

# MEASURING THE EFFECT OF FILE SHARING ON MUSIC PURCHASES\*

ALEJANDRO ZENTNER  
*University of Texas at Dallas*

## ABSTRACT

File sharing may substantially undermine the intellectual property rights of digital goods. This paper concentrates on the music industry. I estimate the effect of music downloads on the probability of purchasing music using a European individual-level cross section of 15,000 people from 2001. A simple comparison of means shows that people who regularly download music online are more likely to buy music. The positive relationship persists when controlling for observed characteristics. However, simultaneity between tastes for music and peer-to-peer usage makes it difficult to isolate the causal effect of music downloads on music purchases. To break that simultaneity, this paper uses measures of Internet sophistication and the speed of the Internet connection as instruments. The results suggest that peer-to-peer usage reduces the probability of buying music by 30 percent. On the basis of my estimates, back-of-the-envelope calculations indicate that—without downloads—sales in 2002 would have been around 7.8 percent higher.

## I. INTRODUCTION

THE global music industry was quite successful during the 1990s. According to the International Federation of the Phonographic Industry (IFPI), album sales grew from US\$24.1 billion in 1990 to US\$39.4 billion in 1996 and remained at a high level until 1999. Those days are over, and the industry is now struggling. Global music sales have been falling for the last 3 years (Figures 1 and 2). Global sales (units) of CDs—the most popular format—fell in 2001 for the first time since its introduction in 1983.

This downturn coincides with the proliferation of online music file sharing. In June 1999, Napster was created, making the work of many artists available for free. Its popularity was immediate. According to Mediametrix, a company that provides Internet rankings and measurement, Napster was the fastest software adoption in history (Associated Press 2000). Given its impact, the

\* I wish to thank Austan Goolsbee, Steven Levitt, and Hugo Sonnenschein for their advice and encouragement. I also thank Nathaniel Baum-Snow, Gary Becker, Dennis Carlton, Carolina Czastkiewicz, Julio Elias, Jonah Gelbach, Thomas Hubbard, Anupam Babu Jena, David Levine, Jose Liberti, Douglas Lichtman, Stan Liebowitz, Laura Martinolich, Kevin Murphy, Alejandro Rodriguez, Chad Syverson, and Frank Wolak for their helpful comments.

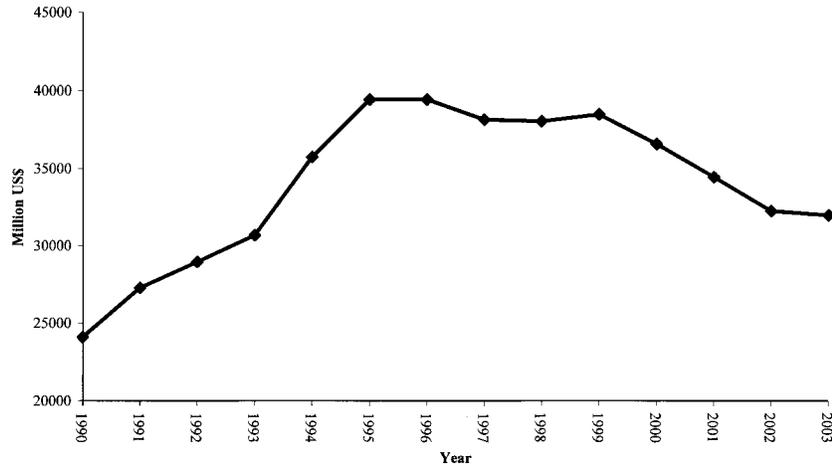


FIGURE 1.—Global music sales in million US\$ (IFPI 2004a). Local currency values are converted to dollars at the exchange rate of each year. Music videos are included since 2001.

Recording Industry Association of America (RIAA) soon filed a motion against Napster in the U.S. District Court of San Francisco for “engaging in or enabling, or facilitating others in copying, downloading, uploading, transmitting, or distributing plaintiffs copyrighted musical compositions and sound recordings, protected by either federal or state law, without express permission of the rights owner” (*A&M Records v. Napster*, 2001 U.S. Dist. LEXIS 2186 [N.D. Cal., March 5, 2001]). Napster was shut down in February 2001. However, many peer-to-peer alternatives for sharing music over the Internet remain available. In 1993, Kazaa claimed to be the most downloaded application, with more than 230 million users worldwide (Black 2003b). According to the Yahoo Buzz Index, an index that measures Internet searches using the Yahoo search engine, “Kazaa” was the number one searched term on the Internet in 2003.<sup>1</sup>

File sharing has an important online presence. In May 2002, IFPI estimated that there were 3 million simultaneous global users and 500 million files available for copying at any given time (IFPI 2002a). In 2002, NetPD, a company that provides protection services to copyright owners whose material is being pirated through the Internet, reported that 3.6 billion files were downloaded monthly, of which around 60–70 percent were music files (Rowan 2002). The most popular albums are available for online sharing almost immediately after release and in some cases, such as Oasis and Eminem, even before (IFPI 2002b). Copy protection technology has been ineffective.

<sup>1</sup> The Yahoo Buzz Index is available at <http://buzz.yahoo.com>.

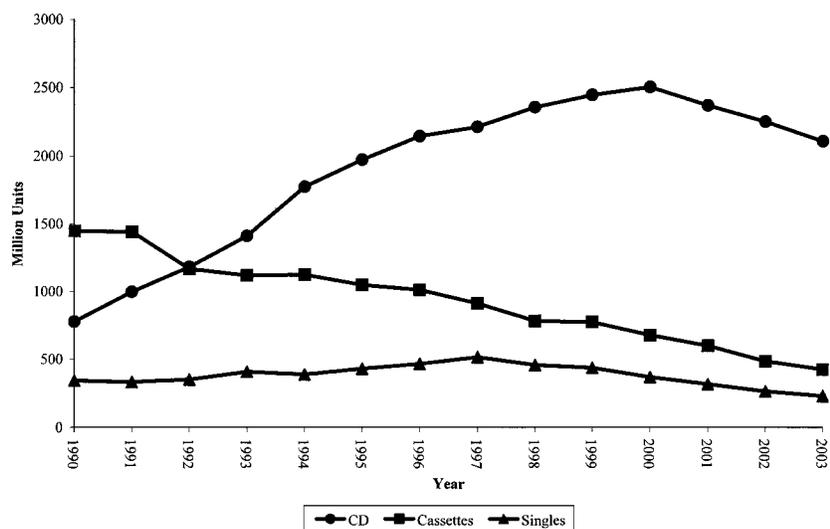


FIGURE 2.—Global music sales in million units (IFPI 2004a)

The development of broadband facilitates music sharing. A soundtrack that takes more than 12 minutes to download with a dial-up connection can be downloaded in as little as 20 seconds with a high-speed connection. Napster and its successors were banned in many universities because the very fast connections apparently induced so much file sharing that there was little available bandwidth left for anything else. In the case of the University of Illinois at Urbana-Champaign, this amounted to 75 percent of the total bandwidth (Ante 2000).

File sharing is not limited to music. The development of fast connections is extending downloading to other digital goods such as movies, software, video games, and books. Some movies are available online during the opening week of theatrical release and before the authorized DVD is available.

Does file sharing reduce music sales? If so, what is the magnitude of the impact of file sharing on music sales? Is file sharing responsible for the recent drop in sales? Inside the music industry, it is generally accepted that music sales have been affected negatively by file sharing, but there has not been much empirical work done to measure the sign and extent of this effect.

The question is important because file-sharing technology may undermine the effective protection offered by copyright. Strong property rights create monopoly distortions, but weak property rights may lead to low levels of creation of artistic work or innovation.<sup>2</sup> The balance between these opposing

<sup>2</sup> Romer (2002) analyzes this trade-off for the case of musical recordings. See Boldrin and Levine (2002) and Klein, Lerner, and Murphy (2002) for a discussion of the effect of file sharing on copyright value.

forces has rarely been empirically addressed. Knowledge of the impact of file sharing on sales is an essential part of the information needed for balancing these factors in the case of digital goods. Another important component for this balancing—which is not addressed in this paper—is the elasticity of the supply of talent.

The objective of this paper is to obtain estimates of the impact of online file sharing on music purchases. The main difficulty in the estimation of the effect of downloads on purchases of music, using individual-level data, is the presence of unobserved heterogeneity in music tastes among individuals. To control for this heterogeneity, it would be ideal to follow individuals through time. In the absence of such panel data, I use a major European cross-sectional database of 15,000 observations from October 2001. With these data, I attempt to identify the causal effect of downloading MP3 files on the probability of buying music.

