

# *Topics in Economic Analysis & Policy*

---

*Volume 6, Issue 1*

2006

*Article 18*

---

## The Impact of Digital File Sharing on the Music Industry: An Empirical Analysis

Norbert J. Michel\*

\*Nicholls State University, [norbert.michel@nicholls.edu](mailto:norbert.michel@nicholls.edu)

Copyright ©2006 The Berkeley Electronic Press. All rights reserved.

# The Impact of Digital File Sharing on the Music Industry: An Empirical Analysis\*

Norbert J. Michel

## Abstract

The first file-sharing software, Napster, was shut down in 2001, but the copying technology's impact on the music industry is still passionately debated. This paper uses micro-level data from the Consumer Expenditure Survey to examine the impact of Internet file sharing on music sales. Music industry representatives argue that the practice decreases CD sales, while supporters of file-sharing allege the practice could actually increase sales. Using household-level data from the Consumer Expenditure Survey, we find support for the claim that file-sharing has decreased sales.

**KEYWORDS:** file sharing, copyright, music industry

---

\*This paper is based on Chapter 6 of my Ph.D. dissertation (University of New Orleans, Dept. of Economics and Finance). I thank my co-chairs, Arja Turunen-Red and Oscar Varela, as well as Gerald Whitney, Stan Liebowitz, and my committee members for their helpful suggestions and comments. I also thank Tracy Foertsch and several anonymous referees. Any errors are solely my responsibility. Norbert J. Michel, Nicholls State University, 310 B Powell, Thibodaux, LA. 70310, norbert.michel@nicholls.edu , 985-448-4223.

## INTRODUCTION

The impact of new copying technology on the music industry has been hotly debated since the launch of the first file-sharing software, Napster, in 1999. Music industry representatives have charged that indiscriminate copying decreases compact disc (CD) sales, while supporters of free file sharing have alleged the practice is mostly innocuous.<sup>1</sup> Although several researchers have examined the impact of copying on CD purchases, the present paper is the first to use the micro-level Consumer Expenditure Survey data to study the possible impact of file-sharing. The findings in this paper are broadly consistent with recent research.

Zentner (2005) finds that cross-country aggregate data support a 14 to 23 percent reduction in CD sales (in the U.S.) from file sharing. Hong (2004) uses a two-sample instrumental variable approach and finds that file sharing significantly decreased music purchases for households with children aged 6 to 17. Liebowitz (2004 – a ) examines the impact of file sharing using aggregate U.S. data and estimates that file sharing may have decreased album sales by as much as 30 percent. Oberholzer and Strumpf (2004) match U.S. record sales data to file-sharing data from a peer-to-peer (P2P) network over a seventeen week period in 2002 and find that the impact of downloads on album sales is statistically indistinguishable from zero.<sup>2</sup> Rob and Waldfogel (2004) use micro-level data on college students' downloads and music purchases, and find that downloads do displace music purchases (by varying degrees). Zentner (2003) finds that file sharing usage reduces the probability of purchasing music by 30 percent. In the present study, using micro-level expenditure data from the Bureau of Labor Statistics' (BLS) Consumer Expenditure Survey (CEX), we find that some U.S. music consumers could have decreased their CD purchases (prior to 2004) by about 13 percent due to Internet file sharing. The remainder of this paper is structured as follows. Section I provides an overview of the data and methodology, Section II discusses summary statistics, Section III presents our regression analysis, and Section IV concludes.

---

<sup>1</sup> On September 17, 2002, the President of the Consumer Electronics Association (CEA) gave a speech in which he said that file sharing is both "legal and moral." A copy of the speech can be found on the CEA's website at [http://www.ce.org/press\\_room/speech.doc](http://www.ce.org/press_room/speech.doc). However, given the recent U.S. Supreme Court decision in *MGM Studios v. Grokster*, the legality of file sharing does not appear to be an open question. Nothing in this paper should be construed to suggest that the author endorses or condones illegal file sharing.

<sup>2</sup> As explained in Oberholzer and Strumpf, the bulk of file sharing (during this period) took place on P2P networks.

## **I. DATA AND METHODOLOGICAL APPROACH**

### **A. DATA**

The CEX is used by the BLS to compute the Consumer Price Index and data is collected in both *interview* and *diary* formats. The diary method (consumers keep a log) primarily collects detailed information on consumer staples, such as food. We use the CEX interview files, where up to 95 percent of total household expenditures are reported, for the years 1995 through 2003. The CEX interview survey uses a rotating panel design, wherein consumers report their purchases from the three-month period prior to their interview month. Interviews are conducted each month but individual consumer units report their purchases no more than four times per twelve-month period. The interview files contain quarterly expenditures for fifteen overlapping quarters.

For example, the first set of interviews in the 2003 CEX interview survey is conducted in January 2003 and captures expenditures from October to December of 2002. The second set of interviews is conducted in February 2003 and captures expenditures from November 2002 to January 2003, and so on. The last set of interviews (the fifteenth quarter) is conducted in March 2004 and captures expenditures from December 2003 to February 2004. In order to derive estimates that are comparable to the music industry's annual sales figures, we derive weighted calendar year estimates of annual mean music expenditures by following the procedure provided in the CEX documentation. This method adjusts the cross-sectional quarterly expenditures to account for the different calendar months reported in each interview.<sup>3</sup> All regression analyses are performed on the cross-sections of quarterly expenditure data.

The unit of observation in the CEX is a "consumer unit," a measure analogous to a household. Each consumer unit's expenditures and demographic information are reported only for those individuals living at the same address. Children at boarding school or living in a college dorm, for example, are not included in their parents' consumer unit (they would comprise their own consumer unit). In this paper, we interchangeably use the terms *consumers* and *households* to refer to consumer units. For all statistics calculated with CEX data, only households coded as "complete income reporters" are used, a designation that indicates information was provided for at least one of the household's major sources of income, such as wages and salaries. Additionally, we drop from our analysis all consumers with missing income or expenditure information, as well as those with missing information on computer ownership. Unless otherwise noted, all statistics reported below employ the appropriate weights (see the Appendix for

---

<sup>3</sup> For more information on this procedure, see page 273 of "2004 Consumer Expenditure Interview Survey Public Use Microdata Documentation," available at <http://www.bls.gov/cex/2004/cex/csxintvw.pdf>.

further details).<sup>4</sup> All nominal dollar amounts are converted to 2003 dollars using the CPI-U for all items less food and energy.

