

## DIGITAL FILE SHARING AND THE MUSIC INDUSTRY: WAS THERE A SUBSTITUTION EFFECT?

NORBERT J. MICHEL

ABSTRACT. Several empirical studies exist that measure the impact of file-sharing services on music sales, and most suggest that there was a negative impact on sales. Still, most of these studies do not examine (at the household level) whether consumers substituted out of music and into movies. This paper uses micro-level data from the Consumer Expenditure Survey (1998 through 2003) to test for this possible substitution effect. The data do not support the hypothesis that music consumers spent less on music because they spent more on either movie tickets or prerecorded movies (purchases or rentals).

### 1. INTRODUCTION

Since the first Internet file-sharing service, Napster, was launched in 1999, industry and consumer advocates have disagreed on the impact of digital file sharing on music sales. Music industry representatives blame the services for their recent sales declines, while users of the services claim virtually no economic loss has occurred. After Napster was shut down in 2001, other services quickly took its place and the debate intensified. The music industry decided to fight illegal file sharing through the courts, eventually deciding to initiate lawsuits against the most egregious copyright violators. Similarly, the Motion Picture Association of America (MPAA), has initiated its own legal campaign out of fear that movies will become as popular as music on the file-sharing services. The legal battles appear to be going in favor of the content industry – on June 27, 2005, the United States Supreme Court ruled that making file-sharing software available so that copyrights can be infringed is not a legitimate business.<sup>1</sup>

Despite these legal battles, file-sharing services continue to exist. Defenders of Internet file sharing argue the services can benefit music sales, and claim that much of the decline in music sales can be explained by other factors. For instance, file-sharing advocates argue that the services can increase artists' popularity, and that declines in music sales can be explained by increased consumer spending on substitute activities, such as movie viewing. Liebowitz (2004) examines aggregate sales of music, videogames, movie box office sales, and prerecorded movie sales, and argues that the aggregate data fail to support a substitution effect.

Michel (2005) and Hong (2004) both use the Consumer Expenditure Survey household-level data (CEX) and find that file sharing brought about a decrease

---

This paper is an extension of Chapter 6 of my Ph.D. dissertation (University of New Orleans, Dept. of Economics and Finance; see Michel, 2003). I thank my co-chairs, Arja-Turunen Red and Oscar Varela, as well as Gerald Whitney, Stan Liebowitz, an anonymous referee and my committee members for their helpful suggestions and comments.

<sup>1</sup>This was the ruling in "Metro-Goldwyn-Mayer Studios v. Grokster."

in music sales, but neither tests for a possible substitution effect.<sup>2</sup> The present paper uses household-level data from the CEX to directly test whether consumers simultaneously spent less on music and more on movie tickets and/or prerecorded movies. The CEX data fail to support the idea that music consumers spent less on music because they spent more on either movie tickets or prerecorded movies (purchases or rentals).<sup>3</sup> The remainder of this paper is structured as follows: Section 2 briefly describes the data and methodology, Section 3 presents average annual expenditures and summary statistics, Section 4 presents the regression model and test results, and Section 5 concludes.

## 2. DATA AND METHODOLOGY

The Consumer Expenditure Survey public-use micro files (CEX) are used to create six separate calendar year samples for 1998 through 2003. To provide comparable statistics, nominal dollar amounts are converted to 2003 dollars using the CPI-U for all items less food and energy. Weighted mean annual expenditures are provided for compact disc (CD), movie ticket, and movie purchase and rental expenditures (referred to as “prerecorded movies”).<sup>4</sup> These weighted statistics are provided for the following three groups of consumers with non-missing expenditures: all consumers, computer-owning consumers, and non-computer-owning consumers. For the main regressions, data from 1998 and 2003 are pooled.

Using the micro-level CEX data does not allow us 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 annual cross sections of consumers’ expenditures. This feature of the data is exploited to identify clusters of possible file-sharing activity. In particular, regressions are run to test for significant changes in the relationship between computer ownership and household expenditures on music, movie tickets, and prerecorded movies, respectively, with computer ownership serving as a proxy for possible file-sharing activity.<sup>5</sup>

Because a computer is a necessary tool for file sharing, if file sharing had nothing to do with declines in music sales, the data should not show any significant change in this relationship.<sup>6</sup> If, on the other hand, computer-owning consumers increased their use of file-sharing services and increased (decreased) their purchases

---

<sup>2</sup>For a thorough discussion of the empirical literature on file sharing and music sales, see Liebowitz (2004).

<sup>3</sup>The same findings hold when video game expenditures are tested, though the sample sizes are much smaller. For space considerations, the results are not presented here (see Appendix for further details).

<sup>4</sup>Summary statistics taken from the CEX are representative of U.S. consumers when the appropriate weights are applied (see Appendix).