Ordinary least squares (OLS) estimates show a positive—although not significant—relationship between peer-to-peer usage and the probability of purchasing music. However, OLS estimates are biased and inconsistent in the presence of unobserved heterogeneity. People who download music may have a greater taste for music. To overcome this obstacle, my approach is to use instrumental variables techniques. This requires an instrumental variable that explains peer-to-peer systems usage but is otherwise unrelated to music purchases. I employ measures of Internet sophistication and access to broadband as instruments. My estimates indicate that peer-to-peer usage reduces the probability of buying music. Using measures of sophistication as instruments, I find that music downloading reduces the probability of buying music by 30 percent.

Because of the large change in the estimated effect found in the instrumented regressions, special attention is given to the validity of the instruments and robustness of the results. First-step regressions show a positive and significant correlation of the instruments with peer-to-peer usage. I present tests of overidentifying restrictions to analyze the exogeneity of the instruments. I also analyze the exclusion restrictions by studying how MP3 file downloading affects purchases of other entertainment goods and by studying the relationship of the instruments with variables that may signal taste for music and the value of time. I analyze the robustness of the results by restricting the analysis to Internet users and to individuals who had an Internet connection before Napster started—to control for the possibility that people went online in order to download music.

The data employed in this paper do not contain the information on the intensity of music purchases and downloading necessary to make a direct analysis of the effect of file sharing on music sales. However, using my estimate for the reduction in the probability of buying music, back-of-the-envelope calculations indicate that—without file sharing—sales in 2002 would have been around 7.8 percent higher in the countries considered.

The paper is organized as follows. Section II provides an overview of the literature. Section III summarizes the main features of the music industry. Section IV describes the data. Section V presents the empirical strategy and results. Section VI discusses the implications of the estimated coefficients for music sales. Finally, Section VII concludes.

## II. RELATED LITERATURE

To the best of my knowledge, the first empirical attempt to measure the effect of online file sharing on music sales is the one prepared in the legal battle against Napster. To measure this effect, the plaintiff hired Soundscan, a company that developed an information system to capture point-of-sale data on music sales in more than 18,000 stores throughout the United States. Soundscan (*A&M Records v. Napster*, 114 F. Supp. 2d 896 [2000]) compared sales means for the first quarter of the years 1997, 1998, 1999 (when Napster was not available) and 2000 (when Napster was available) for all stores, stores within 1 mile of any college or university, stores within 1 mile of the top 40 most-wired universities, and stores near universities that banned Napster use after the first quarter of 2000. From the first quarter of 1999 to the first quarter of 2000, national sales grew 6.6 percent, sales near all universities dropped 2.6 percent, sales near most-wired schools dropped 6.2 percent, and sales near schools where Napster was banned after the first quarter of 2000 fell 8.1 percent. However, as noted by Fader (2000) in the *Napster* case, sales near universities had been falling since 1998, at a time when Napster was not available and national sales were growing, which casts doubts on the conclusion of Soundscan's report. More recent empirical attempts to measure this effect are Liebowitz (2003), Zentner (2005a), Oberholzer-Gee and Strumpf (2005), and Rob and Waldfogel (2004).<sup>3</sup>

Using a time series of aggregate data for the United States, Liebowitz (2003) analyzes the effect of a variety of reasons that could explain the drop in music sales, including income, prices of albums, prices of complements and substitutes, recording formats, interest in music, demographics, and distribution channels. He finds that these alternative reasons cannot explain the observed reduction in sales.

Zentner (2005a) uses Internet and broadband usage by country as proxies for music downloaders and documents that countries with greater Internet and broadband penetration have suffered greater reductions in music sales. Some types of music would be expected to be more susceptible to file sharing than others. For a given country, international repertoire—as opposed to domestic repertoire—is music that it is not produced inside that country. Music produced in the United States is more popular among young individ-

<sup>3</sup> Hui and Png (2003) analyze the effect of commercial piracy on music sales with data from the years 1994–98. See Liebowitz (2005) for a more extensive analysis of the literature including several more recent studies.

uals in international markets who are more likely to share files online, whereas the domestic repertoire contains albums that appeal more to an older and less computer-friendly population. Zentner (2005a) finds some evidence that file sharing may be behind the recent increase in the share of domestic repertoire in total sales.

Using a panel of weekly album sales and information on the weekly number of downloads by album for the United States, Oberholzer-Gee and Strumpf (2005) find that music downloading has an effect on sales that is statistically indistinguishable from zero. To establish causality, they employ technical features related to file sharing as instruments. The high-frequency nature of their identification strategy—across-weeks variation within album—and the durability of music may restrict the interpretation of the results to the lack of contemporaneous substitution.<sup>4</sup>

Rob and Waldfogel (2004) collected individual-level data on album downloads and purchases by 500 college students. They find evidence that each album download reduces music purchases by .2 albums in the OLS specification, and they find a much bigger effect when using instrumental variables. They also address causality issues by using the speed of the Internet connection as an instrument. In addition to individual-level self-reported broadband access, they employ broadband access variation at the school level as an alternative instrument. They also collected information on consumers' valuation of these albums, which allows them to document that downloaded albums are valued less than purchased albums. The restriction to 500 college students in four schools limits generalizations of their results.

### III. THE MUSIC INDUSTRY

Global music sales in 2002 totaled US\$32.2 billion. Forty-one percent of these sales were made in North America, 34.5 percent in Europe, 18.6 percent in Asia—with Japan representing more than 80 percent of Asian sales—3.1 percent in Latin America, and 2.7 percent distributed among Australasia, the Middle East, and Africa. Sales are concentrated in the top markets. The top five countries—the United States, Japan, the United Kingdom, France, and Germany—represent 76.5 percent of global sales, and the top 10—the top five above plus Canada, Italy, Spain, Australia, and Mexico—represent 85 percent (IFPI 2002b).

Sales are also concentrated among a few companies. The four biggest companies—Universal, Sony-BMG, EMI, and Warner—control more than 70 percent of the global market of music sales, with the rest of the market share distributed among many independent record labels. These latter labels,

<sup>4</sup> To address this concern they present a distributed lag model. See Liebowitz (2004) for a critique of the instruments and Rob and Waldfogel (2004) for a more detailed analysis of the identification strategy.

in some cases, have an important presence in an individual country, region, or continent.

Companies and musicians usually negotiate exclusive multiyear contracts. When producing a new album, artists typically receive an up-front payment and a royalty somewhere between 5 and 13 percent of the retail price of the record (Standard & Poor's 2002).

The CD is the most popular music format, representing 72 percent of total international units sales. Sales of singles, LPs, and cassettes continue to be replaced by sales of CD albums (IFPI 2002b). Two new formats, DVD Audio and Super Audio CD, are growing but do not yet have an important share. These new formats have higher sound quality and extra content such as video clips and interviews with the artists. The tendency to include extra content might have been accelerated by the need to differentiate the product from the illegal online substitute.

There is variability in prices across CDs. While the average retail price of a CD in the United States is US\$14.19 (Black 2003a), 28 of the top 50 albums on the *Billboard* charts have a list price between US\$17.98 and US\$19.98, and only seven are listed at US\$14.98 or less (Morris 2002). There is little information on music prices for other countries. In 2001, the European Commission opened an investigation to study the higher prices in Europe and the divergence in prices inside the European Union.<sup>5</sup> In 2003, the average price of a CD in the United Kingdom was US\$16.80 (British Phonographic Industry 2004).

Distribution costs of music represent a very important share of total costs. A CD with a suggested price to consumers of US\$16.98 has a price of US\$10.50 to the retailer (Standard & Poor's 2002). This latter figure includes distribution costs from the record company to the retailer.

The distribution channels have been changing. Music stores have been shrinking as a source of sales and are being replaced by supermarkets, discount stores, department stores, and online retail. In the United States, music stores' share of sales fell from 62 percent in 1991 to 42 percent in 2000 (RIAA 2000). In the United Kingdom, supermarkets increased their share of music sales from 11.2 percent in 1999 to 17.7 percent in 2001 (IFPI 2002b). Online retail (off-line delivery), as a share of total sales, increased from 6 percent in 2001 to 9 percent in 2002 in Germany and from 4 percent to 6 percent in the United Kingdom and remained steady at 3 percent in the United States. Online legitimate delivery became available in 2001, but it is still not an important source of sales. In 2003, sales of music downloads in

<sup>5</sup> It was found that the major labels were including minimum advertised prices linked to cooperative advertising agreements in Germany (Tannock 2003). A lawsuit alleging the same practices was filed in the United States. The case ended with a settlement in which the companies agreed to refund US\$67.4 million to consumers and discontinue their minimum pricing policies (Deutsche 2002).

the United States totaled US\$30 million, which represents .25 percent of total music sales (IFPI 2004b).

Online legitimate delivery and file sharing are possible in MP3 format. The MP3 format is a way to compress audio data without significantly compromising sound quality.