## **B. METHODOLOGICAL APPROACH**

One advantage of using the CEX to study file-sharing activity is that self-selection bias in the data is all but eliminated. For instance, there is little chance that consumers will misreport their music purchases for self-serving reasons related to Internet downloading because the CEX does not ask any questions related to file sharing. Consequently, we are only able to use the CEX to infer what impact file sharing may have had on CD purchases.<sup>5</sup> A major weakness of using the CEX data is that we are unable to directly test the same individuals' expenditure changes across years. However, using the CEX does allow us to directly test for the effects that specific demographic characteristics may have on nationally representative cross sections of consumers' expenditures. We exploit this feature of the data to identify clusters of possible file-sharing activity.

Specifically, we examine the relationship between computer ownership and household expenditures on music. To study this relationship, regressions are run on cross sections from 1995 through 2003, respectively, as well as on pooled data. Because a computer is a necessary tool for file sharing, the data should not (in general) show any significant change in this relationship if file sharing had nothing to do with declines in music sales. If, on the other hand, computer-owning consumers increased their use of file-sharing services and increased (decreased) their purchases of CDs, we would expect to see a positive (negative) change in the relationship between computer ownership and CD purchases.

One possible problem with this approach is that computer owners who did not engage in file sharing may have decreased their music purchases for reasons unrelated to file sharing (yet inherent to owning a computer). Because the CEX does not directly record consumers' file-sharing activity we are unable to more fully control for this possibility. Another possible weakness with this methodology is that broadband usage could be better than computer ownership as a predictor of file-sharing activity. Unfortunately, the CEX does not record broadband access (or even Internet access) until 2001, two years after the first file-sharing service went online. Even though the data quality for broadband use is poor, the regression results using a "high speed Internet access" indicator

---

<sup>4</sup> The CEX weights are designed to generate statistics representative of U.S. consumers, but these weights are generally targeted to hit broader aggregates than the expenditures used in this paper. In fact, it is well-documented that using a fine level of detail for expenditure categories with the CEX frequently results in aggregates that appear smaller than other aggregate measures (see Garner, Janini, Passero, Paszkiewicz, and Vendemia (2003)). It is not surprising, therefore, that the weighted sum of music purchases in the CEX fluctuates between \$3.5 billion and \$3.8 billion during the sample period, representing about 30 percent of aggregate U.S. CD sales.

<sup>5</sup> For a definition of the CEX variable used to represent "CD purchases," see the Appendix.

variable are presented below.<sup>6</sup> Summary statistics for our samples are presented in the next section.

## **II. SUMMARY STATISTICS, 1995 THROUGH 2003**

All of the following summary statistics are presented only for consumers that meet the following criteria: non-missing CD expenditures, non-missing income, complete income reporters and non-missing computer ownership information. Table 1 presents summary statistics for the following consumer groups: (1) *both* computer-owning and non-computer-owning consumers; (2) consumers that own a computer at the time of the survey; and (3) consumers that do not own a computer at the time of the survey. Table 1 provides calendar-year mean CD expenditures as well as year-to-year changes (in absolute and percentage change form) for each consumer group.

Table 1 indicates there is a positive relationship between computer ownership and CD expenditures, and the relationship appears stronger in the beginning of the sample period. Still, the mean CD expenditure for both computer-owners and non-owners follows a generally downward trend from 1995 through 2003. Additionally, the *number* of non-owners is about 50 percent lower in 2003 vs. 1995 even though the overall sample size is increasing. This pattern is consistent with increased U.S. household computer ownership in 2003 vs. 1995.<sup>7</sup> Even when we examine CD expenditure trends around 1999, the year of the Napster launch, we find similar expenditure *patterns* among both computer owners and non-owners.

Both groups' mean expenditure declined from 1995 to 1997, increased in 1998 and 1999, and then started a generally downward trend through 2003. These trends match those of U.S. aggregate CD sales in five of the eight years measured (see Table 2). While most of the computer owning group's year-to-year changes are larger in absolute terms than the non-owning group's changes, the same cannot be said using year-to-year percentage changes. Nonetheless, a careful comparison of these groups' mean expenditure changes suggests that a closer look at the data is warranted.

The year Napster went online (1999), for example, computer owners' mean CD expenditures increased \$0.68 (1 percent), a small, statistically insignificant change. On the other hand, non-computer owners' mean CD expenditure increased \$3.30 (20 percent) in 1999, a statistically significant increase at the 5 percent level of significance. In 2000, computer owners' mean CD expenditure decreased \$4.77, a statistically significant 10 percent decrease. In

---

<sup>6</sup> Of course, Internet *access* and Internet *usage* are not the same, but usage is not recorded in the CEX.

<sup>7</sup> See Day, Janus, & Davis (2005).

the same year, non-computer owners' mean CD expenditure decreased by 8 percent, and the change is statistically insignificant.

In 2001, the year that Napster was shut down, computer owners' mean CD expenditure increased by only \$0.05 (less than 1 percent) and the change is statistically insignificant. In the same year, non-computer owners' mean CD expenditure declined 19 percent (significant at the 5 percent level), a decrease that does not bolster the hypothesis that file sharing caused a decrease in music sales. Yet, in the following two years, the relative changes for these two groups is consistent with file sharing leading to a decrease in music sales. In 2002 and 2003, years in which Napster-like services continued to grow in popularity, computer owners' mean CD expenditures decreased by \$4.79 and \$5.55 (11 and 14 percent), respectively. These changes are statistically significant at the 5 percent level. During the same two years, the mean CD expenditure of the non-computer owning group decreased by only \$0.80 and \$0.22 (5 and 2 percent), respectively, and neither decrease was statistically significant.

Table 3 provides additional summary statistics for the three consumer groups, and shows that computer owners tend to be a bit older, have slightly larger families, and have more than 1.5 times the annual wage and salary income (in most years) as the non-computer owners. Income also follows different trends for the computer owners and non-owners. While income for non-computer owning consumers is almost \$9,000 less in 2003 than in 1995, computer owners' annual income is approximately \$3,500 greater in 2003 vs. 1995. Furthermore, computer owners' income is \$934 higher in 2003 than in 1998 (the year prior to the Napster launch), suggesting income changes may not explain the decline in these consumers' CD purchases. While these summary measures only provide a superficial picture of what is happening in the data, they do warrant a closer look at the relationship between computer ownership and music expenditures.