<sup>5</sup>Internet usage is likely to be a better predictor of file-sharing than computer ownership, but the CEX does not record Internet usage. Starting in 2001, the CEX does include a unique variable for expenditures on Internet access. However, data collection for this variable is sparse, which could explain why the relationship between computer ownership and spending is stronger than the relationship between Internet access and spending (see Michel, 2003, Appendix J). Of course, Internet access and Internet usage are not the same, which also could explain this weak relationship.

<sup>6</sup>One possible exception is that computer owners who did not engage in file sharing decreased their music purchases for reasons unrelated to file sharing. Unfortunately, the CEX does not directly record consumers’ file-sharing activity, so I am unable to more fully control for this possibility.

of CDs, we would expect to see a positive (negative) change in the relationship between computer ownership and CD purchases. Likewise, conducting the same tests on movie ticket and prerecorded movie expenditures provides a check for possible substitution away from music purchases into alternative entertainment goods. If computer-owning consumers did substitute away from music and into movies, then a significant change (of opposite sign) should exist in the data. A complete description of the regression model is presented after a discussion of the summary statistics.

### 3. ANNUAL EXPENDITURES AND SUMMARY STATISTICS

Table 1 lists the annual mean expenditures on CDs, movie tickets, and prerecorded movies. Inflation adjusted expenditures are presented (in 2003 dollars) for all consumers, computer owning consumers, and non-computer owning consumers. The expenditures are not presented relative to income because each consumer group typically spends less than one percent of its income on CDs, movie tickets, and prerecorded movies, respectively. Panel A of Table 1 shows that the overall mean expenditures on CDs for all households decreased each year, and fell from \$51.09 in 1998 to \$34.95 in 2003. The downward trend for music is more severe than for the movie categories.

**Table 1. Mean Expenditures (2003 \$)**

*Panel A - All Consumers*

	1998	1999	98-99	2000	99-00	2001	00-01	2002	01-02	2003	02-03
Mean CD Expenditures	\$51.09	\$47.92	-	\$46.35	-	\$42.19	<i>b</i>	\$38.76	<i>a</i>	\$34.95	-
No. of Obs. (expenditure > 0)	4,921	6,032	-	5,932	-	5,857	-	6,091	-	5,927	-
Mean Movie Ticket Expenditures	\$105.76	\$105.76	-	\$98.82	<i>a</i>	\$94.91	-	\$104.19	-	\$94.37	-
No. of Obs. (expenditure > 0)	8,665	10,431	-	9,539	-	9,916	-	11,248	-	10,486	-
Mean Prerecorded Movie Expenditures	\$74.18	\$71.82	-	\$71.35	-	\$70.62	-	\$78.76	<i>b</i>	\$76.45	-
No. of Obs. (expenditure > 0)	9,561	11,767	-	11,174	-	11,337	-	12,270	-	12,052	-
<hr/>											
<i>Panel B - Computer Owning Consumers</i>											
Mean CD Expenditures	\$82.36	\$71.10	<i>b</i>	\$64.35	<i>b</i>	\$58.81	<i>b</i>	\$51.47	<i>b</i>	\$44.26	<i>b</i>
No. of Obs. (expenditure > 0)	2,870	3,776	-	4,061	-	4,378	-	4,728	-	4,691	-
Mean Movie Ticket Expenditures	\$166.12	\$164.59	-	\$143.00	<i>b</i>	\$135.64	-	\$144.46	-	\$127.67	-
No. of Obs. (expenditure > 0)	4,870	6,428	-	6,443	-	7,315	-	8,687	-	8,442	-
Mean Prerecorded Movie Expenditures	\$105.21	\$98.78	<i>a</i>	\$95.09	-	\$94.59	-	\$101.07	-	\$97.80	-
No. of Obs. (expenditure > 0)	5,202	6,955	-	7,322	-	8,218	-	9,320	-	9,576	-
<hr/>											
<i>Panel C - Non Computer Owning Consumers</i>											
Mean CD Expenditures	\$29.84	\$29.50	-	\$28.12	-	\$21.10	<i>b</i>	\$19.68	-	\$18.43	-
No. of Obs. (expenditure > 0)	2,003	2,189	-	1,797	-	1,391	-	1,269	-	1,166	-
Mean Movie Ticket Expenditures	\$65.08	\$59.72	-	\$54.50	-	\$42.75	<i>b</i>	\$43.69	-	\$35.26	-
No. of Obs. (expenditure > 0)	3,715	3,873	-	2,969	-	2,401	-	2,351	-	1,906	-
Mean Prerecorded Movie Expenditures	\$53.53	\$50.85	-	\$47.47	-	\$40.13	<i>b</i>	\$45.15	-	\$38.83	<i>a</i>
No. of Obs. (expenditure > 0)	4,286	4,686	-	3,700	-	2,927	-	2,743	-	2,326	-

All statistics use the required full-sample and replicate weights supplied with the Consumer Expenditure Survey data, and all dollar amounts are converted to 2003 dollars using the CPI-U for all items less food and energy. The letters "a" and "b" denote statistically significant changes (for the annual mean) at the 10 and 5 percent levels of significance, respectively. The symbol "-" denotes that the change was statistically insignificant.