Sound recordings are originally represented as waves. When the sounds are digitalized, these waves are sampled many times per second and a file is created. Compact disc quality needs a sampling of 44,100 times per second (44.1 KHz). Humans can hear only around 10 percent of the sounds that are recorded on a CD. The MP3 compression system eliminates sounds that are not perceptible to humans and softer sounds when different sounds are playing simultaneously. There are different qualities of MP3 compression (depending on the bit rate of the file). Bit rates between 64 and 192 Kbps are standard on the Internet, but only files above 160 Kbps have quality comparable to CDs.

A 5-minute soundtrack that would take more than 50 MB in CD compression format can be reduced to a file of 5 MB without significantly affecting the sound quality. A compressed file of this size can be downloaded in as fast as 12 minutes with a dial-up Internet connection (56 Kbps modem downstream), 1 minute and 20 seconds with a regular DSL or cable connection (512 Kbps downstream), and 20 seconds with a fast DSL connection (2,000 Kbps downstream). However, the actual downloading speed also depends on the upstream speed connection of the computer providing the file, and the upstream speed is usually lower than the downstream speed.<sup>6</sup>

People can upload (rip) CDs to their PCs' hard drives and listen to music on their computer, compress the files to an MP3 format to reduce the storage memory requirement and to facilitate the sharing of the files over the Internet, and convert the files back to a CD format and burn CDs that can be played in a regular player. The development of these technologies could represent a very significant reduction in costs considering that around 50 percent of music costs are distribution expenses.

There are many alternatives for getting music online. They can be divided into two groups: legitimate and illegitimate (under the current law).

Legitimate companies either own the copyright or make deals with copyright owners to distribute their music. Among the biggest companies are iTunes, Napster, Rhapsody, and the Europe-based OD2. Most services offer

<sup>6</sup> Having a high-downstream-speed connection allow users to download many files at the same time. In addition, in many peer-to-peer services, such as Kazaa, it is possible to download the same file from different users. The program recognizes the same file in different computers, partitions the file in several pieces, and allows downloading of each piece from a different peer. These two ways help circumvent the bottleneck created by a lower upstream speed than downstream speed. Universities have T3 connections that are much faster than residential service (from 3 Mbps to 45 Mbps downstream).

unlimited “streaming audio” and “tethered downloads”<sup>7</sup> for a fixed charge per month of around US\$10 and the possibility of permanent burnable downloads for around US\$1 per song. In 2003, prices of digital tracks in Europe were higher, between US\$1.76 and US\$2.35 per song (Koranteng 2003), but they have decreased in the past 2 years.

Online legitimate digital delivery has been overshadowed by piracy. After the shutdown of Napster, illegitimate music online can still be found globally on peer-to-peer file-sharing services such as Kazaa, BitTorrent, and several others. These services are distributed without charge and allow users to download both licensed and unlicensed files, including music, movies, games, and software. The amount of music available through these services is larger than on any legitimate site.

Since the second half of 2003, the RIAA has been suing individual users who are offering substantial amounts of copyrighted music over peer-to-peer networks. The industry claims that these lawsuits are behind the leveling off of the decline in U.S. music sales in 2004.

#### IV. THE DATA

This paper uses a European consumer mail survey by Forrester Research from October 2001 called Consumer Technographics. Forrester is a business research company specializing in the information economy. The fieldwork was conducted by the market research company Taylor Nelson Sofrés. The sampling methodology is proprietary but is meant to ensure a representative sample in each country. The survey contains more than 70 questions about many different topics, with many subitems and multiple answers. Analogous U.S. data from Forrester have been used extensively in the economic literature related to the Internet (Goolsbee 2000, 2001; Goolsbee and Brown 2002; Goolsbee and Klenow 2002).

The survey includes 22,488 observations and is designed to be representative of the total 16 and older population in seven European countries: France, Germany, Italy, Netherlands, Spain, Sweden, and the United Kingdom.<sup>8</sup> Three of these countries are among the top five music markets, and five are among the top 10. In 2001, they represented 27.8 percent of international music sales (IFPI 2002b).

The database contains a discrete  $\{0, 1\}$  variable indicating purchases of

<sup>7</sup> Streaming audio gives the opportunity to listen to soundtracks without downloading them to the hard drive. Tethered downloads are downloads to the hard drive that cannot be burned into a CD. The music is no longer available when subscription is ended.

<sup>8</sup> I found modest differences in age and gender when comparing these data with country-level censuses. A total of 7,355 individuals have missing values for at least one of the variables. Income is the only missing variable for 2,625 individuals. The results below are similar if these individuals are included in the regressions and income is not included as a control. English proficiency is the only missing variable for 1,083 individuals. The results are similar when including these individuals and not using this variable as a control.

TABLE 1  
SUMMARY STATISTICS

Statistic	Value
Overall population ( $N = 15,133$ ) (%):	
Internet users	50.7
People who regularly download MP3 files	9.0
People who bought music during the last month	39.3
People who bought music during the last month given that they regularly download MP3 files	55.8
People who bought music during the last month given that they do not regularly download MP3 files	37.7
People with DSL or cable Internet connection	4.3
Internet connection at home ( $N = 5,917$ ) (%):	
People who regularly download MP3 files	21.0
People who bought music during the last month	47.1
People who bought music during the last month given that they regularly download MP3 files	55.1
People who bought music during the last month given that they do not regularly download MP3 files	45.0
People with DSL or cable Internet connection	11.0
Probability of buying music:	
France	.33
Germany	.48
Italy	.34
Netherlands	.30
Spain	.27
Sweden	.37
United Kingdom	.53

music—CDs, tapes, or records—during the month prior to the survey for each respondent. This question is broad in the sense that it is not restricted to off-line purchases. For each respondent it also contains information about access to the Internet; purchases of many goods during the last month, including videos, books, software, and groceries; ownership of many electronic goods, including portable stereos, hi-fi stereos, cellular phones, DVD players, MP3 players, CD writers, and game consoles; and demographic variables such as gender, age, work status, education, household size, and household income.

For Internet users, the database contains information on the weekly average number of hours spent online, the number of years that they have been going online and using e-mail, and information about their Internet activity, including checking e-mail, using search engines, purchasing goods online, publishing their own Web pages, participating in online auctions, and downloading MP3 files. For people with an Internet connection at home, the database contains information on the type of connection—DSL, cable, ISDN, or dial-up.

Table 1 presents summary statistics. Across the overall sample (15,133 observations after dropping individuals with missing values), 39.3 percent

bought music during the month prior to the interview, 9.0 percent regularly download MP3 files,<sup>9</sup> and 50.7 percent have Internet access.

Downloading's current illegal status may induce a negative bias in the fraction of people who report downloading music. However, the industry strategy of suing individuals—rather than companies, for contributory and vicarious liability—who are illegally sharing music is recent, and the risk of individuals being prosecuted or fined was probably negligible in October 2001.<sup>10</sup> Only in the middle of 2004 did the industry start suing European individuals (RIAA 2003).

The percentage of people who bought music is much larger among the group who regularly download MP3 files (55.8 percent) than among those who do not (37.7 percent), which suggests that MP3 downloaders have a strong taste for music. Considering only those people who have an Internet connection at home (5,917 observations), 47.1 percent bought music during the month prior to the interview and 21.0 percent regularly download MP3s. Again, the fraction of people who bought music is higher among those who regularly download music (55.1 percent) than among those who do not regularly download music (45.0 percent).

Table 2 compares Internet users and broadband access by country in Forrester's sample with data on the number of Internet users by the International Telecommunication Union (ITU). The ITU does not collect the data by itself, but it compiles information from different country-level surveys.<sup>11</sup> Forrester's data show a higher number of Internet users and broadband users than the data by the ITU.

There is heterogeneity in per capita music sales among the countries considered. According to the IFPI (2002b), in 2001 per capita music sales (CDs,

<sup>9</sup> There are two questions about music downloading in the survey: regularly "download music (MP3)" and regularly do "file sharing (e.g., Napster)." I report the results of a variable by assigning a value of 1 to doing any of these activities. These two variables have a high correlation. The results below are similar using any of these two variables separately.

<sup>10</sup> Because of the design of the questionnaire, it is not possible to know whether some individuals refused to answer if they regularly download music. File sharing is an item in a question with more than 40 items, and individuals are asked to mark all the statements that apply. Regularly "download music (MP3)" and regularly do "file sharing (e.g., Napster)" are two of these statements. An important characteristic of the questionnaire is that the subitems including the question asking if they bought music in the month prior to the survey and the subitem including the question asking if they regularly download music are in different sections. While the first question is asked of the whole sample, the second is asked only of individuals with access to the Internet. There are many questions about completely different topics in between them. This may be important because individuals might strategically misreport their actions if they are supporters of file sharing or are against it. By answering both questions together, they could predict that the results are going to be used to evaluate the effect of file sharing on music sales. This may be a problem for analyses based on surveys that concentrate only on music purchases and downloads.

