piece of information is the phenomenal success of MP3 players such as the iPod which are generally thought to be the provenance of the young. This would seem to indicate that young people are still heavily interested in music. This is somewhat imprecise, however, since young people might be listening to less music while still interested in hearing it on a portable device.

The second piece of information is somewhat more direct. Nielsen has national statistics on television viewing by age group and by year. According to Nielsen Media, time spent viewing television *increased* for all age groups during the period from 1995 to 2005 (an increase of about 13% in total) 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+).[3] At the same time, Internet penetration increased from 7% to 68%.[4] These data appear inconsistent with a claim that the Internet dissuades, in a disproportionate way, young people from using the main form of entertainment in the US.

Finally, the results below, for genres popular with young people are orders of magnitude higher than the Internet impact on radio usage measured in section V of the paper and thus seem unlikely to be entirely or even largely due to the entertainment impact.

We now repeat our econometric analysis for each genre of music for the period 1998-2003. Regressions measuring the impact of file-sharing on record sales 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).

---

[3] From page 16-17 in "Cable TV Facts 2001" and page 18-19 in "Cable TV Facts 2006" published by the Cabletelevision Advertising Bureau, New York, NY.

[4] According to http://www.nua.ie/surveys/how_many_online/n_america.html and http://www.internetworldstats.com/am/us.htm.

For the sake of brevity, only a summary of the results is shown in Table 7. The numbers found in the first three columns to the right of the genre names are normalized coefficients based the six regressions run for each genre. The normalization process consisted of three steps. First, the regression coefficients (for the Internet use variable) were multiplied by the 2003 Internet penetration rate to generate a predicted impact of file-sharing on sales per capita for that genre. Then they were multiplied by the summed DMA population in order to generate a total number of albums per genre impacted by Internet use. Finally, these numbers were further scaled by the summed DMA sales for that genre in 1999 (the last year largely untainted by file-sharing). The resulting numbers give the estimated impact of Internet use on record sales as a percentage of 1999 sales for each genre. For example, the number in Table 7 for the 'average' coefficient in the genre 'Country would imply, at the 2003 Internet penetration rate, that Internet use diminished album sales by 5.99% of the 1999 sales level. For comparison purposes, the actual changes in sales (per capita) by genre are shown in the last two columns.

Table 7: Estimated Internet Impact on Per Capita Sales					Download	Actual Sales	
	Average	Min	Max	Comments	Propensity*	98-03	00-03
Country	-5.99%	-26.08%	22.99%	3 +	0.81	-9.19%	2.71%
Classical	-15.14%	-34.47%	22.89%	1+	0.40	-0.40%	5.07%
Hard Rock	-19.27%	-31.43%	-7.90%		0.51	-13.76%	-20.01%
Jazz	-26.45%	-31.78%	-18.40%		0.87	18.41%	18.45%
Alternative	-39.88%	-50.59%	-20.46%	3@5%, 1@10%	0.86	5.18%	-5.13%
R&B	-48.54%	-54.06%	-44.82%	2 @ 5%, 3@10%	1.37	-14.86%	-26.93%
Rap	-50.51%	-60.53%	-39.25%	2 @ 5%,1@10%	2.84	-14.31%	-31.11%
x @ y% means there were x regressions where coefficient was significant at the y% level; n+ means n regressions had a positive coefficient; *NPD data based on May 2005 through April 2006							

Our main interest is in the relative size of the file-sharing coefficients for different genres. The average adjusted coefficients indicate a fairly large negative impact of file-sharing. It is not a surprise that Rap and R&B (many Rap albums are cross-listed with R&B) have the highest estimated impact from Internet use (file-sharing). The fourth column tells us how many of the six regression specifications provided statistically significant coefficients for each genres The perhaps surprising low estimate of file-sharing for hard rock value may be due to the fact that many hard rock cognoscenti are now beyond middle age and

less likely to engage in file-sharing, contrary to the general opinion held by many, including myself, that hard rock fans remain forever young. Or it may be due to the change in genre definition that occurred in 1999. Finally, the average coefficient for Country music implies that country music fans do walk the line and avoid engaging in much file-sharing, a not implausible finding. The Internet impact on Jazz is somewhat higher than expected

The measure of file-sharing intensity reported by NPD from its observation of file-sharing in its sample of users is largely but not completely consistent with the regression results.[5] The simple correlation between the NPD measure of file-sharing and the absolute value of the normalized coefficients in column 1 is -.73, which is significant at the 6% level, a fairly remarkable level of significance give that there are only seven observations. The correlation between the NPD measured genre impacts and the actual sales changes is -.37 for the 98-03 period used in the regressions and -.62 for the 00-03 period representing the recent decline in album sales and although neither is significant at the 10% level the measured negative relationships are nevertheless large, implying that actual sales changes by genre also were consistent with NPD data, although not as strong as our genre regression results.

The results from the genre regressions seem to strongly support a conclusion that file-sharing is harmful to record sales.

3. Regressions from periods 1998-2001 and 2000-2003

We have chosen the time period 1998-2003 as the most appropriate for our analysis. One may wonder what would happen if the time period were separated into subsamples to shorten the time period and increase the likelihood that fixed effects remain fixed since they have so little time to change. It also may help us to understand the changing impact of file-sharing during the entire interval, which I have

[5] The genre file-sharing information comes from a NPD MusicWatch Digital panel consisting of 12,500 U.S. households who have given NPD permission to observe their computer activity. Panelists agree to install a software application on their hard drives which monitors their digital music activities, including file-sharing. Data are processed, weighted/projected, and reported monthly.

suggested would grow stronger due to increased availability of CD burners, MP3 players, and broadband connections, although the latter failed to provide any extra impact over dial-up in the main regressions.