### **III. REGRESSION ANALYSIS**

#### **A. DIFFERENCE-IN-DIFFERENCES MODEL**

We begin our regression analysis by directly measuring the *change* in the relationship between computer ownership and household expenditures on music from 1998 to 2003. We conduct this test with the difference-in-differences estimator, designed to compare control and treatment groups before and after a particular event (see Wooldridge (2003)). The event is the initiation of the first file-sharing service (Napster) in 1999. The estimator compares the difference in expenditures of a control group (consumers that do not own a computer) and a treatment group (consumers that do own a computer) before and after the event. The test is run on data from the years 1998 *and* 2003, so the estimator represents

the difference in expenditures that could be attributed to file-sharing activity.<sup>8</sup> If file sharing between 1999 and 2003 was largely performed by computer owners who previously purchased significant amounts of music, then we would expect the estimator to be negative and statistically significant. If, on the other hand, computer-owning consumers increased their use of file-sharing services and increased their purchases of CDs, we would expect to see a positive change in the relationship between computer ownership and CD purchases.

It is also reasonable to expect that any significant impact on CD sales from Internet file sharing (in the near term) should be concentrated among avid music consumers. To test this implication using the difference-in-differences estimator, we split the sample into “below average” and “above average” music consumers, where the “average” consumer is defined as one who purchases the median level of CDs. In our full sample, the inflation adjusted median level of household CD purchases is \$33 in 1998 and \$30 in 2003. Our “below average” sub-sample consists of all consumers who spend less than or equal to \$32 on CDs.<sup>9</sup>

The tests are performed by estimating the following equation:

$$\ln(CDexp) = \hat{\beta}_1 + \hat{\beta}_2 \ln(income) + \hat{\beta}_3 I(computer = 1) + \hat{\beta}_4 X' + \hat{\beta}_5 I(Year2003 = 1) + \dots + \hat{\beta}_6 (computer * Year2003). \quad (1)$$

In model (1), the natural logarithm of each consumer’s real CD expenditures serves as the dependent variable ( $\ln(CDexp)$ ). The independent variables of interest are the indicator variables for computer ownership ( $I(computer = 1)$ ) and the year 2003 ( $I(Year\ 2003 = 1)$ ), and the interaction variable ( $computer * Year\ 2003$ ). The  $I(Year\ 2003 = 1)$  variable is a year indicator, set to one for expenditures made in the year 2003, and the  $I(computer = 1)$  variable is a year indicator, set to one for all consumers who own a computer. The coefficient on ( $computer * Year\ 2003$ ) is the difference-in-differences estimator, whereby the year indicator  $I(Year\ 2003 = 1)$  is multiplied by each consumer’s computer indicator ( $I(computer = 1)$ ).

To control for preferences and seasonal variation, the model contains a vector ( $X$ ) of control variables including family size and age, as well as time indicators for the expenditure quarters and years. The CEX family size variable is reclassified so that six or more people represent the largest family size, and the

<sup>8</sup> The CEX data cannot be used to measure annual expenditure changes for the *same* consumers over time, so an underlying assumption is that computer owners in 1998 were not systematically different from computer owners in 2003. This issue is addressed further below.

<sup>9</sup> Defining below average music consumers as those with CD purchases in the first quartile (rather than up to the median) produces results similar to those presented below.



age variable is grouped into four categories: under 31, between 31 and 55, between 56 and 65, and over 65 (the age of the spouses are averaged for married households). Because the CEX definition of before tax income was changed in 2001, the natural logarithm of wage and salary income is used as the independent variable controlling for income ( $\ln(\text{income})$ ).<sup>10</sup> An alternative specification, using the age of the household's children, yields nearly identical results to those presented here (see Michel (2003), Appendix I), and including additional preference variables such as race and region of residence do not materially impact the results presented below. All regressions are run only on consumers that meet the following criteria: non-missing CD expenditures, non-missing income, complete income reporters and consumers with non-missing computer ownership data.

### **B. DIFFERENCE-IN-DIFFERENCES RESULTS**

Table 4 presents the results from running model (1) on pooled data from 1998 and 2003 for the full sample as well as for "below average" and "above average" music consumers. The table is divided into three sets of three columns. For each set, the first column provides the estimated coefficients and standard errors from running model (1). The second and third columns provide the absolute and percentage changes, respectively, for each estimate that can be measured separately in 1998 and 2003.<sup>11</sup>

Table 4 shows that, using the full sample, the coefficient of the difference-in-differences estimator ( $\text{computer} * \text{Year } 2003$ ) is negative and statistically significant at the 5 percent level of significance. In other words, the relationship between computer ownership and CD expenditures weakened from 1998 to 2003. Because the model regresses the natural logarithm of CD expenditures, this estimate (-0.13) suggests that file sharing could have decreased CD sales by about 13 percent between 1998 and 2003. The regression results also confirm the positive relationship between income and CD expenditures, with a statistically significant estimate of 0.11 on the natural log of income. Given that computer owners' income is actually a bit higher in 2003 than 1998, it does not appear that income changes can explain computer owners' decrease in mean CD expenditures. Table 4 also presents the estimates from running model (1) on only the "below average" and "above average" music consumers.

As expected, the difference-in-differences estimator is smaller in magnitude (-0.01) and statistically insignificant when using the below average sample. Running model (1) on only the "above average" music consumers

---

<sup>10</sup> Income is collected on an annual basis in the CEX.

<sup>11</sup> Absolute and percentage changes for the full sample are calculated from coefficient estimates (presented later) shown in Table 5. Changes for the below and above average music consumers are taken from coefficient estimates shown in the Appendix.

produces a difference-in-differences coefficient that is negative (-0.11) and statistically significant at the five percent level. This estimate of the (*computer \* Year 2003*) coefficient is statistically indistinguishable from the estimate on the full sample. These results suggest that any negative impact on CD sales from file sharing was concentrated among the avid music consumers. Nonetheless, these results require further scrutiny because of possible econometric problems.