<sup>11</sup> The date of the year in which the data were collected in each country and the broadness of the measure of Internet access—access at home or overall access—are not known. International Telecommunication Union data are from their Web site (<http://www.itu.int/ITU-D/ict/statistics>).

TABLE 2  
INTERNET PENETRATION (%)

	FORRESTER (October 2001)			INTERNATIONAL TELECOMMUNICATION UNION		
	Access at Home	Overall Access	Broadband	Internet 2001	Internet 2002	Broadband 2002
France	28.6 (.753)	39.2 (.813)	2.3 (.254)	26.3	31.3	2.4
Germany	43.8 (.853)	55.6 (.854)	4.8 (.368)	37.3	42.3	3.9
Italy	40.1 (1.119)	53.1 (1.139)	1.4 (.269)	26.8	30.1	1.5
Netherlands	58.8 (1.335)	66.7 (1.279)	15.2 (.975)	49.0	53.0	7.2
Spain	18.1 (.902)	33.2 (1.102)	1.8 (.311)	18.2	19.3	1.1 <sup>a</sup>
Sweden	58.3 (2.449)	67.0 (2.336)	8.1 (1.357)	51.6	57.3	7.8
United Kingdom	47.8 (.972)	59.9 (.954)	3.8 (.375)	32.9	40.6	2.3

NOTE.—Standard errors are listed in parentheses.

<sup>a</sup>Broadband access for 2001, not 2002.

cassettes, singles, and LPs) per year were 4.79 units in the United Kingdom, 3.53 in Sweden, 3.16 in Germany, 2.89 in France, 2.31 in the Netherlands, 2.02 in Spain, and .85 in Italy. Average per capita unit sales per year over the countries considered in the sample were 2.83. The microdata do not contain sales quantities to match to the IFPI data.<sup>12</sup> Table 1 also reports probabilities of music purchases by country.

## V. EMPIRICAL APPROACH

The goal of this paper is to estimate the effect of peer-to-peer usage on the probability of purchasing music. It was shown in the last section that a comparison of means indicates that individuals who regularly download MP3s are more likely to buy music. This positive relationship persists—although it is not statistically different from zero—when controlling for several individual-level characteristics.

I consider the following model:

$$B_i = b_1 D_i + b_2 \mathbf{X}_i + \varepsilon_i,$$

where for individual  $i$ ,  $B_i$  and  $D_i$  are discrete  $\{0, 1\}$  variables indicating that individual  $i$  bought music during the month prior to the survey and regularly downloads MP3 files, respectively,  $\mathbf{X}_i$  is a vector of observed individual characteristics, and  $\varepsilon_i$  represents the error. Column 1 of Table 3 presents OLS estimates for respondents with and without Internet access. Feasible generalized least squares is employed to allow for heteroskedasticity across in-

<sup>12</sup>The microdata probably include purchases of illegitimate copies of music (commercial piracy). Seasonality in music sales and the fact that Forrester does not include individuals 15 years old and younger make it difficult to compare probabilities computed from the microdata with data on quantities from the IFPI. To my knowledge, information on country-level seasonality is not available.

TABLE 3  
ORDINARY LEAST SQUARES (OLS) AND INSTRUMENTAL VARIABLES:  
PROBABILITY OF PURCHASING MUSIC

	OLS	INSTRUMENTAL VARIABLES			
	(1)	(2)	(3)	(4)	(5)
Download MP3 files	.0135 (.0153)	-.4982* (.2485)	-.3195* (.1246)	-.3271* (.1288)	-.3843* (.1736)
Age	-.0033** (.0003)	-.0035** (.0003)	-.0034** (.0003)	-.0037** (.0007)	-.0044** (.0010)
Log of income	-.0030 (.0070)	-.0064 (.0074)	-.0052 (.0071)	-.0075 (.0132)	-.0035 (.0196)
No Internet access	.0134 (.0103)	.0065 (.0111)	.0089 (.0105)	N.A.	N.A.
Listen to music while online	.0667** (.0115)	.1391** (.0371)	.1138** (.0209)	.1218** (.0207)	.1181** (.0260)
Household size	-.0143** (.0038)	-.0127** (.0040)	-.0133** (.0039)	-.0094 (.0062)	-.0186* (.0091)
Confidence in English	.0076* (.0037)	.0102* (.0040)	.0093* (.0038)	.0148* (.0064)	.0116 (.0094)
Male	.0496** (.0086)	.0633** (.0111)	.0586** (.0093)	.0975** (.0181)	.1211** (.0265)
Hours of TV	-.0003 (.0004)	-.0002 (.0004)	-.0003 (.0004)	.0000 (.0008)	-.0003 (.0011)
Hours of Internet	-.0024** (.0008)	.0040 (.0032)	.0018 (.0018)	.0020 (.0017)	.0026 (.0021)
Student	.0273 (.0182)	.0701* (.0282)	.0552** (.0213)	.0352 (.0291)	.0371 (.0405)
Work full-time	.0165 <sup>+</sup> (.0095)	.0052 (.0113)	.0091 (.0100)	.0056 (.0158)	-.0112 (.0232)
Education	-.0072 (.0053)	-.0111 <sup>+</sup> (.0058)	-.0097 <sup>+</sup> (.0055)	-.0057 (.0094)	.0057 (.0135)
Own Walkman	.0773** (.0087)	.0801** (.0090)	.0791** (.0088)	.0783** (.0144)	.0718** (.0203)
Own MP3 player	.0215 (.0204)	.0510 <sup>+</sup> (.0262)	.0407 <sup>+</sup> (.0221)	.0202 (.0330)	.0273 (.0455)
Own hi-fi stereo	.0665** (.0092)	.0614** (.0096)	.0632** (.0093)	.0284 (.0212)	.0223 (.0318)
Own CD writer	-.0013 (.0117)	.0674 <sup>+</sup> (.0356)	.0434* (.0205)	.0487 <sup>+</sup> (.0253)	.0643 <sup>+</sup> (.0360)
Overidentifying restrictions test ( <i>P</i> -value)	N.A.	N.A.	.102	.453	.271
<i>N</i>	15,133	15,133	15,133	5,917	3,257

NOTE.—The dependent variable is a dummy variable for buying music during the last month. Column 2: download MP3 files is instrumented (instruments: have a DSL or cable Internet connection). Column 3: download MP3 files is instrumented (instruments: publish own Web page, participate in online auctions, ask for technical support online, read computer magazines, how long using Internet, and how long using e-mail). Column 4: same instruments as column 3 and the regression is restricted to Internet users, Column 5: same instruments as column 3 and the regression is restricted to Internet users with more than 2 years going online. Other controls (dummies) included for all regressions are region (53 regions within countries), ownership of goods (TV, widescreen TV, dish, cable TV, set-top box for TV, VCR, PC, digital assistant, handheld game, wireless application protocol phone, camcorder, printer, digital camera, digital camcorder, Web camera, game console, DVD drive, DVD player, scanner, and mobile phone), purchases of goods (books, videos, video games, software, toys, sporting goods, clothing, footwear, jewelry, gifts, flowers, event tickets, electronics, groceries, beer, wine, or spirits, tobacco, health and beauty), and read magazines (women, family, home, cooking, travel, health, sports, motor, men, TV, celebrity, news, business, finance, and nature). Robust standard errors are in parentheses.

<sup>+</sup> Significant at the 10% level.

\* Significant at the 5% level.

\*\* Significant at the 1% level.

dividuals. Household income was standardized to gross annual euros per year, and its logarithm was included in the regression. Age is measured in years. Confidence in English is measured from 1 to 5, with 5 being “very confident.” Education takes values from 1 to 3 for lower, middle, and upper levels. Household size has a cap of five or more people in the household. Hours of television and hours of Internet (only for Internet users) are the average hours per week and take the values 2, 7, 12, 17, 22, 27, and 30 (with a cap of 30 or more hours). All the other controls are dummies.

Note that prices of music are not included in the analysis. As the coefficient is identified in the cross section, the relevant variation in prices is the variation across regions. In a concentrated music market, it is possible that there exists some price variation across regions. To my knowledge, regional price information is not available. The regressions include 53 within-country regional dummies that may account for this possibility.<sup>13</sup>

The downloaded online product can theoretically be a complement for music in regular formats (CDs, cassettes, LPs) as the means comparison and the OLS regression suggest.<sup>14</sup> Consumers may get to know albums by downloading some tracks from the Internet and then decide whether or not to buy the album. Downloading takes time, but this cost could be lower than the cost of getting to know music by going to the store and listening to different albums before purchasing. Also, not having an MP3 player, not being able to burn CDs in order to listen to the downloaded copy on a regular stereo system, or some quality difference between the copy and the original may induce the purchase of the album.