In order to gauge the best compare the sub-periods with the main period one adjustment to the regressions in Table 2 needs to be made. That table included the impact of music radio play. I do not have radio information for any years except 1998 and 2003 and thus must run the subsamples without it. Table 2X presents the results for regressions identical to Table 2, but without the music radio variable.

The average coefficient on Internet penetration drops (in absolute terms) to -2.02. Not surprisingly, the smaller coefficients have lower levels of statistical significance although the coefficients are similar to the "Low" estimates in Table 5 and provide the same overall conclusion about the impact of file-sharing.

Table 2X: 1998-2003 Differences in Record Sales per Capita (no radio)

	Naïve	Weighted by	OLS		Pop weighted	
	OLS	Cov&Pop	Cov>.6	Cov>.75	Cov>.6	Cov>.75
Internet Penetration	-1.386	-1.925	-1.933	-1.702	-2.378	-2.196
	(1.22)	(2.09)**	(1.62)	(1.46)	(2.28)**	(1.87)*
Income Change (000)	0.008	0.025	0.028	0.034	0.028	0.038
	(0.83)	(2.41)**	(1.89)*	(1.98)*	(2.47)**	(2.82)***
Population Change (00,000)	-0.001	-0.020	-0.022	-0.026	-0.022	-0.028
	(0.07)	(1.74)*	(1.43)	(1.77)*	(1.94)*	(2.66)**
Share of Yng Change	5.929	5.340	8.809	8.788	4.519	4.286
	(1.94)*	(1.64)	(2.14)**	(1.59)	(1.37)	(1.14)
Share of Males Change	0.007	-1.737	-4.285	-3.425	-2.491	-1.444
	0.00	(0.84)	(1.53)	(1.07)	(1.02)	(0.60)
Change in Share of 55+	1.799	1.740	6.111	3.266	1.794	-0.322
	(0.82)	(0.80)	(1.89)*	(0.79)	(0.64)	(0.14)
Change in College Grad	2.197	3.513	5.270	8.839	2.912	3.317
	(1.06)	(1.19)	(1.55)	(1.57)	(0.97)	(0.93)
Change in Hispanics	-0.417	0.097	-0.051	-0.230	1.015	0.302
	(0.23)	(0.06)	(0.03)	(0.14)	(0.53)	(0.18)
Change in Blacks	-3.237	-1.109	-0.041	-1.987	-0.091	-1.031
	(1.93)*	(0.81)	(0.02)	(1.07)	(0.06)	(0.63)
Constant	0.174	0.414	0.325	0.156	0.663	0.559
	(0.25)	(0.71)	(0.45)	(0.21)	(1.05)	(0.87)
Observations	92	92	61	40	61	40
R-squared	0.18	0.2	0.36	0.46	0.23	0.37

1 city removed; Robust t statistics in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%

I have chosen to split the 1998-2003 period into two subperiods: 1998-2001 and 2000-2003. By overlapping the years in this way this choice of periods makes the two periods equal in length. More importantly, since file-sharing had its first growth spurt in 2000 and became fully entrenched by 2001, it is important to include 2001 in the first period in order to capture some impact from file-sharing but the second period should begin in 2000 so that too much damage is not done to the assumption that the first year occurs before file-sharing exists, which is the centerpiece of my methodology. Note, however, that by including some file-sharing activity in the first year, the second term on the rhs of equation (3) will not be zero, causing our results to be biased. Our measure of file-sharing, the first term on the rhs of the equation, will always overstate the true size of file-sharing and will do so even more for cities with a larger share of Internet users in 2000. The first factor will tend to bias the Internet coefficient downward and the second will as well if across cities 20003 Internet use is positively correlated with 2000 Internet use, which it is. Thus the measured coefficients in the latter period are likely to understate the true coefficients.

Table 2Y: 1998-2001 Differences in Record Sales per Capita						
		Weighted by	OLS		Pop weighted	
	OLS	Cov&Pop	Cov>.6	Cov>.75	Cov>.6	Cov>.75
Internet Penetration	-0.253	-0.751	-1.449	-1.716	-0.999	-0.941
	(0.45)	(1.45)	(2.15)**	(2.13)**	(1.72)*	(1.38)
Income Change (000)	0.013	0.030	0.044	0.046	0.038	0.040
	(1.52)	(3.44)***	(3.87)***	(3.38)***	(3.56)***	(3.21)***
Population Change (00,000)	-0.014	-0.016	-0.036	-0.031	-0.016	-0.018
	(0.59)	(1.21)	(1.55)	(1.20)	(1.31)	(1.27)
Share of Yng Change	1.408	1.565	1.921	3.238	1.371	1.473
	(1.12)	(1.17)	(1.09)	(1.33)	(0.86)	(0.73)
Share of Males Change	1.683	1.588	2.016	1.882	1.837	1.126
	(1.19)	(1.04)	(1.08)	(0.77)	(1.03)	(0.50)
Change in Share of 55+	0.510	0.038	0.942	0.911	-0.126	-0.986
	(0.54)	(0.04)	(0.72)	(0.51)	(0.10)	(0.60)
Change in College Grad	-0.383	-1.782	-3.811	-4.093	-2.893	-2.708
	(0.34)	(1.38)	(2.13)**	(1.92)*	(1.73)*	(1.40)
Change in Hispanics	0.457	-0.438	-1.160	-2.119	-0.499	-1.031
	(0.39)	(0.40)	(0.92)	(1.42)	(0.39)	(0.68)
Change in Blacks	0.259	0.439	0.531	1.393	0.455	0.857
	(0.27)	(0.45)	(0.48)	(0.81)	(0.41)	(0.52)
Constant	0.153	0.427	0.885	1.127	0.556	0.549
	(0.45)	(1.32)	(2.15)**	(2.26)**	(1.49)	(1.26)
Observations	92	92	67	52	67	52
R-squared	0.08	0.16	0.26	0.31	0.23	0.26
1 city removed; Robust t statistics in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%						

The results from the first part of this period, 1998-2001 are show in Table 2Y. Although the average Internet coefficient, -1.17, is more than half of the full period value in Table 2X and is high enough to explain the entire period decline in record sales, the Stata robust (RREG) regressions and examination of influential observations indicate that these results are influenced by outliers and that the coefficients would be closer to -.45 if influential observations were removed.[6] Thus it seems that the impact of the Internet during this early period of time is probably best thought to be in the lower portion of the range of -.4 to -.9.