Movie ticket expenditures trend slightly downward until 2002, and then decrease again in 2003. Consumers spent, on average, \$105.76 on movie tickets in 1998 and \$94.37 in 2003. The overall trend for the final category, prerecorded movies, is basically flat; the average household expenditure in 1998 was \$74.18 and only 3 percent higher in 2003 (\$76.45). Despite this overall trend, there was a statistically significant increase (at the five percent level of significance) of approximately \$8 in the average prerecorded movie expenditure for all households. This increase is not, however, concentrated among computer owners.

For the computer-owning households, spending on all three categories of goods exhibit a downward trend, with the sharpest declines in the music category. The average computer-owning consumer spent significantly less (at the five percent level) on CDs in each successive year of the sample period. As of 2003, the group's average expenditure on CDs was nearly 50 percent less than in 1998 (\$44.26 vs. \$82.36). The average movie ticket expenditure also trended downward, from \$166.12 in 1998 to \$127.67 in 2003 (although the change was statistically significant only in 2000). As for the final category of goods, the trend in computer-owning households' prerecorded movie spending was also slightly downward, starting at \$105.21 in 1998 and ending up at \$97.80 in 2003. Of course, these trends suggest that the goods are complements rather substitutes.

These trends would not exist in the data if consumers substituted movie ticket and/or prerecorded movie purchases for a significant portion of their music purchases. To be sure, these averages are not sufficient to attribute a causal relationship to file sharing and music expenditure declines. Nonetheless, the average computer-owning household's expenditure did not increase for either of the goods often mentioned as possible substitutes for CDs (movie tickets and prerecorded movies). In fact, expenditures for all three categories trend slightly downward even for the non-computer owning households.<sup>7</sup> As an added check, Table 2 provides summary statistics for several demographic variables.

**Table 2 - Mean Income, Age and Family Size (2003 \$)**

*Panel A - All CUs*

	1998	1999	2000	2001	2002	2003
Annual Income	\$47,325	\$48,679	\$48,884	\$51,470	\$51,550	\$52,684
Income Change	-	2.86%	0.42%	5.29%	0.16%	2.20%
Reference Person Age	42.33	42.61	42.50	42.53	43.16	43.26
Family Size	2.66	2.67	2.68	2.69	2.65	2.65
No. of Obs. (expenditure > 0)	13,280	16,371	15,533	15,809	17,295	16,892

*Panel B - Computer Owning CUs*

Annual Income	\$60,647	\$61,227	\$59,440	\$59,935	\$59,199	\$59,323
Income Change	-	0.96%	-2.92%	0.83%	-1.23%	0.21%
Reference Person Age	41.58	42.30	42.05	42.31	42.63	42.93
Family Size	2.81	2.82	2.82	2.79	2.76	2.73
No. of Obs. (expenditure > 0)	6,864	9,355	9,825	11,109	12,757	13,062

*Panel C - Non Computer Owning CUs*

Annual Income	\$33,417	\$29,441	\$30,855	\$31,550	\$29,365	\$29,336
Income Change	-	-11.90%	4.80%	2.25%	-6.93%	-0.10%
Reference Person Age	43.35	43.22	43.47	43.34	44.93	44.53
Family Size	2.51	2.48	2.45	2.48	2.33	2.35
No. of Obs. (expenditure > 0)	6,301	6,829	5,495	4,411	4,232	3,621

All income figures are converted to 2003 dollars using the CPI-U for all items less food and energy. Each statistic is reported for consumers with non-missing expenditures in at least one of the three expenditure categories (CDs, movie tickets, or prerecorded movies). The 1998 to 1999 income change for non computer owning consumers is statistically significant at the 5 percent level of significance; none of the other changes in annual income are statistically significant.

The statistics presented on Table 2 are not suggestive of any major demographic shifts in the CEX sample that would explain the declines in spending on music and movies. For example, Table 2 shows that mean annual income for all consumers rose steadily over the six year period, from \$47,325 in 1998 to \$52,684 in 2003.<sup>8</sup> The average computer-owning household's income fluctuated slightly from year-to-year,

<sup>7</sup>These complementary trends are consistent with aggregate CD, movie ticket and prerecorded movie sales (see Liebowitz, 2004).

<sup>8</sup>Since the CEX definition of before tax income changed in 2001, wage and salary income is used to measure annual income.

and ended the six-year sample period lower relative to where it began (\$59,323 in 2003 vs. \$60,647 in 1998). Non-computer owning households also experienced a slight decline in income, from \$33,417 in 1998 to \$29,336 in 2003. As for the average age and family size, households with and without computers are fairly similar. For households that do not own a computer, the average age is a bit older and the average size is a bit smaller.