However, radios advertise music, and the technology is available to burn good-quality CDs from MP3 files in order to listen to the downloaded copy on a regular player. The presence of unobserved heterogeneity in music tastes across individuals may be an obstacle to the identification of the causal effect. The approach of this paper is to use instrumental variables techniques. To

<sup>13</sup> Despite many missing observations, for each individual the survey contains brackets for the size of the population in the city (not available for Germany). The brackets are different for every country and do not coincide with the within-country regions. For example, for Italy the brackets are as follows: fewer than 10,000, 10,000–30,000, 30,000–100,000, and more than 100,000 individuals. In a few cases a specific city is listed. For the listed cities, I imputed the real population. For the other cities, I imputed the median value of the range of the bracket. For the top bracket I tried different values. The magnitude of the effect of downloading in the regressions below is similar when I include the size of the city as an additional control and measures of Internet sophistication are used as instruments. The coefficient is estimated more imprecisely but is also similar in magnitude when including size as a control and instrumenting with the speed of the connection. The coefficient and standard error in this latter case are similar when including or excluding the size of the city as a control, which suggests that the increased imprecision is a consequence of the missing observations.

<sup>14</sup> It should be noted that this is valid for both legal and illegal downloads. A related question to the one analyzed in this paper—that to my knowledge has not been addressed in the literature—is whether legal downloads reduce sales of music in traditional format.

achieve the goal requires an instrumental variable that explains peer-to-peer systems usage but is otherwise unrelated to music purchases.

To account for the presence of unobserved heterogeneity, I consider the following model:

$$B_i = b_1 D_i + b_2 X_i + \theta_i + v_i,$$

where, for individual  $i$ ,  $\theta_i$  represents unobserved taste for music and  $v_i$  represents the error. If  $\theta_i$  and  $D_i$  are not orthogonal, OLS gives a biased and inconsistent estimate of  $b_1$ . One way to overcome this problem is to find an instrumental variable  $Z_i$  that explains  $D_i$  but is orthogonal to  $\theta_i$ .<sup>15</sup>

The first instrument that I consider is access to a broadband connection.<sup>16</sup> The idea is that it is considerably faster to download a music file with a high-speed connection, which reduces the cost in time of downloading music.

Column 1 of Table 4 shows that having a broadband connection significantly increases the probability of downloading music. The first stage shows who downloads music. Young individuals,<sup>17</sup> students, males, owners of MP3 players, owners of CD writers, and individuals who listen to music while online are more likely to be downloaders.

The second stage presented in column 2 of Table 3 suggests a reduction of 50 percent in the probability of buying music. Purchases of music are negatively correlated with age and household size and positively correlated with listening to music while online,<sup>18</sup> being male, being a student, the level of confidence in English, owning CD writer and an MP3 player, and owning complements of music in traditional format, such as a Walkman and a hi-fi stereo.

A problem with this instrument is that it is a choice variable, and it is

<sup>15</sup> Estimation of models of limited dependent variables (LDVs) with endogenous continuous regressors has been discussed extensively in the literature (Newey 1987; Rivers and Vuong 1988). However, here the endogenous regressor is a dummy variable. Angrist (2001) shows different ways of dealing with dichotomous endogenous regressors in models of LDVs. Using linear instrumental variables with a local average treatment effect interpretation is one of the proposed alternatives. Kane, Rouse, and Staiger (1999) show that instrumenting a dummy regressor measured with error biases the coefficient away from zero (see note 10).

<sup>16</sup> Having a broadband connection might be correlated with a high downloading intensity. If there was misreporting (see note 10), I do not see a reason to believe that the unobserved downloading intensity would be correlated with this hypothetical misreporting in a particular way. Only recently has the RIAA started suing U.S. individuals who share large amounts of music online.

<sup>17</sup> Note that the age variable is not significant. Including the variable age implicitly assumes that the marginal effect computed at different ages is the same. The substitution of the variable age for dummy variables by years of age does not change the result of the second-stage regressions.

<sup>18</sup> This is not necessarily streaming audio by the context in which the question is asked. The question is "do you ever do any of the following while online?" One of the options is "listen to music/radio," and there are many other options, such as "have the TV on," "have friends over," and "read magazines."

TABLE 4  
FIRST-STEP ORDINARY LEAST SQUARES REGRESSIONS FOR THE PROBABILITY OF  
DOWNLOADING MUSIC ONLINE

	(1)	(2)
Have broadband connection	.0847** (.0174)	N.A.
Publish own Web page	N.A.	.0636** (.0224)
Read computer magazines	N.A.	.0683** (.0111)
Participate in online auctions	N.A.	.0523** (.0203)
Ask for technical support online	N.A.	.0475** (.0143)
How long using the Internet	N.A.	.0056 (.0049)
How long using e-mail	N.A.	-.0022 (.0047)
Age	-.0003 (.0002)	-.0003 (.0002)
Log of income	-.0059 (.0041)	-.0038 (.0041)
No Internet access	-.0125** (.0045)	-.0071 (.0051)
Listen to music while online	.1413** (.0093)	.1355** (.0093)
Household size	.0028 (.0024)	.0040 <sup>+</sup> (.0023)
Confidence in English	.0053* (.0021)	.0031 (.0021)
Male	.0266** (.0047)	.0186** (.0047)
Hours of TV	.0003 (.0002)	.0003 (.0002)
Hours of Internet	.0117** (.0007)	.0103** (.0008)
Student	.0835** (.0134)	.0845** (.0133)
Work full-time	-.0212** (.0052)	-.0223** (.0052)
Education	-.0070* (.0030)	-.0088** (.0029)
Own Walkman	.0051 (.0046)	.0059 (.0045)
Own MP3 player	.0564** (.0156)	.0559** (.0154)
Own hi-fi stereo	-.0101* (.0040)	-.0090* (.0040)
Own CD writer	.1311** (.0089)	.1236** (.0089)
$R^2$	.30	.31
$F$ -test (global test of excluded in the second-stage instruments)	N.A.	15.3 ( $P = .000$ )

NOTE.—The dependent variable is a dummy variable for regularly download MP3 files (mean among overall population = .09). Other controls (dummies) included for all regressions are region (53 regions within countries), ownership of goods (TV, widescreen TV, dish, cable TV, set-top box for TV, VCR, PC, digital assistant, handheld game, wireless application protocol phone, camcorder, printer, digital camera, digital camcorder, Web camera, game console, DVD drive, DVD player, scanner, and mobile phone), purchases of goods (books, videos, video games, software, toys, sporting goods, clothing, footwear, jewelry, gifts, flowers, event tickets, electronics, groceries, beer, wine, or spirits, tobacco, health and beauty), and read magazines (women, family, home, cooking, travel, health, sports, motor, men, TV, celebrity, news, business, finance, and nature). Robust standard errors are in parentheses.  $N = 15,133$ .

<sup>+</sup> Significant at the 10% level.

\* Significant at the 5% level.

\*\* Significant at the 1% level.

probably related to the interest in downloading music.<sup>19</sup> However, if people who have broadband connections have a strong taste for music—and subscribing to a broadband connection in order to download music may signal

<sup>19</sup> A better instrument would be variation in broadband availability at the regional level. Regions are within countries, and there are a total of 53 regions in the sample. However, to my knowledge, data on European broadband availability at the within-country regional level is not available. Moreover, the database is not precise about which cities are considered in each region to match broadband availability at the regional level. For example, for the United Kingdom the regions are North, Midlands, and South.

a strong taste for music—the bias goes in the direction of underestimating the effect of downloads on the probability of buying music. In fact, in 2001 many broadband service providers explicitly advertised the ability to download music quickly. On the other hand, if—for a given taste for music—people who buy broadband connections are people who are going to substitute more, then the bias goes in the direction of overestimating the average reduction in the probability of buying.

Measures of Internet sophistication could be employed as alternative instruments. The presumption is that downloading MP3 files requires a high degree of Internet sophistication. In the database there is information about many uses of the Internet and information on the number of years individuals have been using the Internet and e-mail. There is also a dummy variable for individuals who read computer magazines. A large number of years using the Internet and e-mail and some uses of the Internet such as “publish own Web pages,” “participate in online auctions,” and “ask for technical support online” and “read computer magazines” might signal a high degree of Internet sophistication. These variables could increase the probability of downloading music but might be otherwise unrelated to an unobserved taste for music or entertainment.

Column 2 of Table 4 reports the first-step regression of “download MP3” on the proposed instruments. As noted in Table 1, the percentage of people who download MP3 files across the overall population is 9 percent. The first stage shows that sophisticated Internet users have a higher probability of downloading music online. The probability value of an *F*-test of the global significance of the excluded in the second-stage variables is reported at the bottom of Table 4. Table 4 also shows that being a student, being male, listening to music while online, and the ownership of complements of music in digital format—MP3 player and CD writer—increase the probability of downloading music.