[6] Indeed, there was one outlier, Richmond, Virginia, whose removal reduces the coefficient in the OLS regression where coverage>.6 to -.23 and where coverage >.75 to -.39. There was no similar influential observation that would raise the coefficient.

On the other hand, the period from 2000-2003 shown in Table 2Z reveals an average Internet coefficient of -1.39 and this value largely holds up after examining the impact of unusually influential observations. This value, although not quite as large as the full sample, is considerably larger than required to explain the entire decline in record sales.

There is one year of overlap between these two periods. Nevertheless, the general story told by these sets of regressions is consistent between the subperiods and the full period. The first period coefficients should be smaller since file-sharing propensity among Internet users is smaller. The second half should have a high coefficient, as high as the coefficient for the entire period. Nevertheless, the coefficients are not far enough apart to reject the hypothesis that they are the same. Secondly, the coefficients in table 2Y are biased downward somewhat due to the fact that file-sharing had already begun. Thus the results are not surprising.

We conclude from this examination that file-sharing has a negative impact in both the early portion and the later portion of the full 5-year period 1998-2003. But the impact in the later period is about twice as strong as the impact in the early period, which is consistent with the expectation that both the quantity of file-sharing and its impact on sales was increasing during the period.

Table 2Z: 2000-2003 Differences in Record Sales per Capita						
	Naïve	Weighted by	OLS		Pop weighted	
	OLS	Cov&Pop	Cov>.6	Cov>.75	Cov>.6	Cov>.75
Internet Penetration	-0.466	-1.002	-0.829	-1.411	-1.600	-2.107
	(0.90)	(1.60)	(1.22)	(1.85)*	(1.91)*	(2.33)**
Income Change (000)	0.001	0.004	0.002	0.006	0.007	0.012
	(0.15)	(0.71)	(0.32)	(0.73)	(1.06)	(1.39)
Population Change (00,000)	-0.452	-0.386	-0.392	-0.276	-0.384	-0.413
	(3.33)***	(4.13)***	(2.74)***	(1.64)	(4.31)***	(4.58)***
Share of Yng Change	-0.074	-0.614	0.129	-1.073	-0.191	-0.216
	(0.07)	(0.48)	(0.11)	(0.54)	(0.13)	(0.09)
Share of Males Change	-0.428	-1.458	-3.580	-3.337	-2.599	-3.134
	(0.32)	(0.88)	(1.68)*	(1.57)	(1.11)	(1.19)
Change in Share of 55+	-0.995	-1.564	-1.256	-2.581	-1.639	-2.875
	(1.12)	(1.65)	(1.17)	(2.01)*	(1.32)	(2.25)**
Change in College Grad	0.439	0.698	2.866	2.043	0.647	-0.481
	(0.54)	(0.58)	(1.93)*	(1.19)	(0.35)	(0.22)
Change in Hispanics	1.257	1.656	2.653	3.384	1.046	1.822
	(1.18)	(1.40)	(2.07)**	(2.48)**	(0.68)	(1.21)
Change in Blacks	-0.591	-1.329	-0.949	-0.442	-2.191	-2.322
	(0.62)	(1.18)	(0.63)	(0.25)	(1.48)	(1.44)
Constant	-0.274	0.048	-0.153	0.238	0.392	0.716
	(0.81)	(0.13)	(0.36)	(0.51)	(0.80)	(1.38)
Observations	95	94	67	50	67	50
R-squared	0.11	0.21	0.24	0.28	0.29	0.37
1 city removed; Robust t statistics in parentheses; * significant at 10%; ** significant at 5%; *** significant at 1%						

4. Regressions with both dialup and broadband

Table 2BB provides regressions identical to those in Table 2 in the paper except that the Internet usage variable is decomposed into the dialup and broadband components. Comparing the two tables reveals that the Internet coefficient in Table 2 is always between the two coefficients in Table 2BB which are themselves quite similar to one another. It is somewhat surprising that dial-up has slightly higher coefficients than broadband but it is possible that dial-up users have different characteristics that are associated with more file-sharing, ceteris paribus, than broadband users, and the regression does not control for this.