In 2003, for instance, the average age and family size for households with a computer were 42.93 years old and 2.73 individuals, respectively.<sup>9</sup> For households that report not owning a computer, the average age and family size were 44.53 and 2.35, respectively. These differences appear minor, but our main regressions do test for whether these changes alone could explain the documented declines in spending on music, movie tickets, and prerecorded movies.

#### 4. REGRESSION MODEL AND RESULTS

**4.1. Regression Model.** The regression framework used in this paper employs 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 associated with file-sharing activity.<sup>10</sup> The tests are run separately for the three categories of goods, and they are also run on only those consumers who made purchases in all three categories.

Using the pooled data from years 1998 and 2003, the following weighted least squares regression (WLS) is run for three categories of entertainment expenditures:

$$\log(ENTEXP) = \hat{\beta}_1 + \hat{\beta}_2CMP + \hat{\beta}_3X' + \hat{\beta}_4Y03 + \hat{\beta}_5CMPINT \quad (1)$$

Model (1) is run separately using three categories of entertainment expenditures for its dependent variable ( $\log(ENTEXP)$ ). These dependent variables are as follows: the natural log of CD expenditures, the natural log of movie ticket expenditures, and the natural log of prerecorded movie expenditures (purchases and rentals). The key independent variables are  $CMP$ ,  $Y03$ , and  $CMPINT$ .

The  $Y03$  variable is a year dummy, set to one for expenditures made in the year 2003, and the coefficient on  $CMPINT$  is the difference-in-differences estimator. To compute  $CMPINT$ , the year dummy ( $Y03$ ) is multiplied by each consumer's  $CMP$  dummy (set to one for consumers owning a computer). Because the model regresses the natural logarithm of CD expenditures, the coefficient on  $CMPINT$  measures the change in the relationship between computer ownership and entertainment expenditures between 1998 and 2003. To control for preferences, the model contains a vector ( $X$ ) of control variables, including income, family size and age. Because the CEX definition of before tax income was changed in 2001, the natural

---

<sup>9</sup>For households with married individuals, the mean age is calculated by averaging the age of the spouses.

<sup>10</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.

logarithm of wage and salary income is used as the independent variable controlling for income.

The CEX family size variable is reclassified so that six or more people represent the largest family size, and the age variable is grouped into four categories: under 31, between 31 and 55, between 56 and 65, and over 65 (for married households, the age of the spouses are averaged). Alternative specifications, using the age of the household's children, as well as race and region variables, yield results nearly identical to those presented below (available from the author). While testing the model on only younger households may be more appropriate, the CEX is not the best data from which to perform such an analysis due to a lack of usable observations.<sup>11</sup>

#### 4.2. WLS Regression Results.

**Table 3, Difference-in-Differences Estimates on Pooled Expenditures, 1998 to 2003 (2003\$)**

Dependent Variable (in Natural Logs):	CD Expenditures	Movie Ticket Expenditures	Prerecorded Movie Expenditures
<i>Independent Variables:</i>			
Intercept	2.36 (0.1051) <i>0.0000</i>	1.17 (0.1174) <i>0.0000</i>	1.77 (0.0997) <i>0.0000</i>
lny	0.11 (0.0102) <i>0.0000</i>	0.19 (0.0135) <i>0.0000</i>	0.12 (0.0089) <i>0.0000</i>
CMP	0.16 (0.0296) <i>0.0070</i>	0.16 (0.0295) <i>0.0000</i>	0.04 (0.0281) <i>0.1800</i>
Family Size	-0.01 (0.0072) <i>0.3560</i>	0.03 (0.0068) <i>0.0000</i>	0.09 (0.0071) <i>0.0000</i>
Age	-0.04 (0.0221) <i>0.0480</i>	0.05 (0.0159) <i>0.0060</i>	-0.080 (0.0125) <i>0.0000</i>
Y03	-0.04 (0.04548) <i>0.4420</i>	0.01 (0.0441) <i>0.7840</i>	0.14 (0.0370) <i>0.0010</i>
CMPINT	-0.15 (0.0486) <i>0.0050</i>	0.016 (0.0485) <i>0.7460</i>	0.015 (0.0379) <i>0.6890</i>
No. of Observations:	9,721	16,631	19,342
R <sup>2</sup> :	0.029	0.0650	0.0480

Table 3 presents WLS estimates of model (1) using data from 1998 and 2003, and employs the CEX full-sample and replicate weights. The first result column presents results using the natural log of CD expenditures as the dependent variable. The second and third column results are obtained using the natural log of movie ticket expenditures and the natural log of prerecorded movie expenditures (purchase and rental), respectively, as the dependent variable. In each column, for the respective independent variables, the first row presents the estimated coefficient, the second row the standard error, and the third row the *p*-value. The key coefficient is the difference-in-differences estimator (CMPINT), an interaction of a computer-ownership indicator variable and a time-indicator for the year 2003. See the text for a description of the remaining independent variables in model (1).