In column 3 of Table 3, “download MP3,” is instrumented with the proposed variables. The coefficient on “download MP3” is negative and significant. It indicates that downloading MP3 files reduces the probability of buying music by 32 percent.

In the case of a dummy independent variable, a dummy treatment, and a dummy instrument, instrumental variables has a local average treatment effect interpretation (Angrist 2001). When using the speed of the connection as an instrument, the estimate measures the average treatment effect for those who would be induced to download if broadband is provided. Downloading intensity is not available in the database. Having a broadband connection may induce a high downloading intensity and thus generate a big reduction in the probability of buying music. This may explain a lower estimated effect when using measures of Internet sophistication as an instrument. The broadband instrument is interesting for forecast analysis if it is assumed that broadband will be ubiquitous in the near future.

The instruments “how long using the Internet” and “how long using e-mail” are not significant in the first step. Even having a high global significance of the instruments— $F$ -value of 14.7<sup>20</sup>—in the first-step regression, it may be better to exclude them from the analysis in order to avoid the risk of dealing with weak instruments. The results do not change much when these instruments are excluded from the analysis.

To further analyze the robustness of the results, columns 4 and 5 of Table 3 restrict the regressions to people with Internet access at home and to Internet users with more than 2 years of online experience, respectively. The regression with individuals with more than 2 years of Internet experience is performed to exclude people who acquired an Internet connection after Napster was introduced—to control for the possibility that people went online in order to download music.<sup>21</sup> The estimated effects of downloading music on the probability of buying music are similar when restricting to these groups.

Given that the number of instruments exceeds the number of endogenous variables, it is possible to test for overidentifying restrictions. The probability values for this test are reported in Table 3. The results support the exclusion of the instruments from the purchase equation. Given the large change in the estimated coefficient in the instrumented regressions in comparison with the OLS estimate, special scrutiny of the instruments is necessary.

To overstate the negative effect of downloads on music purchases, the instruments would have to be correlated with a weak taste for music. Another possibility is that the instruments were correlated with an unobserved low value of time. Downloading music from an illegitimate site is free of charge. However, there are other costs associated with the process of downloading. An important one is the cost of time involved in this process. The risk is that sophisticated Internet users are people with a low value of time and therefore—for a given taste for music—more willing to substitute music purchases for freely downloaded music.

Table 5 presents OLS regressions of goods that may be related to a taste for music or entertainment on the instruments. Sophisticated Internet users may have a tendency to use complements of music in digital (nontraditional) formats—MP3 players and CD writers—and substitute away from complements of music in regular formats—Walkmans and hi-fi stereos. Sophisticated Internet users may also have a tendency to use the computer to watch videos or to play games, which induces a negative correlation between the instru-

<sup>20</sup> A rule of thumb for being concerned about weak instruments in the case of a single endogenous regressor is to have an  $F$ -value lower than 10 (Staiger and Stock 1997).

<sup>21</sup> Napster was launched 2 years and 4 months before the fielding of the survey. The coefficient is similar but the standard error is bigger when restricting to Internet users with more than 3 years and with more than 4 years of Internet experience.

ments and having a DVD player or a game console.<sup>22</sup> The mean values of having these goods at home are reported in the last row of Table 5.

Publishing a Web page is positively correlated with having an MP3 player and a CD writer but is negatively correlated with owning a Walkman and a hi-fi stereo.<sup>23</sup> As suggested before, this may be explained by a negative effect of music downloading on purchases of complements of music in regular formats. Excluding “own a Web page” from the list of instruments results in a similar estimated impact of downloading MP3 on the probability of purchasing music.

Table 5 also reports an OLS regression of the number of hours that respondents watch television, an activity that may be thought to be negatively correlated with the value of time, on the instruments. The number of hours of television watched per week correlates negatively with “publish own Web pages” and “ask for technical support online,” which suggests that Internet-sophisticated people do not place an especially low value on time. However, this analysis may not be valid if using the Internet and watching television are substitutes.

To further analyze the exogeneity of the instruments, I study how purchases of goods that may signal a taste for entertainment are affected by MP3 file downloading. Table 6 presents instrumental variables regressions of purchases of books, videos and DVDs,<sup>24</sup> video games, and audiovisual electronics. If the instruments were selecting individuals with weak taste for entertainment or individuals with low probability of buying entertainment goods, a negative impact of the predicted values of MP3 file downloading on purchases of these goods would be expected. None of the regressions show that downloading music reduces the probability of buying other entertainment goods. This suggests that the negative effect of downloading music on the probability

<sup>22</sup> The regressions control for “have a CD writer” and “have an MP3 player.” Owning an MP3 player may be thought to reduce music purchases by allowing downloaders to listen to the downloaded files away from the computer. Also, as MP3 files can be converted back to a CD format, burning CDs may be expected to reduce music purchases too. Unfortunately, the survey does not ask people if they regularly burn CDs. The survey does ask about the ownership of CD writers. The regressions in Table 3 show a positive effect of owning MP3 players and CD writers on the probability of buying music. However, these variables may also be correlated with the unobserved heterogeneity in tastes for music. Among music downloaders, having a CD writer reduces the probability of buying music by 5.6 percent and having an MP3 player reduces the probability of buying by 2 percent.

<sup>23</sup> The regressions in Table 3 include complements of music as controls. However, these decisions might be determined jointly. An interpretation of this negative correlation with complements of music in regular format and positive correlation with complements of music in digital format is that downloading affects not only sales of music but also the whole regular-format music market.

<sup>24</sup> Sharing films on the Internet has some importance now but was nearly nonexistent in 2001. Sharing films requires fast connections because the files are big (around 600 MB). Another important difference between downloading songs and movies is that repeated play of movies does not seem to be as usual as repeated play of records. This characteristic may explain the popularity of rental of movies but not of records.

TABLE 5  
CORRELATION OF INSTRUMENTS WITH THE PROBABILITY OF HAVING GOODS THAT SIGNAL TASTE FOR  
ENTERTAINMENT AND WITH THE VALUE OF TIME (Ordinary Least Squares)

	Walkman	Hi-Fi Stereo	MP3 Player	CD Writer	DVD	Game Console	Hours of TV
Publish own Web page	-.0422 <sup>+</sup> (.0220)	-.0296* (.0147)	.0243 <sup>+</sup> (.0141)	.0720** (.0216)	.0070 (.0182)	.0144 (.0211)	-.7659 <sup>+</sup> (.3999)
Read computer magazines	-.0022 (.0126)	-.0080 (.0085)	.0031 (.0067)	.0813** (.0128)	.0140 (.0096)	-.0006 (.0119)	.1758 (.2390)
Participate in online auctions	-.0044 (.0152)	.0115 (.0091)	-.0060 (.0080)	.0799** (.0154)	.0199 (.0124)	-.0023 (.0144)	-.5017 <sup>+</sup> (.2789)
Ask for technical support online	.0142 (.0196)	-.0127 (.0115)	-.0005 (.0120)	.0056 (.0200)	.0175 (.0175)	.0131 (.0191)	.2300 (.3764)
How long using the Internet	-.0039 (.0062)	-.0084 <sup>+</sup> (.0045)	.0056 <sup>+</sup> (.0030)	.0033 (.0057)	.0022 (.0043)	-.0018 (.0056)	.1298 (.1113)
How long using e-mail	.0075 (.0060)	.0060 (.0042)	-.0052 <sup>+</sup> (.0027)	.0029 (.0055)	-.0020 (.0041)	-.0119* (.0054)	-.3118** (.1064)
Age	-.0042** (.0003)	-.0039** (.0003)	.0002 (.0001)	-.0003 (.0002)	-.0005** (.0002)	-.0035** (.0003)	.0442** (.0064)
Log of income	-.0088 (.0070)	.0324** (.0061)	.0036 (.0032)	.0137** (.0051)	.0118** (.0043)	-.0236** (.0059)	-.5196** (.1426)
No Internet access	-.0168 (.0110)	-.0206* (.0086)	.0107* (.0048)	-.0103 (.0079)	-.0050 (.0067)	.0100 (.0095)	.4591* (.2056)
Listen to music while online	.0427** (.0107)	-.0012 (.0069)	.0023 (.0051)	.0396** (.0104)	.0150 <sup>+</sup> (.0079)	-.0022 (.0101)	-.3070 (.2001)