Table 2BB: 1998-2003 Differences in Record Sales per Capita						
	Naïve	Weighted by	OLS		Pop weighted	
	OLS	Cov&Pop	Cov>.6	Cov>.75	Cov>.6	Cov>.75
Dialup	-0.433	-2.280	-2.268	-3.407	-2.989	-2.777
	(0.40)	(2.14)**	(1.68)	(2.76)**	(2.54)**	(2.24)**
Broadband	-0.654	-1.861	-2.149	-2.753	-2.246	-2.370
	(0.61)	(1.84)*	(1.55)	(2.25)**	(1.98)*	(1.89)*
Income Change (000)	0.005	0.020	0.025	0.029	0.026	0.035
	(0.61)	(2.22)**	(1.73)*	(1.77)*	(2.21)**	(2.37)**
Population Change (00,000)	-0.009	-0.012	-0.012	-0.002	-0.018	-0.022
	(0.48)	(0.80)	(0.65)	(0.06)	(1.51)	(1.56)
Change in Music Radio	-0.351	-0.814	-1.098	-1.775	-0.422	-0.497
	(0.42)	(1.42)	(1.52)	(2.15)**	(0.81)	(0.68)
Share of Yng Change	4.579	4.770	8.055	7.012	4.604	4.353
	(1.68)*	(1.58)	(2.06)**	(1.48)	(1.38)	(1.13)
Share of Males Change	-1.351	-2.819	-5.434	-5.429	-2.780	-1.840
	(0.62)	(1.20)	(1.83)*	(1.51)	(1.05)	(0.69)
Change in Share of 55+	1.481	1.376	5.242	1.392	1.564	-0.600
	(0.72)	(0.68)	(1.62)	(0.42)	(0.56)	(0.26)
Change in College Grads	2.669	3.234	4.832	7.069	2.634	3.132
	(1.31)	(1.14)	(1.52)	(1.48)	(0.84)	(0.84)
Change in Hispanics	0.831	-0.085	-0.193	-2.109	0.812	-0.218
	(0.48)	(0.05)	(0.09)	(0.98)	(0.40)	(0.11)
Change in Blacks	-2.970	-1.294	-0.653	-1.475	-0.314	-1.104
	(1.79)*	(0.94)	(0.41)	(0.79)	(0.20)	(0.63)
Constant	-0.434	0.330	0.218	0.600	0.760	0.701
	(0.61)	(0.55)	(0.29)	(0.99)	(1.14)	(1.09)
Observations	89	89	60	40	60	40
R-squared	0.18	0.24	0.4	0.53	0.25	0.38
Robust t statistics in parentheses; * significant at 10%; ** significant at 5%; *** significant at						

5. Individual Year Regressions on Record Sales

Table Z presents results from individual year regressions on the per capita sales of sound recordings for 1998 and Table Q for 2003. These regressions are run with the same demographic variables as in the main text. There are some clear patterns in these pure cross section results.

Table Z: 1998 Record Sales per Capita						
	Naïve	Weighted	OLS		Pop weighted	
	OLS	by	Cov>.6	Cov>.75	Cov>.6	Cov>.75
Internet	1.902	2.352	3.354	3.831	2.489	2.463
Penetration	(2.38)**	(2.22)**	(2.64)**	(2.63)**	(1.69)*	(1.30)
Income	0.007	0.007	-0.007	-0.019	0.006	0.007
(000)	(0.65)	(0.56)	(0.43)	(1.10)	(0.40)	(0.32)
Population	0.003	0.002	0.002	0.002	0.002	0.001
(00,000)	(1.87)*	(2.12)**	(1.49)	(1.87)*	(2.30)**	(1.98)*
Music	-0.151	-0.267	-0.242	-0.079	-0.452	-0.671
Radio	(0.59)	(0.94)	(0.73)	(0.17)	(1.25)	(1.20)
Share of	-2.136	-2.085	-2.351	0.948	-1.874	2.223
Yng	(1.60)	(1.47)	(1.17)	(0.48)	(0.99)	(0.95)
Share of	1.550	0.498	0.671	1.020	0.131	0.195
Males	(1.05)	(0.30)	(0.31)	(0.40)	(0.06)	(0.08)
Share of	-2.995	-2.731	-2.636	-1.991	-2.073	0.322
55+	(2.30)**	(2.09)**	(1.51)	(1.10)	(1.15)	(0.14)
College	2.541	2.571	2.822	4.195	2.872	3.343
Grads	(1.93)*	(1.77)*	(1.41)	(1.79)*	(1.46)	(1.47)
Hispanics	0.186	-0.006	0.105	-0.263	0.036	0.145
	(0.35)	(0.01)	(0.17)	(0.46)	(0.06)	(0.21)
Blacks	-0.183	-0.154	-0.336	0.089	0.208	0.744
	(0.41)	(0.31)	(0.48)	(0.11)	(0.25)	(0.68)
Constant	2.345	3.031	3.156	1.465	3.386	2.006
	(1.70)*	(1.97)*	(1.54)	(0.71)	(1.64)	(0.88)
Observations	91	91	60	41	60	41
R-squared	0.49	0.54	0.55	0.68	0.61	0.69

Cov=Coverage, Pop=population; t-statistics based on robust standard errors in parentheses (not for RREG); * significant @ 10%; ** significant @ 5%; *** significant @ 1%

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.

Income appears to have no clear impact on record sales. Income's impact may be picked up instead by the Internet variable and the college variable, which is positively related to record sales. Cities with large populations also have higher record sales per capita in 1998, although this is not found in 2003. Music

radio listening appears to have a negative impact although the result is much stronger in 2003 than in 1998.