<sup>11</sup>For example, only about seven percent of all 2003 CEX households have a reference person age 25 or below (before any distinction is made between computer owners and non-owners who purchased a particular type of good). As mentioned above, using the age of households' children as an alternative variable does little to change the main results of the paper.

Table 3 presents the three main test results.<sup>12</sup> The first column of results, from running model (1) with the natural log of CD expenditures as the dependent variable, shows that the coefficient on the difference-in-differences estimator (*CMPINT*) is negative and statistically significant (at the one percent level). Because the model is in logs, this estimate suggests that increased file-sharing activity could have resulted in a 15 percent decline in CD purchases between 1998 and 2003. The second and third columns present results using the natural log of movie ticket purchases and prerecorded movie expenditures, respectively, as the dependent variable. In these two cases, the *CMPINT* coefficient is statistically indistinguishable from zero, with *p*-values near 70 percent and point estimates of about 0.01.

As for the control variables, Table 3 shows that income is positively related to all three expenditure categories, with a slightly larger coefficient for movie ticket purchases. For movie ticket purchases, the coefficient on the log of income suggests that a one percent decline in income would lead to a 0.19 percent decline in movie ticket purchases. Because the computer owning households' income was approximately 2.2 percent less in 2003 than 1998, the coefficient on income explains less than one percent of the decline in computer owners' expenditures on movie tickets.<sup>13</sup> Income changes (as well as family size and age differences) also appear to be too small to have had any impact on the sales of CDs and/or prerecorded movies. The results on Table 3 also affirm the positive relationship between computer ownership and expenditures on each of the goods (as expected based on the summary statistics in Table 1, the *CMP* dummy variable is positive).

Table 4 presents partial results (only for the *CMPINT* coefficient) from running model (1) on data from only consumers who purchased all three types of goods (CDs, movie tickets, and prerecorded movies). Though the sample size is much smaller, leading to larger standard errors, these results for the *CMPINT* coefficient are fairly similar to those presented in Table 3. While no longer significant at the one percent level, the *CMPINT* coefficient is statistically significant at the ten percent level, and the corresponding estimate for movie tickets and prerecorded movies are (as before) indistinguishable from zero.

Whether using consumers who spent on any of the three goods, or only consumers who spent on all three goods, there is no evidence of a substitution effect between CDs, movie tickets and prerecorded movies. At the very least, there is a clear (statistically significant) negative change in the relationship between computer ownership and CD purchases in the CEX data. There is no distinguishable change in the relationship between computer ownership and either movie ticket or prerecorded movie expenditures. Of course, these results require further scrutiny before they can be used to suggest file sharing caused music sales to decline.

---

<sup>12</sup>The main test results from model (1) employ all appropriate CEX weights (see Appendix), but running model (1) on non-weighted data does not materially impact the results presented below.

<sup>13</sup>To be more precise, the average change in computer owning consumers' income appears to explain only about 0.76 percent of the decline in the average expenditure for movie tickets ( $.16 + .19 \times 2.18 = 0.76$ ).

**Table 4, Difference-in-Differences Estimates on Pooled Expenditures 1998 to 2003 (2003\$)**  
**(Only Includes Consumers Who Purchased CDs, Movie Tickets, *And* Prerecorded Movies)**

Dependent Variable (in Natural Logs):	CD Expenditures	Movie Ticket Expenditures	Prerecorded Movie Expenditures
CMPINT	-0.15 (0.0875) <i>0.1000</i>	0.026 (0.0786) <i>0.7440</i>	-0.002 (0.0798) <i>0.9780</i>
No. of Observations:	4,937	4,937	4,937
R <sup>2</sup> :	0.028	0.0520	0.0730

Table 4 presents (partial) results from model (1) using data from 1998 and 2003; it employs the CEX full-sample and replicate weights. The first result column presents results using the natural log of CD expenditures as the dependent variable. The second and third column results are obtained using the natural log of movie ticket expenditures and the natural log of prerecorded movie expenditures (purchase and rental), respectively, as the dependent variable. In each column, for the respective independent variables, the first row presents the estimated coefficient, the second row the standard error, and the third row the  $p$ -value. The coefficient presented is the difference-in-differences estimator (CMPINT), an interaction of a computer-ownership indicator variable and a time-indicator for the year 2003. The model is run on only those consumers with expenditures on *all three* goods categories (CDs, movie tickets, and prerecorded movies (purchase and rental)).