Household size	.0349** (.0038)	-.0014 (.0030)	.0025 (.0017)	.0010 (.0030)	-.0018 (.0025)	.0552** (.0035)	-.5282** (.0730)
Confidence in English	.0213** (.0037)	.0042 (.0030)	-.0029 <sup>+</sup> (.0016)	.0046 (.0028)	-.0021 (.0023)	-.0056 <sup>+</sup> (.0031)	-.1273 <sup>+</sup> (.0714)
Male	-.0386** (.0086)	.0288** (.0075)	-.0042 (.0037)	-.0050 (.0064)	-.0096 <sup>+</sup> (.0052)	.0008 (.0070)	.7057** (.1728)
Hours of TV	.0019** (.0004)	.0000 (.0004)	-.0002 (.0002)	-.0004 (.0003)	.0005 <sup>+</sup> (.0003)	.0015** (.0004)	N.A.
Hours of Internet	-.0016* (.0008)	-.0003 (.0005)	.0012* (.0005)	.0044** (.0008)	-.0007 (.0006)	-.0006 (.0008)	.1136** (.0149)
Student	.0749** (.0161)	-.0387** (.0121)	.0099 (.0083)	-.0118 (.0150)	-.0549** (.0111)	-.0481** (.0165)	-.8211* (.3236)
Work full-time	.0389** (.0094)	.0141 <sup>+</sup> (.0074)	-.0023 (.0040)	-.0109 (.0073)	.0005 (.0060)	.0197* (.0082)	-1.4893** (.1804)
Education	.0151** (.0052)	.0126** (.0045)	-.0014 (.0023)	.0006 (.0039)	-.0042 (.0033)	-.0229** (.0044)	-.4453** (.1022)
Mean of dependent variable	.51	.79	.03	.19	.08	.27	16.09

NOTE.—Other controls (dummies) included for all regressions are region (53 regions within countries), ownership of goods (TV, widescreen TV, dish, cable TV, set-top box for TV, VCR, PC, digital assistant, handheld game, wireless application protocol phone, camcorder, printer, digital camera, digital camcorder, Web camera, game console, DVD drive, DVD player, scanner, and mobile phone), purchases of goods (books, videos, video games, software, toys, sporting goods, clothing, footwear, jewelry, gifts, flowers, event tickets, electronics, groceries, beer, wine, or spirits, tobacco, health and beauty), and read magazines (women, family, home, cooking, travel, health, sports, motor, men, TV, celebrity, news, business, finance, and nature). Robust standard errors are in parentheses.  $N = 15,133$ .

<sup>+</sup> Significant at the 10% level.

\* Significant at the 5% level.

\*\* Significant at the 1% level.

TABLE 6  
 PROBABILITY OF PURCHASING GOODS THAT SIGNAL TASTE FOR ENTERTAINMENT: INSTRUMENTAL VARIABLES

	Books	Videos and DVDs	Video Games	Audiovisual Electronics
Download MP3 files	.0759 (.1181)	.1724 <sup>+</sup> (.1021)	.3681** (.1025)	.2190* (.0964)
Purchase music	.1343** (.0087)	.1071** (.0067)	.0341** (.0056)	.0285** (.0055)
Age	.0027** (.0003)	-.0015** (.0003)	-.0009** (.0002)	.0004 <sup>+</sup> (.0002)
Log of income	.0216** (.0076)	-.0057 (.0055)	-.0152** (.0047)	.0054 (.0045)
No Internet access	-.0177 (.0110)	.0167* (.0080)	.0052 (.0067)	.0060 (.0067)
Listen to music while online	-.0034 (.0198)	-.0254 (.0168)	-.0552** (.0171)	-.0350* (.0157)
Household size	-.0020 (.0040)	-.0016 (.0032)	.0061* (.0028)	-.0003 (.0027)
Confidence in English	.0253** (.0039)	.0029 (.0028)	-.0023 (.0023)	-.0037 (.0024)
Male	-.0400** (.0098)	.0143* (.0069)	.0111 <sup>+</sup> (.0059)	.0115 <sup>+</sup> (.0061)
Hours of TV	-.0029** (.0004)	.0009** (.0003)	.0008** (.0003)	.0003 (.0003)
Hours of Internet	.0016 (.0017)	-.0000 (.0014)	-.0029* (.0014)	-.0003 (.0014)
Student	-.0042 (.0210)	-.0489** (.0165)	-.0548** (.0158)	-.0045 (.0152)
Work full-time	-.0024 (.0101)	-.0015 (.0077)	-.0109 <sup>+</sup> (.0066)	.0002 (.0066)
Education	.0609** (.0057)	.0021 (.0042)	.0013 (.0036)	.0009 (.0035)
Own Walkman	.0217* (.0090)	.0062 (.0064)	-.0065 (.0054)	.0134* (.0054)
Own MP3 player	.0310 (.0204)	-.0109 (.0184)	-.0106 (.0160)	.0105 (.0165)
Own hi-fi stereo	.0164 (.0108)	-.0067 (.0065)	-.0152** (.0052)	-.0017 (.0057)
Own CD writer	-.0301 (.0197)	-.0450** (.0164)	-.0463** (.0160)	-.0299* (.0151)

NOTE.—Download MP3 files is instrumented (instruments: publish own Web page, participate in online auctions, ask for technical support online, read computer magazines, how long using Internet, and how long using email). Other controls (dummies) included for all regressions are region (53 regions within countries), ownership of goods (TV, widescreen TV, dish, cable TV, set-top box for TV, VCR, PC, digital assistant, handheld game, wireless application protocol phone, camcorder, printer, digital camera, digital camcorder, Web camera, game console, DVD drive, DVD player, scanner, and mobile phone), purchases of goods (books, videos, video games, software, toys, sporting goods, clothing, footwear, jewelry, gifts, flowers, event tickets, electronics, groceries, beer, wine, or spirits, tobacco, health and beauty), and read magazines (women, family, home, cooking, travel, health, sports, motor, men, TV, celebrity, news, business, finance, and nature). Robust standard errors are listed in parentheses.  $N = 15,133$ .

<sup>+</sup> Significant at the 10% level.

\* Significant at the 5% level.

\*\* Significant at the 1% level.

of buying music is not driven by instruments that pick individuals with a weak taste for entertainment goods.

One alternative explanation for the recent drop in music sales is that prices of substitutes (DVDs and video games) have been dropping. The inclusion of prices of DVDs and video games in the analysis is essential if time variation is the source of identification. Note that Table 6 shows that purchasing music is positively correlated with buying DVDs and games. The positive effect of downloading on purchases of other entertainment goods may indicate that downloading free music induces purchases of alternative entertainment goods.

File sharing is an activity concentrated in the younger population, which makes it difficult to estimate the effect of downloading for different age groups. Fifteen percent of the population 40 years old and younger downloads music. On the other hand, the percentage of downloaders among the population aged 40 and older is only 4 percent. When separating the analysis for individuals younger than 40 from the 40-and-older population, the coefficient is similar to the results presented so far for the younger group. For the older population, the coefficient is smaller but estimated imprecisely.

As mentioned before, having a high-speed connection considerably reduces the cost in downloading time. This suggests that the effect of downloading on sales should be larger for downloaders with broadband. Despite the estimates presenting high standard errors, separating MP3 downloaders with and without broadband shows a big reduction in purchases among the first group.<sup>25</sup>

## VI. EFFECTS OF ONLINE MUSIC DOWNLOADING ON MUSIC SALES

Instrumenting with measures of Internet sophistication, the results in the last section indicate that downloading MP3 files online reduces the probability of buying music during the month prior to the survey by 30 percent. Downloaded music may be shared with people who do not download MP3 files and affect their purchases as well. In this event, the estimates understate the effect of online music downloading on music sales.

The database does not contain information on quantities of music purchased or on intensities of music downloads to calculate what music sales would have been in the absence of music downloading. Moreover, the probability of buying music and intensity of purchases may vary in complicated ways.

Another obstacle in calculating the impact on sales comes from the possibility of having downloaders incorrectly measured. Forrester's data overestimate the number of Internet users compared to the ITU (Table 2), which suggests that the number of downloaders may be overestimated. That said,

<sup>25</sup> See Zentner (2005b) for this analysis and other results not reported here.

and with the goal of having a back-of-the-envelope estimate of the predicted effect on sales, in this section I investigate this counterfactual.

Across the overall population, 9 percent regularly downloaded music in 2001.<sup>26</sup> Therefore, if both digital music users and nonusers had the same propensity to buy music, the effect on the music industry would be a reduction in music sales units of 2.7 percent.

But digital music users have a greater propensity to buy music, which indicates that a correction for the heterogeneity in the groups is needed. It was shown in Table 1 that the probability of buying music for people who download music is higher than the probability for nonusers of peer-to-peer systems. Also, using my estimate with a treatment on the treated interpretation, the proportion of music downloaders who would have bought music would have been around 30 percent higher if the possibility of downloading music did not exist. This suggests that the probability of buying music for downloaders would have been around double the probability of buying for nondownloaders.<sup>27</sup>

As there are no data on quantities of music purchased, an assumption about the number of units bought is needed. A probably conservative assumption is that users and nonusers of peer-to-peer systems, when reporting that they bought music during the month prior to the survey, bought the same quantity of units. With this assumption, the estimated impact on units sold—at the 2001 level of file sharers—would be a reduction of 4.9 percent.<sup>28</sup> If users of peer-to-peer systems not only have greater propensity to buy music but also would have bought more units conditional on buying, this would be an underestimate of the impact.