Table Q: 2003 Record Sales per Capita						
	Naïve	Weighted	OLS		Pop weighted	
	OLS	by	Cov>.6	Cov>.75	Cov>.6	Cov>.75
Internet	3.362	2.103	2.925	1.846	1.721	0.891
Penetration	(4.59)***	(2.28)**	(2.54)**	(1.51)	(1.55)	(0.73)
Income	0.003	0.005	-0.002	0.004	0.001	0.007
(000)	(0.57)	(0.71)	(0.23)	(0.36)	(0.19)	(0.77)
Population	-0.001	-0.001	-0.001	-0.002	-0.001	-0.001
(00,000)	(0.60)	(0.95)	(1.07)	(1.70)*	(0.72)	(0.99)
Music	-0.731	-0.909	-0.910	-1.048	-0.951	-1.057
Radio	(2.69)***	(2.87)***	(2.25)**	(2.16)**	(2.72)***	(2.42)**
Share of	-4.924	-5.359	-6.057	-4.198	-5.544	-3.328
Yng	(3.25)***	(2.93)***	(3.96)***	(1.20)	(2.81)***	(0.87)
Share of	-4.963	-3.174	-6.517	-4.292	-4.295	-2.280
Males	(2.69)***	(1.20)	(1.96)*	(1.41)	(1.13)	(0.56)
Share of	-2.014	-2.876	-4.029	-3.400	-3.455	-2.225
55+	(1.68)*	(2.01)**	(2.94)***	(1.56)	(2.04)**	(0.88)
College	-0.126	1.507	1.489	2.356	2.327	2.648
Grads	(0.15)	(1.02)	(1.04)	(0.90)	(1.44)	(1.15)
Hispanics	1.757	1.406	1.485	1.315	1.248	1.106
	(4.00)***	(3.09)***	(2.80)***	(2.09)**	(2.78)***	(1.94)*
Blacks	1.402	1.341	1.202	2.115	0.916	1.392
	(2.33)**	(2.00)**	(1.23)	(1.93)*	(1.18)	(1.48)
Constant	5.846	6.154	8.094	6.796	7.272	5.768
	(3.04)***	(2.52)**	(3.04)***	(2.38)**	(2.46)**	(1.65)
Observations	92	92	66	47	66	47
R-squared	0.5	0.55	0.54	0.6	0.6	0.63
Cov=Coverage, Pop=population; t-statistics based on robust standard errors in parentheses (not for RREG); * significant @ 10%; ** significant @ 5%; *** significant @ 1%						

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 seems slightly negative in 2003. 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. The minority variables show no relationship to record sales in 1998 but it appears to be positive in 2003.

6. Individual Year Regressions on Television and Radio Usage

Table T provides the yearly results for television viewing. Internet use does not show up with any consistent result. The main demographic results are consistent with findings elsewhere in the literature. These would be that old individuals and minorities both watch more television. Also, viewership is greater in larger cities although the economic size of the coefficient is not large (an increase of ten million individuals in a city would lead to an approximate increase in viewing of only 30 minutes. Income appears unimportant. It also appears to be the case that cities with a larger share of males have greater television usage and although there is some indication of a negative relationship between share of population with college degrees and television use, it is not strong. Income seems to have no clear effect.

Table T: TV 1998 and 2003										
	----------1998----------					----------2003----------				
	Cov & POP	Cov > .60	Cov > .75	Cov > .60	Cov > .75	Cov & POP	Cov > .60	Cov > .75	Cov > .60	Cov > .75
Share Broadband						-1.694	-3.214	0.607	-2.666	0.346
						(1.12)	(1.54)	(0.28)	(1.44)	(0.15)
Share Dialup	0.656	0.695	-0.798	0.872	-0.659	-1.432	-1.845	0.837	-2.176	0.025
	(0.62)	(0.56)	(0.38)	(0.56)	(0.29)	(0.92)	(0.94)	(0.44)	(1.37)	(0.01)
Income (000)	-0.007	-0.003	0.011	-0.002	0.013	0.002	0.011	0.008	0.007	0.004
	(0.57)	(0.17)	(0.49)	(0.12)	(0.43)	(0.20)	(1.01)	(0.49)	(0.65)	(0.28)
Population (00,000)	0.005	0.007	0.006	0.006	0.006	0.0038	0.0049	0.0044	0.0044	0.0042
	(4.17)***	(3.23)***	(2.97)***	(2.71)***	(2.09)**	(3.66)***	(2.97)***	(2.48)**	(1.94)*	(1.82)*
55 and over	2.5187	2.5659	3.6950	2.1095	3.6833	2.819	5.538	5.678	8.113	7.358
	(1.25)	(1.15)	(1.12)	(0.95)	(1.19)	(0.96)	(1.91)*	(1.77)*	(3.29)***	(2.43)**
Share with College Degree	-2.720	-3.019	-3.314	-3.399	-3.566	-0.717	1.333	-2.678	2.125	-1.499
	(1.79)*	(1.54)	(1.37)	(1.40)	(1.13)	(0.34)	(0.55)	(0.80)	(1.03)	(0.52)
Share Males	3.640	2.737	2.362	2.839	2.570	4.920	6.545	4.880	8.284	7.576
	(1.58)	(1.09)	(0.62)	(1.00)	(0.66)	(1.16)	(1.80)*	(1.02)	(2.25)**	(1.68)*
Hispanic Share	0.996	1.187	0.926	1.194	0.978	0.829	1.016	1.624	0.909	1.390
	(1.93)*	(1.86)*	(1.25)	(1.75)*	(1.13)	(1.68)*	(1.59)	(2.24)**	(1.30)	(1.76)*
Black Share	3.090	3.201	3.024	3.287	3.052	3.453	3.582	4.983	4.151	5.109
	(5.75)***	(4.18)***	(2.43)**	(4.06)***	(2.55)**	(4.61)***	(3.75)***	(4.46)***	(4.91)***	(4.71)***
Share Young	0.220	0.190	1.856	-0.049	1.616	-2.014	0.329	-0.384	1.730	1.047
	(0.10)	(0.08)	(0.47)	(0.02)	(0.42)	(0.71)	(0.13)	(0.11)	(0.72)	(0.33)
Constant	2.108	2.303	1.813	2.418	1.741	2.608	0.057	-0.165	-1.767	-1.926
	(1.25)	(1.20)	(0.62)	(1.11)	(0.57)	(0.88)	(0.02)	(0.05)	(0.71)	(0.62)
Observations	95	62	42	62	42	95	73	58	73	58
R-squared	0.54	0.50	0.45	0.43	0.40	0.46	0.47	0.53	0.51	0.54

The yearly radio usage regressions share some of the same type of results. Large cities have audiences that listen to more hours of radio. Minorities appear to listen to more radio. Young people appear to listen to less radio. There is no apparent impact of Internet use.