**4.3. Two Stage Least Squares (2SLS) Estimates.** The main complicating factor for the above tests is that computer owners in 1998 and 2003 are not necessarily two homogenous groups. To begin with, the CEX data show that more and more consumers are purchasing computers. As seen on Table 1, the number of computer owning consumers with expenditures on CDs, movie tickets, and prerecorded movies has nearly doubled for each good category. Simultaneously, the corresponding number of non-computer owners in the CEX has dropped nearly in half. For example, the number of computer owning households with expenditures on CDs was 2,870 in 1998, and 4,691 in 2003. On the other hand, the corresponding number of consumers without a computer was 2,003 in 1998, and 1,166 in 2003.

Although the income and demographic data are fairly uniform in the CEX, the difference-in-differences estimator in (1) could be biased downward if, for example, computer owners in 2003 purchased computers specifically to download music (or movies). Because a panel with the same computer owning consumers included for all six years cannot be constructed with CEX data, possible endogeneity is checked for by using a two-stage least squares (2SLS) approach. The following three instruments are used: (1) an indicator for households whose reference person has some college education, (2) an indicator for households with children, and (3) a linear combination of both.<sup>14</sup>

All instruments are highly correlated with computer ownership, but Hausman tests are somewhat inconclusive – for movie ticket and prerecorded movie purchases

<sup>14</sup>Because numerous regression packages are unable to run 2SLS with the CEX supplied *replicate* weights, the 2SLS estimates are run with non-weighted data and employ heteroscedastic-robust standard errors. In earlier research, I also compiled synthetic cohorts by including in the sample only those consumers with similar demographic characteristics in each time period. Experimentation with synthetic cohorts based on income, age, family type, and family size, respectively, produced results similar to the WLS results on Table 3, although the sample sizes were much smaller.



– on whether there is an endogeneity problem in the data. There is no such ambiguity when using CD expenditures (with either of the instruments). Because of the uncertainty using movie purchases, the 2SLS estimates of the difference-in-differences estimator are presented below.

Table 5. Two-Stage Least Squares on Pooled Expenditures, 1998 and 2003 (2003S)

Instrument for Computer Owners:	CD Expenditures Households			Movie Ticket Expenditures Households			Prerecorded Movie Expenditures Households		
	College Education	With Children	Combination	College Education	With Children	Combination	College Education	With Children	Combination
<i>2SLS Coefficients of CMPINT</i>	0.20	-0.15	0.07	0.88	0.56	0.78	0.06	-0.68	-0.10
<i>Standard Error</i>	(0.2062)	(0.2605)	(0.1636)	(0.2217)	(0.3198)	(0.1806)	(0.1677)	(0.3235)	(0.1493)
<i>p-value</i>	0.3320	0.5550	0.6810	0.0000	0.0820	0.0000	0.7290	0.0360	0.5140
<i>p-value on residual</i>	0.1180	0.8810	0.1800	0.0000	0.0750	0.0000	0.8680	0.0250	0.4020
<i>OLS Coefficient of CMPINT</i>	-0.12			-0.01			0.03		
<i>Standard Error</i>	(0.0408)			(0.0375)			(0.0329)		
<i>p-value</i>	0.0050			0.8800			0.3470		
No. of Observations:	9,721	9,721	9,721	16,631	16,631	16,631	19,342	19,342	19,342

Table 5 presents OLS and 2SLS estimates of the difference-in-differences coefficient (CMPINT) from model (1). The OLS coefficient of CMPINT interacts an indicator variable for computer ownership with an indicator variable for the year 2003. The 2SLS estimates use the following three instrument sets: (1) an indicator for a household whose reference person reports at least some college education; (2) an indicator for whether a household includes children (up to the age of 17); and, (3) a linear combination of the first two indicators. The statistic “p-value on residual” is from the residual used in a Hausman test (for the corresponding expenditure category and instrument set). Because numerous regression packages are unable to run 2SLS with the CEX supplied *replicate* weights, the OLS and 2SLS estimates are on non-weighted data and employ heteroscedastic-robust standard errors. The results are divided into three groups of three columns. The first group is obtained running OLS and 2SLS on CD expenditures, the second on movie ticket expenditures, and the third on prerecorded movie expenditures (purchase and rental).

Table 5 shows that when either of the instruments are used with CD expenditure data, Hausman tests cannot reject that the 2SLS estimates are statistically the same as their OLS counterparts. Using CD expenditures, the 2SLS coefficients for the difference-in-differences estimator are statistically indistinguishable from zero using each of the instruments (though Hausman tests suggest that the 2SLS estimates are not necessary here). For movie ticket expenditures, Hausman tests suggest that 2SLS estimates are needed only when the “college education” instrument is included in the instrument set.<sup>15</sup> Using movie ticket purchases, the 2SLS estimates suggest that there is a (statistically significant) positive change in the relationship between computer ownership and movie ticket purchases – a finding that is not consistent with the CEX averages (see Table 1).