The main complicating factor is that computer owners in 1998 and 2003 are not necessarily two homogenous groups. Indeed, Table 1 demonstrates that the raw number of computer-owning consumers in our sample increased from 2,670 in 1998 to 4,306 in 2003, while the number of non-owners decreased from 1,737 in 1998 to 1,008 in 2003. Furthermore, the difference-in-differences estimator in (1) could be biased downward if computer owners in 2003 purchased computers specifically to download music. The proliferation of CD burning, for example, highlights the possibility that computers were purchased specifically to copy music CDs. On the other hand, because consumers use the same technology to convert CDs to the MPEG format, it is possible that this activity could contribute to a stronger relationship between computer ownership and CD purchases. File sharing proponents have also suggested that music sales decreased as consumers substituted into other types of entertainment goods, such as prerecorded movies and video games. However, Michel (2005) concludes there is no evidence that *CEX households'* decrease in music purchases corresponds to an increase in spending on movie tickets, prerecorded movies or video games.

Another possibility is that computer owners' mean CD expenditures have decreased through time as they purchased legal downloads instead of CDs. Fans of independent label music, for example, have been able to download music through the Internet Underground Music Archive (IUMA) website since 1993. By 1998, IUMA claimed to be indirectly responsible for gross sales of approximately \$1 million per year (See Krasilovsky & Shemel, 2000, p. 449). Still, given that the market for digital downloads is still developing (Apple launched its iTunes service for Macintosh users on April 28, 2003), it is likely that this issue will be much more important using more recently released versions of the CEX than are used in this paper. Regardless, as of the 2003 release, the CEX does not directly capture purchases of digitally downloaded music.<sup>12</sup> To summarize, there are at least four possible issues which could mitigate the findings presented in this paper, and the most serious issue appears to be the fact that computer owners in 1998 and 2003 are not necessarily two homogenous groups.

Ideally, panel data (with the same computer-owning consumers included for all four years) would be used to control for this problem, but such a panel

---

<sup>12</sup> In fact, Nielsen SoundScan, the company that tracks U.S. music purchases, did not begin tracking digital downloads until the second half of 2003.

cannot be constructed with CEX data. Alternatively, two-stage least squares (2SLS) could be used to control for possible endogeneity, but finding suitable instruments within the CEX has proved most difficult.<sup>13</sup> Because neither option is workable, we further analyze the relationship between computer ownership and CD expenditures using data that predates the existence of online file sharing. If, for example, a negative change in the relationship between CD purchases and computer ownership exists in the data prior to the initiation of the first file-sharing service, our interpretation of the results in Table 4 would be weakened.

### C. REGRESSIONS ON ANNUAL AND POOLED DATA

In this section of the paper, we analyze the relationship between computer ownership and CD expenditures in the following two ways: on a year-by-year basis and with pooled data from 1995 to 2003. To perform these tests, we simply modify our main regression model. For the regressions using the respective annual data, we use the following model:

$$\ln(CDexp) = \hat{\beta}_1 + \hat{\beta}_2 \ln(income) + \hat{\beta}_3 I(computer = 1) + \hat{\beta}_4 X' . \quad (2)$$

As in model (1), the natural logarithm of each consumer's real CD expenditures serves as the dependent variable ( $\ln(CDexp)$ ). The key independent variable is ( $I(computer = 1)$ ), an indicator set to one for consumers owning a computer. Model (2) also includes the same independent variable for income ( $\ln(income)$ ) and the same vector ( $X$ ) of control variables as the original model. For the regression using the pooled data from 1995 to 2003, we simply modify model (2) by adding a full set of year indicators (with 2003 as the reference year) and interacting the computer indicator with the respective year indicators (for a total of eight interaction variables).

Table 5 shows that the coefficient on the computer indicator variable is positive (with a point estimate of 0.15) and statistically significant (at the five percent level of significance) in 1995, four years before any online file sharing services existed. The parameter estimate then decreases to 0.12 in 1996, but increases to 0.15 and 0.16 in 1997 and 1998, respectively (and remains statistically significant). In 1999, the year that Napster went online, the coefficient estimate is still statistically significant but it decreases to 0.11. The estimate then decreases to 0.03 in 2000 and is no longer statistically significant. In 2001, the year that Napster was shut down, the point estimate on the computer indicator increases to 0.12 and is statistically significant at the five percent level.

---

<sup>13</sup> Earlier versions of this paper included 2SLS with admittedly poor instruments, as well as a synthetic cohort analysis which suffered from small sample problems. These test results, though less dependable, did nothing to change the conclusions drawn from the tests presented in the paper.

While this increase does coincide with Napster being shut down, at least one independent survey suggests that Internet file sharing actually became more widespread between 2000 and 2002, a period during which Napster replacements proliferated.<sup>14</sup> Table 5 also demonstrates that in both 2002 and 2003 the parameter estimates for the computer indicator are close to zero and statistically insignificant. There is some evidence to suggest that Internet file sharing continued to grow in popularity during these two years, but several estimates suggest a slowdown in activity during 2003. This possible slowdown in file-sharing activity has been linked to the increase in lawsuits filed (late in 2003) by the recording industry against individuals engaged in file sharing.<sup>15</sup>

Table 6 presents the regression results from using pooled data on the entire sample period (1995 through 2003). These results are consistent with the previously discussed findings. The coefficient on the natural log of income, for example, is 0.10 and is statistically significant at the five percent level. The coefficient on the computer indicator variable is 0.02 and statistically insignificant. The estimates on the interaction terms (the respective year indicators multiplied by the computer indicator) are stronger in the pre-file-sharing period than in the post-file-sharing period. Overall, at the very least, our test results suggest that the relationship between computer ownership and music purchases (in the CEX) weakened after Internet file sharing became a viable option for music purchasers, and this change is concentrated among the heaviest music purchasers. We find no evidence (in the CEX) that file sharing led to a widespread increase in music purchases. If such a change were present in the data, we would expect to have seen a positive change in the relationship between computer ownership and CD purchases.