Because city characteristics changes very slowly, unlike Internet usage, many of the demographic variables lose their impacts in the first difference regressions shown in the main text since the fixed effects will pick them up. New York and Los Angeles are large in both 1998 and 2003. San Francisco is rich in both years. And so forth.

Table R: Radio in 1998 and 2003										
	----------1998----------					----------2003----------				
	Cov & POP	Cov > .60	Cov > .75	Cov > .60	Cov > .75	Cov & POP	Cov > .60	Cov > .75	Cov > .60	Cov > .75
Share Broadband						-0.695	-0.349	-0.337	-0.439	-0.416
						(1.61)	(0.85)	(0.62)	(0.89)	(0.64)
Share Dialup	-0.001	0.086	-0.379	0.040	-0.581	0.090	0.057	0.256	-0.025	0.225
	0.00	(0.22)	(0.80)	(0.08)	(1.16)	(0.23)	(0.15)	(0.51)	(0.06)	(0.42)
Income (000)	0.005	0.002	0.006	0.004	0.004	0.000	0.000	0.001	0.001	0.001
	(1.33)	(0.32)	(0.98)	(0.59)	(0.68)	(0.14)	(0.18)	(0.32)	(0.25)	(0.31)
Population (00,000)	0.001	0.002	0.002	0.002	0.002	0.0023	0.0024	0.0025	0.0023	0.0025
	(3.55)***	(3.36)***	(3.16)***	(2.47)**	(3.48)***	(5.89)***	(5.14)***	(4.82)***	(3.84)***	(3.87)***
55 and over	0.6577	0.1081	1.0289	0.4977	1.5183	0.614	1.315	1.599	1.322	1.604
	(1.18)	(0.12)	(1.52)	(0.73)	(2.26)**	(0.86)	(1.99)*	(1.98)*	(1.97)*	(1.83)*
Share with College Degree	-0.689	-0.794	-0.663	-0.640	-0.410	-0.396	-0.386	-0.679	-0.347	-0.698
	(1.46)	(1.36)	(1.00)	(0.85)	(0.59)	(0.77)	(0.79)	(0.96)	(0.59)	(0.85)
Share Males	-0.069	-0.549	-0.328	-0.473	-0.126	1.363	1.680	1.618	1.882	1.959
	(0.09)	(0.70)	(0.38)	(0.53)	(0.15)	(1.28)	(1.74)*	(1.26)	(1.87)*	(1.50)
Hispanic Share	0.523	0.474	0.213	0.469	-0.131	0.388	0.440	0.442	0.386	0.382
	(2.41)**	(2.14)**	(0.64)	(2.23)**	(0.70)	(2.15)**	(2.38)**	(1.94)*	(2.06)**	(1.72)*
Black Share	0.441	0.338	0.213	0.392	0.061	0.570	0.590	0.557	0.590	0.553
	(2.36)**	(1.46)	(0.68)	(1.56)	(0.24)	(2.99)***	(2.97)***	(2.14)**	(2.61)**	(1.82)*
Share Young	-0.603	-1.151	0.868	-0.840	1.905	-1.357	-0.683	-0.602	-0.768	-0.724
	(0.75)	(1.06)	(0.82)	(1.10)	(2.25)**	(2.15)**	(1.13)	(0.72)	(1.17)	(0.79)
Constant	2.898	3.529	2.659	3.228	2.252	2.439	1.784	1.686	1.749	1.588
	(5.13)***	(4.39)***	(3.78)***	(4.77)***	(3.41)***	(3.08)***	(2.70)***	(1.94)*	(2.62)**	(1.84)*
Observations	92	60	41	60	41	92	71	57	71	57
R-squared	0.42	0.38	0.37	0.33	0.44	0.63	0.55	0.57	0.50	0.54

Several of these demographic variables, particularly share of minority population, are very influential in the individual year regressions, consistent with previous research (Trac Media Services, 2001). But in the first difference regressions the demographic factors appear fairly weak. Cities with increasing shares of males appear to have decreases in television viewing. Cities with increasing shares of blacks appear to experience increases in television viewing. Income growth and increases in college degrees appear to be positively related to increased television viewing. Increases in population cities seem to decrease television viewing.

7. Regressions including Pueblo Colorado

Below is Table 2 without removing the Pueblo/Colorado Springs observation.

Table 2' without removing colorado springs: 1998-2003 Differences in Record Sales per Capita						
	Naïve	Weighted by	OLS		Pop weighted	
	OLS	Cov&Pop	Cov>.6	Cov>.75	Cov>.6	Cov>.75
Internet Penetration		-2.680	-3.387	-4.446	-2.878	-2.885
		(2.37)**	(1.92)*	(2.00)*	(2.56)**	(2.18)**
Income Change (000)		0.022	0.028	0.036	0.027	0.036
		(2.31)**	(1.86)*	(1.84)*	(2.35)**	(2.51)**
Population Change (00,000)		-0.009	-0.006	0.004	-0.015	-0.020
		(0.58)	(0.33)	(0.16)	(1.32)	(1.48)
Change in Music Radio		-0.726	-0.740	-1.155	-0.367	-0.383
		(1.24)	(0.92)	(1.07)	(0.67)	(0.54)
Share of Yng Change		5.127	8.810	7.962	4.779	4.713
		(1.71)*	(2.23)**	(1.67)	(1.45)	(1.26)
Share of Males Change		-2.631	-4.794	-4.768	-3.143	-2.390
		(1.13)	(1.50)	(1.32)	(1.21)	(0.89)
Change in Share of 55+		1.183	4.915	1.042	1.463	-0.623
		(0.58)	(1.51)	(0.28)	(0.54)	(0.27)
Change in College Grads		3.985	5.860	9.118	3.485	4.256
		(1.41)	(1.78)*	(1.86)*	(1.15)	(1.18)
Change in Hispanics		0.673	1.159	-0.254	1.558	0.679
		(0.38)	(0.44)	(0.10)	(0.74)	(0.33)
Change in Blacks		-1.069	-0.315	-1.740	-0.056	-0.818
		(0.74)	(0.18)	(0.84)	(0.04)	(0.46)
Constant		0.627	0.910	1.370	0.829	0.820
		(0.92)	(0.88)	(1.05)	(1.23)	(1.11)
Observations		90	61	41	61	41
R-squared		0.21	0.33	0.37	0.22	0.3
Robust t statistics in parentheses						
* significant at 10%; ** significant at 5%; *** significant at 1%						