For prerecorded movies, Hausman tests suggest that 2SLS estimates are needed in only one instance: when the “households with children” instrument is used. In this case, the 2SLS estimate of *CMPINT* suggests that there is a statistically significant negative relationship between computer ownership and prerecorded movie purchases. The direction of this estimate is consistent with the CEX data (for computer owning consumers), but the size of the coefficient appears implausible (-0.68) because the average consumer in the CEX spent about the same on prerecorded movies in 1998 and 2003, while the typical computer owning household in the CEX spent only about 7 percent less in 2003 vs. 1998.

This 2SLS estimate, though, suggests that the average computer owning consumer spent almost 70 percent less on prerecorded movies – about ten times as large as what is in the CEX data. Regardless, there is little evidence in these 2SLS estimates that suggest a significant substitution effect between movie and CD expenditures. If anything, one could argue that these test results indicate a

<sup>15</sup>For movie ticket expenditures, when the “households with children” instrument is used, the failure to reject that the 2SLS and OLS models are the same is marginal.

substitution between movie tickets and prerecorded movies. However, re-running these tests on only consumers who purchased all three types of goods lends little support to this conclusion.

Table 6 presents the 2SLS results using only those consumers with expenditures on CDs, movie tickets, and prerecorded movies. Naturally, the smaller sample size produces much larger standard errors. Still, as before, Hausman tests cannot reject that the 2SLS estimates are statistically the same as their OLS counterparts when CD expenditure data are used.<sup>16</sup> Notwithstanding any implausibly large (or small) point estimates, the possible substitution between movie tickets and prerecorded movies all but disappears in this more restrictive sample.

**Table 6, Two-Stage Least Squares on Pooled Expenditures, 1998 and 2003 (2003S)**  
(Only Includes Consumers Who Purchased CDs, Movie Tickets, And Prerecorded Movies)

Instrument for Computer Owners:	CD Expenditures			Movie Ticket Expenditures			Prerecorded Movie Expenditures		
	College Education	Households With Children	Combination	College Education	Households With Children	Combination	College Education	Households With Children	Combination
<i>2SLS Coefficients of CMPINT</i>	0.52	-0.38	0.13	1.33	-0.24	0.63	0.43	-0.87	-0.14
<i>Standard Error</i>	(0.3628)	(0.3994)	(0.2667)	(0.4439)	(0.4704)	(0.3161)	(0.4099)	(0.4721)	(0.3092)
<i>p-value</i>	0.1490	0.3470	0.6380	0.0030	0.6140	0.0450	0.2890	0.0640	0.6410
<i>p-value on residual</i>	0.0750	0.4910	0.3790	0.0100	0.5850	0.0410	0.3030	0.0510	0.5860
<i>OLS Coefficient of CMPINT</i>	-0.10			0.02			0.02		
<i>Standard Error</i>	(0.0643)			(0.0722)			(0.0734)		
<i>p-value</i>	0.1040			0.8320			0.7690		
No. of Observations:	4,937	4,937	4,937	4,937	4,937	4,937	4,937	4,937	4,937

Table 6 presents OLS and 2SLS estimates of the difference-in-differences coefficient (CMPINT) from model (1). The OLS coefficient of CMPINT interacts an indicator variable for computer ownership with an indicator variable for the year 2003. The 2SLS estimates use the following three instrument sets: (1) an indicator for a household whose reference person reports at least some college education; (2) an indicator for whether a household includes children (up to the age of 17); and, (3) a linear combination of the first two indicators. The statistic “p-value on residual” is from the residual used in a Hausman test (for the corresponding expenditure category and instrument set). Because numerous regression packages are unable to run 2SLS with the CEX supplied *replicate* weights, the OLS and 2SLS estimates are on non-weighted data and employ heteroscedastic-robust standard errors. The results are divided into three groups of three columns. The first group is obtained running OLS and 2SLS on CD expenditures, the second on movie ticket expenditures, and the third on prerecorded movie expenditures (purchase and rental). All tests are performed using only data from consumers with expenditures on all three goods categories (CDs, movie tickets, and prerecorded movies (purchase and rental)).

For instance, none of the corresponding estimates on movie tickets and prerecorded movies have opposite signs.<sup>17</sup> Most importantly for the purpose of this paper, there is little evidence of any substitution between CD and movie expenditures (either movie tickets or prerecorded movies). The only statistically significant estimate on movie ticket expenditures is found when the “college education” instrument is included. Although the 2SLS estimate of the *CMPINT* coefficient on movie tickets is positive and statistically significant (depending on the instrument used), there is no corresponding significant negative estimate for either CD or prerecorded movie expenditures. Even ignoring statistical significance, there is no clear finding that the corresponding signs of these 2SLS estimates are of opposite direction.