#### **D. A NOTE ON BROADBAND INTERNET ACCESS IN THE CEX**

The main tests in this paper use the relationship between computer ownership and music purchases to identify possible clusters of Internet file sharing. It stands to reason that exploiting the connection between broadband Internet use and music purchases would also be a good approach. Unfortunately, the CEX did not collect information on Internet access until 2001, two years after the first file-sharing service went online. And, between 2001 and 2003, the CEX did not collect spending or usage information on broadband, only whether the household's utility bills include such spending.

There is also a serious discrepancy in the CEX broadband variables. For example, 849 households (in 2003) responded that they had paid for *broadband* access, yet 759 of these households responded that they had not paid for *Internet*

---

<sup>14</sup> See Rainie, Madden, Hess and Mudd (2004).

<sup>15</sup> Unfortunately, the estimates of file-sharing activity are so varied that even broad generalizations of trends may be in error. For a thorough discussion, see Liebowitz (2004 – b).

*access* (of any kind).<sup>16</sup> For the sake of completeness, we construct an indicator variable – a far from perfect indicator variable – for broadband access by using the responses to both of these questions. For all households who reported they had not paid for Internet access, the “high speed” indicator was set to zero. For the remaining households who reported they had paid for broadband Internet access, the indicator was set to one. This indicator variable was then added to regression model (2), and the results are reported in Table 7. The estimated effect for the high speed indicator is a statistically significant -0.24 for 2001, but is insignificant for the other two years, as well as in the pooled sample. Given these test results and the poor data quality for this variable, it is probably best to refrain from drawing any conclusions about the relationship between music purchases and broadband Internet access in the CEX. Regardless, it is impossible to measure the change in this relationship (in the CEX) before and after the advent of Internet file sharing.

#### IV. CONCLUSIONS

Even though Napster was shut down in 2001, the impact of digital file sharing on the music industry is still passionately debated. Opponents of file sharing charge that indiscriminate Internet copying decreases music sales, while supporters of free file sharing argue the practice is harmless, at worst, and may even increase music sales. This paper uses micro-level data from the Consumer Expenditure Survey to examine the impact of Internet file-sharing on music sales, and its findings are broadly consistent with recent research.

Our micro-level data test results suggest that file sharing may have reduced album sales (between 1999 and 2003) by as much as 13 percent for some music consumers. At minimum, our test results indicate that the relationship between computer ownership and music purchases (in the CEX) weakened after Internet file sharing became a viable option for music purchasers. No similar negative change exists in the data prior to the initiation of the first file-sharing service. Furthermore, we present evidence that this weakened relationship is concentrated among the heaviest music purchasers, and we find no evidence that file sharing led to a widespread increase in music purchases. Nonetheless, our results should be used carefully when predicting the *long-term* viability of the music industry in an environment where record labels (or artists) compete directly with free file-sharing services. In the current market, most music consumers purchase their music in the CD format, and digital downloads are not as highly substitutable for CDs as for other digital downloads.

---

<sup>16</sup> The CEX respondents are first asked whether their bills include charges for Internet access, and later asked if their bills include charges for DSL or ISDN access.

## **APPENDIX**

The CEX sample methodology is known as a stratified random sample (or a complex sample), whereby the U.S. population is divided into strata and then random samples are drawn from each strata. To control for survey design characteristics, we use the CEX-supplied full-sample and replicate weights. The full-sample weights are used in weighted least squares regressions to correct for heteroscedasticity and to ensure that estimates are representative of the U.S. population. The replicate weights are employed to ensure that estimators are unbiased and test statistics are valid because variances within strata tend to be more homogenous than those found in a simple random sample. According to Wooldridge (2002, p. 596), when stratified samples are partitioned based on exogenous variables, standard non-weighted estimators on the stratified sample are consistent and asymptotically normal. However, the BLS does not release detailed strata information, and Brogan (1998) and Landis, Lepkowski, Eklund, and Stehouwer (1982) have shown that ignoring the weighting and sample design schemes of complex survey data can lead to biased and inefficient estimators, as well as invalid statistical inferences (see also Sharon L. Lohr (1999)). Non-weighted results (available from the author) do not materially change the results presented in this paper.

For the summary statistics on Table 1, because both positive and negative mean expenditure changes are present, and since a standard software package cannot be used to compute the significance tests, two-tailed tests are constructed for all changes rather than individual one-tailed tests. The CEX variable used for music purchases includes expenditures on “compact discs, tapes, needles, and records not from a club.” The CEX designation “complete income reporters” indicates information was provided for at least one of the household’s major sources of income. The results presented in Table A-1 are used to calculate the changes in the coefficients (from 1998 to 2003) shown in Table 4 for the below average and above average samples, respectively.

## **REFERENCES**

Brogan, D. (1998) “Pitfalls of Using Standard Statistical Software Packages for Sample Survey Data” In: Armitage and Colton, eds., *Encyclopedia of Biostatistics*, New York: John Wiley & Sons.

Day, Jennifer Cheeseman, Alex Janus and Jessica Davis (2005) “Computer and Internet Use in the United States: 2003” U.S. Census Bureau Special Study P23-208, October, <http://www.census.gov/prod/2005pubs/p23-208.pdf>.

Garner, Thesia, George Janini, William Passero, Laura Paszkiewicz, and Mark Vendemia (2003) "The Consumer Expenditure Survey in Comparison: Focus on Personal Consumption Expenditures" Prepared for the Federal Economic Statistics Advisory Committee Meeting, March 17, 2003.

Hong, S. (2004) "The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster" Stanford University Working Paper, November.

Krasilovsky, W.M. & S. Shemel (with contributions by John M. Gross) (2000) This Business of Music: the Definitive Guide to the Music Industry (8th edition), New York: Billboard Books.

Landis, R., J. Lepkowski, S. Eklund and S. Stehouwer (1982), "A Statistical Methodology for Analyzing Data From A Complex Survey: The First National Health and Nutrition Examination Survey" U.S. Department of Health and Human Services Pub. No. 82-1366, Vital and Health Statistics. Series 2, No. 92.

Liebowitz, S. J. (2004a) "Peer-to-peer networks: Creative Destruction or just Plain Destruction?" University of Texas at Dallas, School of Management, Working Paper, July.

Liebowitz, S. J. (2004b) "File-Sharing: Creative Destruction or just Plain Destruction?" University of Texas at Dallas, School of Management, Working Paper, December. <http://som.utdallas.edu/capri/destruction.pdf>

Lohr, S. L. (1999) Sampling: Design and Analysis, Pacific Grove, CA: Duxbury Press.