8. Construction of the Data Set

Nielsen SoundScan data:

Data on album sales for the top 100 Nielsen DMAs (out of 210) were purchased for the years 1998 through 2003. The top 100 DMAs included 83% (this is the same percentage as the overall coverage) of the total population. The data from one DMA (Paducah, Kentucky) could not be matched with any Census metro areas and was removed from the analysis. The SoundScan data for DMA record sales are primarily based on information from electronic scanners in retail outlets. These data include information on the sale of albums by city, by year and by genre. These albums include CDs sold over the Internet, where the zip code of the delivery is used to determine the DMA where the sale is recorded. 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). Although current SoundScan data now include digital downloads (based on the zip code of the credit card), digital downloads played no role in the analysis since they were a trivial component of the market even as late as 2003 (they were not even counted by SoundScan until the second half of 2003).

RIAA data on record sales indicate a larger decline than the SoundScan data in part because RIAA data are more inclusive by including outlets particularly hard hit in recent years—record clubs and direct sales—which are not included in Nielsen statistics. Album sales in the US through 2005 have fallen over 30% from their 1999 peak using RIAA data but only 19% using Nielsen SoundScan data.

Census Current Population Survey Data

There were supplementary surveys of Internet Usage in 98, 00, 01, 03 which were a continuation of surveys on compute usage that occurred in 94 and 97. These data are publicly available. The variables chosen (from a much larger set): Home Internet Use; Type of Internet connection, Age of respondent, Household Family Income, Name of metro area (MSA), Weighting Variable, Sex of Respondent, educational attainment, minority status. The Census MSAs were matched as closely as possible to the Neilsen DMAs based on the names of the cities and maps which indicated which counties were included in a DMA. DMAs often contain more counties than MSAs and often the sum of the MSAs did not make up a very large percentage of the DMA population.

[More detail on Census data follows this summary]

Arbitron Data:

I was given access to a computer hooked up to the Arbitron data base. My research assistant was able to print out data for each station contained in the Metro Area, as well as its format, average persons, average ranking, and weekly and daily Time Spent Listening. Beyond the station-specific information, there are aggregates of each of the aforementioned values for each Metro Area. Arbitron has a listing of the Arbitron Metro Areas and how they relate to Nielsen DMAs. The average (weighted) time spent listening

to radio and time spent listening to music radio was calculated for the average person per day in each DMA.

Television Data:

I purchased data from Nielsen Media that included population figures and measurements of People Using Television (PUT) for the 100 Nielsen DMAs that Nielsen SoundScan had provided. PUTs were measured as % people tuned to television for a typical hour. I multiplied these values by 24 hours to get the number of hours (measured using regular decimals) that the average person in the DMA was watching television, since this is the same basis upon which the Arbitron data were measured.

Combining the data:

The Census data are reported by MSA or PMSA (Metropolitan Statistical Area or Primary Metropolitan Statistical Area). Nielsen data are reported by DMA, or designated market area, in which every county in the US is assigned to a DMA based on the television stations that receive a preponderance of television viewing in that county. DMAs and MSAs areas are not identical and the definitions change over time (MSAs as population changes and DMAs as television stations and viewing habits change). There are 210 Nielsen DMAs which cover every television household in the US, but I used data only for the top 100 DMAs. The number of Census MSAs and PMSAs listed in the data number 241, and they do not cover every household in the US. Nielsen DMAs are larger than census MSAs or PMSAs and include a larger population.

Nielsen provides maps which show every county and the DMAs to which the counties are assigned. By using these maps it is possible to place census MSAs into Nielsen DMA (when they do not have the same name). This is how the two data sources were combined. In some cities, such as New York, 12 Census MSAs were combined to form one Nielsen DMA. For most DMAs, a single MSA with the same name was the closest correspondence. 212 of the 241 census MSAs were used.

The maps also list counties by Nielsen Metro ratings areas, which are supposed to generally conform to Census MSAs. In general, DMAs contain more counties than do MSAs. For example, the 12 MSAs in the

New York DMAs represent 22 counties, but there are 29 counties in the New York DMA. The counties not in the MSA are generally smaller than those in the MSA. In general, Census MSAs tend to represent about 83% of the population of the Nielsen DMAs. This ratio is higher for the larger cities than it is for the smaller cities with the unweighted average coverage ratio being 73% and a weighted average of 83%. These left out individuals (~30% of the population) are poorer, have fewer broadband connections, lower Internet penetration rates, and lower income as shown in the rightmost column of Table 1.

Since population is used in the denominator of the dependent variable, the population statistics used are from Nielsen, to match the sample upon which record sales are based. The other statistics, such as Internet use, broadband use, income, and so forth, come from the Census MSAs.

We could not find the MSAs to match one of the Nielsen DMAs (Paducah), reducing the data set to 99. The radio data had missing information for three of the DMAs and the music radio data has missing information for an additional DMA so that sometimes the number of observations is reduced to 95 or 96. There were a small number of cities with no sampled blacks and thus these were missing observations for the share of blacks.