## 5. CONCLUDING REMARKS

The impact of copying on sales of originals is an age-old public policy debate that has once again surfaced due to technological change. As more researchers have examined the relationship between Internet file sharing and music sales, there appears to be mounting evidence that digital copying negatively impacted music sales

<sup>16</sup>Although, the failure to reject is marginal when using the “college education” instrument.

<sup>17</sup>Using the linear combination of instruments for the estimates on prerecorded movies does produce a negative point estimate (-0.14), but because its standard error is 0.31, this estimate cannot be considered strictly negative.

during the last few years. Some of the more recent empirical work in this area (such as Michel, 2005, and Hong, 2004) uses the household-level Consumer Expenditure Survey data (CEX) to test whether digital file sharing may have negatively affected music sales. These micro-level studies do not, however, test whether consumers substituted out of music and into movies.

Supporters of Internet file sharing argue that this sort of substitution effect could explain the declines in music sales that surfaced shortly after the first file-sharing service (Napster) was launched in 1999. The present paper contributes to the literature by using the CEX data to test whether computer owners, serving as a proxy for possible file sharers, simultaneously spent less on music and more on movie tickets and/or prerecorded movies. The test results presented herein show that the CEX data fail to support such a substitution effect, though there are data limitations using the CEX to test for file-sharing activity. For example, while there is evidence of a significantly weaker relationship between computer ownership and music purchases, it is impossible to separate file-sharing computer owners from non-file-sharing computer owners using the CEX. Still, there is no evidence in the CEX that computer owners spent less on music because they spent more on either movie tickets or prerecorded movies.

#### REFERENCES

- Brogan, D.** (1998), "Pitfalls of Using Standard Statistical Software Packages for Sample Survey Data", In *Encyclopedia of Biostatistics*, Armitage and Colton (eds.), New York: John Wiley & Sons.
- Hong, S.** (2004), "The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster", Stanford University Working Paper, November.
- Michel, N. J.** (2005), "The Impact of Digital File Sharing on the Music Industry: A Theoretical and Empirical Analysis", Working Paper, Nicholls State University.
- Michel, N. J.** (2003), "A Theoretical and Empirical Analysis of the Impact of the Digital Age on the Music Industry", Graduate Dissertation, University of New Orleans, available at <http://www.uno.edu/theses/available/etd-11212003-100808/>.
- 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 Publication No. 82-1366, Vital and Health Statistics. Series 2, No. 92.
- Liebowitz, S. J.** (2004), "Pitfalls in Measuring the Impact of File-Sharing", Working paper, University of Texas at Dallas, School of Management.
- Wooldridge, J.** (2003), *Introductory Econometrics; A Modern Approach* (2nd ed.), US: Thomson-South-Western.
- Wooldridge, J.** (2002), *Econometric Analysis of Cross Section and Panel Data*, Cambridge, MA: MIT Press.

### Appendix – Consumer Expenditure Survey

This paper uses the Consumer Expenditure Survey (CEX) interview survey data files, in which consumers respond to questions regarding their purchases and demographic characteristics during the three months prior to their interview. Rather than a simple random sample, the CEX data are collected based on a stratified sample design, whereby two primary sampling units (PSUs) exist per stratum. To ensure that sample sizes from the CEX data are representative of the U.S. population, each Consumer Unit (a measure analogous to a household) is assigned a full-sample weight. To allow for more precise variances and standard errors to be calculated, the data are supplied with 44 half-sample replicate weights. 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. However, 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. Because the BLS does not release detailed strata information, all statistics calculated herein, unless otherwise noted, employ the appropriate weights.

For the summary statistics, the replicate weights are used to construct 90 and 95 percent confidence intervals around the changes in mean annual expenditures. For both summary statistics and regressions, only observations coded as “complete income reporters” are used (a designation that indicates at least one major income earner for the household was interviewed). The unit of observation in the CEX is referred to as the “consumer unit.” In this paper, we interchangeably refer to consumer units as either “consumers” or “households.” All expenditures are compiled by aggregating the reported values, according to universal classification codes (UCC), in the appropriate monthly expenditure (MTAB) files.

The UCC code provided for music purchases includes expenditures on the following items: compact discs, tapes, needles or records not from a club. The UCC code provided for movie ticket purchases includes expenditures on the following items: admission fees for movie, theater, concert, opera or other musical series (single admissions and season tickets). The UCC codes provided for prerecorded movie purchases and rentals include expenditures on the following items: video cassettes, tapes and discs. As noted in the beginning of the paper, test results using video games are not presented in the main text. Because the resulting sample sizes are so small (running model (1) on video game expenditures results in a sample of only 2,847 observations), and because the test results are similar to those for movies anyway, these results are omitted for space considerations. One possible reason that so many CEX households have missing video game expenditures is that the UCC attempts to record purchases on “TV, computer games and computer game software,” perhaps missing other types of video games.