Michel, N. (2003) "A Theoretical and Empirical Analysis of the Impact of the Digital Age on the Music Industry" Ph.D. Dissertation, University of New Orleans, <http://www.uno.edu/theses/available/etd-11212003-100808/>.

Michel, N. (2005) "Digital File Sharing and the Music Industry: Was There a Substitution Effect?" Review of Economic Research on Copyright Issues, Vol. 2(2), pp. 41-52.

Oberholzer, F., and K. Strumpf (2004) "The Effect of File Sharing on Record Sales; An Empirical Analysis" Working Paper, March.

Rainie, Lee, Mary Madden, Dan Hess and Graham Mudd (2004) "RE: The state of music downloading and file-sharing online" Pew Internet Project and

Comscore Media Metrix Data Memo, April,  
[http://www.pewinternet.org/pdfs/PIP\\_Filesharing\\_April\\_04.pdf](http://www.pewinternet.org/pdfs/PIP_Filesharing_April_04.pdf).

Rob, Rafael and Joel Waldfogel (2006) "Piracy On The High C's: Music Downloading, Sales Displacement, And Social Welfare In A Sample Of College Students" Forthcoming, *Journal of Law and Economics*.

Wooldridge, J. (2003) *Introductory Econometrics; A Modern Approach* (2<sup>nd</sup> ed.), US: Thomson-South-Western.

Wooldridge, J. (2002) *Econometric Analysis of Cross Section and Panel Data*, MA: MIT Press.

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 (2003), "Measuring the Effect of Online Music Piracy on Music Sales. University of Chicago, Department of Economics, Working Paper, June.



**Table 1, Mean CD Expenditures (2003 \$)**

	1995	1996	1997	1998	1999	2000	2001	2002	2003
<i>Panel A - All Consumers</i>									
Mean CD Expenditure:	\$35.32	\$26.96	\$20.11	\$29.34	\$32.59	\$31.08	\$31.10	\$29.10	\$26.75
Difference (year-to-year):	-	(8.36)**	(6.86)**	9.23**	3.25**	(1.51)	0.03	(2.00)	(2.35)*
% change:	-	(0.24)	(0.25)	0.46	0.11	(0.05)	0.00	(0.06)	(0.08)
No. Observations:	3,753	3,920	4,194	4,407	5,392	5,349	5,283	5,501	5,314
<i>Panel B - Computer Owners</i>									
Mean CD Expenditure:	\$65.52	\$48.18	\$34.98	\$48.84	\$49.52	\$44.76	\$44.80	\$40.01	\$34.46
Difference (year-to-year):	-	(17.34)**	(13.20)**	(13.86)**	0.68	(4.77)**	0.05	(4.79)**	(5.55)**
% change:	-	(0.26)	(0.27)	0.40	0.01	(0.10)	0.00	(0.11)	(0.14)
No. Observations:	1,717	2,024	2,414	2,670	3,474	3,780	4,084	4,418	4,306
<i>Panel C - Non-Computer Owners</i>									
Mean CD Expenditure:	\$24.94	\$17.76	\$12.42	\$16.91	\$20.21	\$18.63	\$15.15	\$14.35	\$14.13
Difference (year-to-year):	-	(7.19)**	(5.34)**	(4.50)**	(3.30)**	(1.58)	(3.48)**	(0.80)	(0.22)
% change:	-	(0.29)	(0.30)	0.36	0.20	(0.08)	(0.19)	(0.05)	(0.02)
No. Observations:	2,036	1,896	1,780	1,737	1,918	1,569	1,199	1,083	1,008

Annual mean CD expenditures and confidence intervals are calculated using the procedures defined in the Consumer Expenditure Survey documentation. Confidence intervals are estimated at both the 90 and 95 percent confidence levels to test whether the year-to-year changes in mean CD expenditures are statistically significant. The \* and \*\* symbols denote statistically significant changes at the 10 and 5 percent levels of significance, respectively. The \*\* superscript on the 1996 difference, for example, denotes a statistically significant change (at the 5 percent level) in annual expenditures from 1995 to 1996. Parentheses denote negative differences and percentage changes. All statistics use the required full-sample and replicate weights supplied with the Consumer Expenditure Survey data.

Annual mean CD expenditures are presented for consumers that report owning at least one computer (Panel B), consumers that report they do not own a computer (Panel C), and for both groups (Panel A). Statistics for all three groups include only those respondents with non-missing CD expenditures and non-missing income, as well as only those respondents coded as "complete income" reporters. The "complete income" code is used by the CEX to indicate information was provided for at least one of the consumer's major sources of income. All expenditures are in 2003 dollars.

**Table 2, Music Industry Sales (2003 \$)**  
(all figures in millions)

Year:	1995	1996	1997	1998	1999	2000	2001	2002	2003
CDS	\$11,322	\$11,651	\$11,367	\$12,720	\$13,989	\$14,082	\$13,401	\$12,215	\$11,233
<i>net units shipped</i>	723	779	753	847	939	943	882	803	746
Total Retail Units	\$14,875	\$14,699	\$14,028	\$15,277	\$15,920	\$15,264	\$14,265	\$12,793	\$11,854
<i>net units shipped</i>	1,113	1,137	1,063	1,124	1,161	1,079	969	860	798