9. Details on Internet Census (CPS) Variables

Census data:[7]

Number of observations in Dec 1998: 122,935
Number of observations in Aug 2000: 121,745
Number of observations in Sept 2001: 143,300
Number of observations in Oct 2003: 140,058

Census Variables Used:

1) Internet Use Variable: HESIU3 (1998-2000) and HESINT1 (2001-2003)
2) Type of Internet connection: HESINT2A(2003), not available in 1998, 2000
3) Age Variable PRTAGE
4) Household Family Income HUFAMINC
5) Census MSA variable GTMSA

[7] Number of observations is for all ages. This is slightly larger than the number reported by the Census which only includes individuals above age 3.

6) Weighting Variable PWSSWGT
7) Sex of Respondent PESEX
8) Race of Respondent PERACE
9) Hispanic indicator PRHSPNON
10) Highest level of School completed PEEDUCA

Getting Current Population Survey data (technically joint between the Bureau of the Census and the Bureau of Labor Statistics):
Website access: http://dataferrett.census.gov/ . Use the dataferret utility to find data sets under Current Population Survey and then pick "Internet and Computer". You then get to pick from the four years with surveys. The following list provides the names of the variable chosen from the data set and their definitions.

Census variables

1) HESIU3 (1998-2000) and HESINT1 (2001-2003), internet variables:
   - HESIU3 refers to the question "Does anyone in this household use the Internet from home?" and the variable label is "Internet home use, computer or WebTV owners, y/n." HESIU3 has three values: -1 (Not in universe), 1 (Yes), and 2 (No). So, all the responses -1 (Not in universe) were changed to represent the response No. This was recommended by Eric Newburger and Kurt Bauman who are specialists in the Education and Social Stratification Branch Population Division.
   - HESINT1 refers to the question "Does anyone in the household connect to the Internet from home?" and the variable label is "Internet home use, y/n"
2) HESINT2A: 2003 Type of Internet connection. Choices are:
   - Not in universe; Regular dial-up telephone line; Mobile phone, PDS, or pager connection; A DSL Line (Digital Subscriber Line); A Cable modem; Satellite; A fixed wireless connection such a MMDS (Multi-Media Distribution System); Something else
   - We lumped all types of non-dialup together into broadband. Only Cable and DSL were of any consequence.
3) HUFAMINC, the household family in come variable:
   - The variable label is "Household-total family income in past 12 months."
   - The census variable has values from -3 to 14 or -3 to 16 (in 2003). Positive numbers indicated income ranges. Negative values indicated missing values: -3 means "refused", -2 means "don't know" and -1 is blank.
   - We created approximate income values for MSAs by creating the product of the number of individuals in a category and the mean of the income category and averaging over all individuals. These means of the categories were different in 2003 because the range of income categories changes.
4) PRTAGE, the age variable:
   - The variable label is "Demographics-age top coded at 90 years old." In fact, it was often top coded at 80.

- We are interested in two age groups: Old (those over 55) and Young (12-29 years). Since Nielsen TV data is for age 2 and up a category (baby) based on 0 and 1 year of age was created.
- All the observations in the data sets were recoded to fit the appropriate age group.

5) GTMSA, the census cities variable:
   - The variable label is "Geography-MSA code"
6) PESEX: Sex Indicator [1=Male; 2=Female]
   - Two groups—Male and Female recoded to 0=Female.
7) PRHSPNON, Hispanic Indicator
   - 'Demographics-hispanic/non-hispanic origin'
   - Coded: 1 'Hispanic'; 2 'Non-Hispanic'. Recoded Non-Hispanic to 0.
8) PERACE or PTDTRACE (2003): Race Variable
   - 'Demographics-race of respondent'
   - 1998 Coded: 1 'White'; 2 'Black'; 3 'American Indian, Aleut, Eskimo'; 4 'Asian or Pacific Islander'.
   - Recoded to 1 Black; 0 Everyone else
   - 2003 Coded: 1 'White Only'; 2 'Black Only'; 3 'American Indian, Alaskan Native Only'; 4 'Asian Only'; 5 'Hawaiian/Pacific Islander Only'; 6 'White-Black'; 7 'White-AI'; 8 'White-Asian'; 9 'White-Hawaiian'; 10 'Black-AI'; 11 'Black-Asian'; 12 'Black-HP'; 13 'AI-Asian'; 14 'Asian-HP'; 15 'W-B-AI'; 16 'W-B-A'; 17 'W-AI-A'; 18 'W-A-HP'; 19 'W-B-AI-A'; 20 '2 or 3 Races'; 21 '4 or 5 Races'
   - Recoded: 1 to Black Only, everyone else to 0 (other part black categories are very small).
9) PEEDUCA
   - 'Demographics-highest level of school completed'
   - Coded: -1 'Not in Universe' ; 31 'Less Than 1st Grade' ; 32 '1st,2nd,3rd Or 4th Grade' ; 33 '5th Or 6th Grade'; 34 '7th Or 8th Grade'; 35 '9th Grade'; 36 '10th Grade'; 37 '11th Grade'; 38 '12th Grade No Diploma'; 39 'High School Grad-Diploma Or Equiv (GED)'; 40 'Some College But No Degree'; 41 'Associate Degree-Occupational/Vocationl'; 42 'Associate Deg.-Academic Program'; 43 'Bachelor"s Degree(ex:ba,ab,bs)'; 44 'Master"s Degree(ex:MA,MS,MEng,MEd,MSW)'; 45 'Professional School Deg(ex:MD,DDS,DVM)'; 46 'Doctorate Degree(ex:PhD,EdD)'
   - Recoded 43, 44, 45, 46=1. Everyone else 0.
10) PWSSWGT, weighting variable:
   - The variable label is "Weight-second stage weight (rake 6 final step weight)."