Source: Recording Industry Association of America

**Table 3, Consumer Demographics**

	1995	1996	1997	1998	1999	2000	2001	2002	2003
<i>Panel A - All Consumers</i>									
Annual Income:	\$54,080	\$55,149	\$56,793	\$58,551	\$60,327	\$59,846	\$60,929	\$61,947	\$62,995
Difference (year-to-year):	-	\$1,068	\$1,644	\$1,758	\$1,777	-\$481	\$1,082	\$1,018	\$1,048
% change:	-	0.02	0.03	0.03	0.03	-0.01	0.02	0.02	0.02
Age:	39.31	39.07	39.61	39.54	39.70	39.60	40.36	41.08	41.34
Family Size:	2.77	2.75	2.79	2.78	2.84	2.85	2.81	2.80	2.76
No. Observations:	3,753	3,920	4,194	4,407	5,392	5,349	5,283	5,501	5,314
<i>Panel B - Computer Owning Consumers</i>									
Annual Income:	\$65,986	\$66,235	\$67,611	\$68,554	\$70,770	\$68,571	\$67,666	\$67,957	\$69,488
Difference (year-to-year):	-	\$250	\$1,376	\$943	\$2,216	-\$2,199	-\$904	\$290	\$1,532
% change:	-	0.00	0.02	0.01	0.03	-0.03	-0.01	0.00	0.02
Age:	40.43	40.13	40.44	40.08	40.70	40.37	41.15	41.48	41.67
Family Size:	3.00	2.95	2.90	2.89	2.96	2.96	2.89	2.90	2.85
No. Observations:	1,717	2,024	2,414	2,670	3,474	3,780	4,084	4,418	4,306
<i>Panel C - Non Computer Owning Consumers</i>									
Annual Income:	\$44,515	\$43,268	\$42,330	\$43,466	\$42,133	\$39,715	\$38,846	\$38,015	\$35,625
Difference (year-to-year):	-	-\$1,247	-\$938	\$1,136	-\$1,333	-\$2,418	-\$870	-\$830	-\$2,390
% change:	-	-0.03	-0.02	0.03	-0.03	-0.06	-0.02	-0.02	-0.06
Age:	38.42	37.93	38.50	38.71	37.94	37.82	37.77	39.46	39.96
Family Size:	2.59	2.54	2.64	2.61	2.64	2.59	2.55	2.40	2.39
No. Observations:	2,036	1,896	1,780	1,737	1,918	1,569	1,199	1,083	1,008

Table 3 presents weighted mean income, age of reference person, and family size for consumers that report owning at least one computer (Panel B), consumers that report they do not own a computer (Panel C), and for both groups (Panel A). Statistics for all three groups include only those respondents with non-missing CD expenditures and non-missing income, as well as only those respondents coded as "complete income" reporters. The "complete income" code is used by the CEX to indicate information was provided for at least one of the consumer's major sources of income. The Age variable is provided for the "reference person." The BLS defines the reference person as "the first member mentioned by the respondent when asked to 'Start with the name of the person or one of the persons who owns or rents the home.'" All income figures are in 2003 dollars.

Table 4, Difference in Differences, Natural Log of Real CD Expenditures (2003\$)

<i>Independent Variable</i>	Full Sample	Change (raw) from 1998 to 2003	Percentage Change	Average Consumers	Change (raw) from 1998 to 2003	Percentage Change	Average Consumers	Change (raw) from 1998 to 2003	Percentage Change
Intercept	-191.71** (48.63)	57.21	0.27	-40.83 (40.33)	-53.48	-4.26	-52.81 (56.27)	-27.37	-0.69
ln(income)	0.11** (0.01)	0.03	0.33	0.02** (0.01)	0.03	7.34	0.04** (0.01)	0.01	0.17
Family Size	0.00 (0.01)	0.04	2.00	0.01 (0.01)	0.02	7.54	-0.01 (0.01)	0.01	0.57
Age	-0.04* (0.02)	-0.01	-0.33	-0.02* (0.01)	0.00	-0.10	0.00 (0.02)	-0.03	-1.74
I(computer=1)	0.15** (0.03)	-0.17	-1.06	0.05** (0.02)	-0.04	-0.59	0.08** (0.03)	-0.11	-1.41
I(Year 2003=1)	-0.52** (0.12)	-	-	-0.07 (0.10)	-	-	-0.09 (0.14)	-	-
computer*Year 2003	-0.13** (0.05)	-	-	-0.01 (0.03)	-	-	-0.11** (0.03)	-	-
No. of Observations:	9,721			5,040			4,681		
R <sup>2</sup> :	0.06			0.05			0.01		

Table 4 presents WLS estimates of model (1) using data from 1998 and 2003, and employs the CEX full-sample and replicate weights (output for time dummies is suppressed for space considerations). The results are divided into three sets of three columns. For each set, the first column provides the estimated coefficients and standard errors (in parentheses) from running model (1); the second and third columns provide the absolute and percentage changes, respectively, for each estimate that can be measured separately in 1998 and 2003. The first set of three columns presents results on the full sample. The second and third sets provide results from running the model on the "below average" music consumers and the "above average" music consumers, respectively. The coefficient of interest is the difference-in-differences estimator (computer\*Year2003), an interaction of the computer-ownership indicator variable and the time-indicator for the year 2003. See the text for a description of the remaining independent variables in model (1). For all regressions, the dependent variable is the natural log of real CD expenditures (2003\$). The Appendix presents regression results from running the model separately on the below and above average samples using data from 1998 and 2003, respectively.

\* Statistically significant at the 10 percent level, \*\* Statistically significant at the 5 percent level

**Table 5, Regressions on Individual Years of Data, (2003\$)**

<i>Independent Variable</i>	1995	1996	1997	1998	1999	2000	2001	2002	2003
Intercept	-6.35 (4.90)	-55.55 (74.43)	-17.9 (89.14)	-215.82** (90.15)	-132.5** (66.81)	-91.44 (70.71)	-50.29 (89.48)	-10.41 (84.08)	-158.61** (78.60)
ln(income)	0.09** (0.02)	0.06** (0.02)	0.11** (0.02)	0.09** (0.01)	0.09** (0.02)	0.1** (0.02)	0.11** (0.01)	0.10** (0.01)	0.12** (0.01)
Family Size	-0.04** (0.01)	-0.01 (0.01)	-0.04** (0.01)	-0.02* (0.01)	-0.01 (0.01)	0.00 (0.01)	0.00 (0.01)	0.02* (0.01)	0.02** (0.01)
Age	-0.07** (0.03)	-0.02 (0.03)	-0.02 (0.03)	-0.03 (0.03)	-0.04 (0.03)	-0.06** (0.02)	-0.07** (0.02)	-0.08** (0.02)	-0.04** (0.02)
I(computer=1)	0.15** (.03)	0.12** (.04)	0.15** (.04)	0.16** (0.03)	0.11** (0.03)	0.03 (0.03)	0.10** (0.03)	0.02 (0.03)	-0.01 (0.04)
No. of Observations:	3,753	3,920	4,194	4,407	5,392	5,349	5,283	5,501	5,314
R <sup>2</sup> :	0.05	0.04	0.06	0.07	0.05	0.05	0.06	0.05	0.06

Table 5 presents WLS estimates of model (2) using data from 1995 through 2003, respectively, and employs the CEX full-sample and replicate weights (output for time dummies is suppressed for space considerations). In each column, for the respective independent variables, the first row presents the estimated coefficient and the second row the standard error. The coefficient of interest is the indicator variable I(computer=1), an indicator for whether the household owns at least one computer. For all regressions, the dependent variable is the natural log of real CD expenditures (2003\$).

\* Statistically significant at the 10 percent level

\*\* Statistically significant at the 5 percent level

**Table 6, Pooled Regressions (2003\$)**

<i>Independent Variable</i>		<i>Independent Variable</i>	
Intercept	2.18** (0.08)	I(Year 2001=1)	-0.02 (0.05)
ln(income)	0.10** (0.01)	I(Year 2002=1)	0.01 (0.05)
Family Size	-0.01** (0.00)	computer*Year 1995	0.12** (0.05)
Age	-0.05** (0.01)	computer*Year 1996	0.08 (0.06)
I(computer=1)	0.02 (0.04)	computer*Year 1997	0.13** (0.06)
I(Year 1995=1)	0.06 (0.05)	computer*Year 1998	0.13** (0.05)
I(Year 1996=1)	0.04 (0.05)	computer*Year 1999	0.09* (0.05)
I(Year 1997=1)	0.02 (0.05)	computer*Year 2000	0.01 (0.05)
I(Year 1998=1)	0.04 (0.04)	computer*Year 2001	0.09 (0.05)
I(Year 1999=1)	0.03 (0.04)	computer*Year 2002	0.01 (0.05)
I(Year 2000=1)	0.07 (0.05)		
No. of Observations: 43,113			
R <sup>2</sup> : 0.06			

Table 6 presents WLS estimates of a modified version of model (2) on six years of data (1995 through 2003), and employs the CEX full-sample and replicate weights. In each column, for the respective independent variables, the first row presents the estimated coefficient and the second row the standard error. Model (2) is modified by adding indicator variables for the first eight years of data (YEAR = 1995 through Year = 2002) and by including interaction variables, whereby each year indicator is multiplied by the computer indicator variable. For all regressions, the dependent variable is the natural log of real CD expenditures (2003\$).

\* Statistically significant at the 10 percent level

\*\* Statistically significant at the 5 percent level

<http://www.bepress.com/bejeap/topics/vol6/iss1/art18>

Michel: The Impact of Digital File Sharing on the Music Industry

**Table 7, Broadband as Additional Control Variable**

<i>Independent Variable</i>	2001	2002	2003	2001 through 2003
Intercept	2.28** (0.20)	2.21** (0.15)	1.82** (0.14)	2.07** (0.11)
ln(income)	0.11** (0.02)	0.10** (0.01)	0.13** (0.01)	0.11** (0.01)
Family Size	0.00 (0.01)	0.02* (0.01)	0.02** (0.01)	0.02** (0.01)
Age	-0.09** (0.03)	-0.08** (0.02)	-0.04* (0.02)	-0.07** (0.02)
I(computer=1)	0.10** (0.04)	0.02 (0.04)	0.01 (0.04)	0.03 (0.04)
I(High Speed Net=1)	-0.24** (0.11)	-0.28 (0.23)	0.24 (0.19)	-0.06 (0.12)
Year = 2001	- - -	- - -	- - -	0.06 (0.05)
Year = 2002	- - -	- - -	- - -	0.02 (0.05)
computer*Year 2001	- - -	- - -	- - -	0.06 (0.06)
computer*Year 2002	- - -	- - -	- - -	-0.02 (0.05)
No. of Observations:	3,998	4,986	4,839	13,823
R <sup>2</sup> :	0.03	0.05	0.06	0.05

Table 7 presents WLS estimates of a modified version of model (2) on three respective years of data as well as on pooled data (2001 through 2003), and employs the CEX full-sample and replicate weights (output for time dummies is suppressed for space considerations). In each column, for the respective independent variables, the first row presents the estimated coefficient and the second row the standard error. Model (2) is modified by adding an indicator variable for whether the household reports paying for high speed Internet access (High Speed Net=1). Indicators for years 2001 and 2002, as well as interaction variables, are included when running the model on pooled data. See the main text for data quality concerns involving CEX households' high speed access. For all regressions, the dependent variable is the natural log of real CD expenditures (2003\$).

\* Statistically significant at the 10 percent level

\*\* Statistically significant at the 5 percent level

**Table A-1, Additional Regressions on Split Sample**

<i>Independent Variable</i>	1998		2003		Change (raw) from 1998 to 2003		Change (raw) from 1998 to 2003		Percentage Change	
	Average	Below	Average	Above	Average	Above	Average	Above	Average	Above
Intercept	-12.54 (65.63)	-66.02 (57.35)	-53.48	-4.26	-39.80 (81.93)	-67.18 (56.85)	-27.37	-0.69		
ln(income)	0.00 (0.01)	0.04** (0.01)	0.03	7.34	0.03*	0.04** (0.01)	0.01	0.17		
Family Size	0.00 (0.01)	0.02* (0.01)	0.02	7.54	-0.01 (0.01)	-0.01 (0.01)	0.01	0.57		
Age	-0.02 (0.02)	-0.02 (0.02)	0.00	-0.10	0.02	-0.01 (0.01)	-0.03	-1.74		
I(computer=1)	0.07** 0.02	0.03 0.03	-0.04	-0.59	0.08** 0.03	-0.03 0.03	-0.11	-1.41		
No. of Observations:	2,119	2,921	2,288	2,393						

The results in Table A-1 are used to calculate the raw and percentage changes for the model coefficients presented in Table 4. The results in this table are from WLS estimates of model (2) using the "below average" and "above average" samples in 1998 and 2003, respectively. The model employs the CEX full-sample and replicate weights (output for time dummies is suppressed for space considerations). In each column, for the respective independent variables, the first row presents the estimated coefficient and the second row the standard error. For all regressions, the dependent variable is the natural log of real CD expenditures (2003\$).

\* Statistically significant at the 10 percent level

\*\* Statistically significant at the 5 percent level