In October 2001 Napster was over and Kazaa was not yet an option. In 2002, the number of users increased from 3 million to 5 million (IFPI 2003). Assuming that this magnitude is representative of the increase in the number of downloaders in the European countries considered, and that these new downloaders have the same probability of buying as people who were downloaders in October of 2001, sales in 2002 would have been around 7.8 percent higher.

In summary, if music downloading reduces the probability of buying by 30 percent, if 15 percent of the population downloads music, if downloaders are twice as likely to buy music than nondownloaders, and if—conditional on buying—downloaders and nondownloaders buy the same quantity of units, then sales in 2002 would have been 7.8 percent  $(.3 \times [(15 \times 2)/115])$  higher.

<sup>26</sup> Legal online digital delivery was nearly nonexistent in 2001.

<sup>27</sup>  $(55.8 + 30)/37.7$ .

<sup>28</sup>  $[(.09 \times 2)/109] \times (-0.3)$ .

## VII. CONCLUSION

Global music sales have experienced a large drop during the last 4 years. This paper uses a European individual-level database to measure the impact of online music downloads on the probability of buying music and finds that file sharing reduces this probability. Using measures of Internet sophistication as instruments, downloads may explain a 30 percent reduction in the probability of buying music.

The estimates in this paper are an important component of any welfare analysis of file sharing or copyright. The interest is not exclusive to the music industry. Other digital copyrighted goods such as movies, software, games, and books are also being shared online. The development of fast connections is likely to increase the impact of file sharing on sales of these goods. This is going to become an increasingly important issue in the next few years.

Downloading copyrighted material is illegal under the current legal system. The music industry is fighting file sharing in court.<sup>29</sup> In the United States, music piracy has been legally fought on the basis of contributory and vicarious liability (Landes and Lichtman 2003). Under these doctrines, copyright holders sue parties that in some way contribute to or benefit from the infringing conduct, instead of suing individuals. However, the new peer-to-peer systems have proven to be more difficult to fight legally because they do not require a central server to operate (Varian 2000) and have alternative legitimate uses. The other difficulty is that many of these new systems are established in countries with different legal systems: Kazaa is registered in the South Pacific island nation of Vanuatu, the software distributor is in Australia, and the servers are in the Netherlands (Associated Press 2003).

After running into difficulties shutting down file-sharing systems, the RIAA changed strategy and has been “gathering evidence and preparing lawsuits against individual computer users who are illegally offering large amounts of copyrighted music over peer-to-peer networks” (RIAA 2003). In Europe, the industry started suing individuals only in the middle of 2004.

There is controversy about the effect of this strategy on the number of downloads. While suing individuals who offer music—as opposed to individuals who download files—may reduce the number of files available to download, it is not clear whether this would actually affect the number of downloads. This is important when considering the public-good nature of the files offered online. In addition, while it appears that the number of users has decreased for some popular sites such as Kazaa, the legal strategy appears to have induced individuals to use alternative and less popular sites and forums where the risk of being prosecuted may be lower. The music industry

<sup>29</sup> Many countries impose taxes on blank CDs and burners. Another way to combat piracy is to flood the net with decoy files (Lichtman and Jacobson 2000).

claims that the recovery of sales that occurred in the United States in 2004 may be explained by the success of the new legal strategy.

## REFERENCES

- Angrist, Joshua. 2001. "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice." *Journal of Business and Economic Statistics* 19:2–16.
- Ante, Spencer E. 2000. "Inside Napster." *BusinessWeek*, August 14. [http://www.businessweek.com/2000/00\\_33/b3694002.htm](http://www.businessweek.com/2000/00_33/b3694002.htm).
- Associated Press. 2000. "Napster Usage Quadruples Despite Legal Uncertainty." September 11. <http://archives.cnn.com/2000/TECH/computing/09/11/napster.traffic.ap/index.html>.
- . 2003. "Kazaa Sues Movie, Recording Industries." *Chicago Tribune*, January 2.
- Black, Jane. 2003a. "Big Music's Broken Record." *BusinessWeek*, February 13. [http://www.businessweek.com/technology/content/feb2003/tc20030213\\_9095\\_tc078.htm?campaign\\_id=search](http://www.businessweek.com/technology/content/feb2003/tc20030213_9095_tc078.htm?campaign_id=search).
- . 2003b. "Play Taps for Music Pirates." *BusinessWeek*, June 26. [http://www.businessweek.com/technology/content/jun2003/tc20030626\\_3034\\_tc078.htm](http://www.businessweek.com/technology/content/jun2003/tc20030626_3034_tc078.htm).
- Boldrin, Michele, and David Levine. 2002. "The Case against Intellectual Property." *American Economic Review: Papers and Proceedings* 92: 209–12.
- British Phonographic Industry. 2003. *BPI Statistical Handbook*. London: BPI.
- Deutsche, Claudia H. 2002. "Suit Settled over Pricing of Music CD's at 3 Chains." *New York Times*, October 1.
- Fader, Peter. 2000. "Expert Report." *A&M Records, Inc. v. Napster, Inc.* 114 F. Supp. 2d 896.
- Goolsbee, Austan. 2000. "In a World without Borders: The Impact of Taxes on Internet Commerce." *Quarterly Journal of Economics* 115:561–76.
- . 2001. "Competition in the Computer Industry: Online versus Retail." *Journal of Industrial Economics* 49:487–99.
- Goolsbee, Austan, and Jeffrey Brown. 2002. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy* 110:481–507.
- Goolsbee, Austan, and Peter Klenow. 2002. "Evidence on Learning and Network Externalities in the Diffusion of Home Computers." *Journal of Law and Economics* 45:317–44.
- Hui, Kai-Lung, and Ivan Png. 2003. "Piracy and the Legitimate Demand for Recorded Music." *Contributions to Economic Analysis and Policy*, vol. 2, art. 11, pp. 1–22. <http://www.bepress.com/bejeap/contributions/vol2/iss1/art11/>.

- IFPI (International Federation of the Phonographic Industry). 2002a. "Music Piracy Report." London: IFPI.
- . 2002b. "The Recording Industry in Numbers." London: IFPI.
- . 2003. "Fighting Piracy and Growing a Legal Online Business." <http://72.14.203.104/search?q=cache:w-3-iUQ0TG8J:www.ifpi.org/site-content/press/inthedia09.html%22&hl=en&gl=us&Ct=clnk&cd=2>.
- . 2004a. "The Recording Industry in Numbers." London: IFPI.
- . 2004b. "The Recording Industry World Sales." London: IFPI.
- Kane, Thomas J., Cecilia Elena Rouse, and Douglas Staiger. 1999. "Estimating Returns to Schooling When Schooling Is Misreported." Working Paper No. 7235. National Bureau of Economic Research, Cambridge, Mass.
- Klein, Benjamin, Andres Lerner, and Kevin Murphy. 2002. "The Economics of Copyright 'Fair Use' in a Networked World." *American Economic Review: Papers and Proceedings* 92:205–8.
- Koranteng, Juliana. 2003. "Value-Added Tax Poses Problems for EU Digital Sales." *Billboard*, June 28.
- Landes, William, and Douglas Lichtman. 2003. "Indirect Liability for Copyright Infringement: Napster and Beyond." *Journal of Economic Perspectives* 17:113–24.
- Lichtman, Douglas, and David Jacobson. 2000. "Anonymity a Double-Edged Sword for Pirates Online." *Chicago Tribune*, April 13.
- Liebowitz, Stan. 2003. "Will MP3 Downloads Annihilate the Record Industry? The Evidence So Far." Working paper. University of Texas at Dallas, School of Management, June.
- . 2004. "Peer-to-Peer Networks: Creative Destruction or Just Plain Destruction?" Working paper. University of Texas at Dallas, School of Management, July.
- . 2005. "Pitfalls in Measuring the Impact of File Sharing." Working paper. University of Texas at Dallas, School of Management, July.
- Morris, Chris. 2002. "Declaration of Independents." *Billboard*, June 8.
- Newey, Whitney. 1987. "Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables." *Journal of Econometrics* 36:231–50.
- Oberholzer-Gee, Felix, and Koleman Strumpf. 2005. "The Effect of File Sharing on Record Sales. An Empirical Analysis." Working paper. University of North Carolina, Chapel Hill.
- RIAA (Recording Industry Association of America). 2000. "The Recording Industry Association of America's 2000 Consumer Profile." <http://www.riaa.com/news/marketingdata/pdf/2000consumerprofile.pdf>.
- . 2003. "Recording Industry to Begin Collecting Evidence and Preparing Lawsuits against File 'Sharers' Who Illegally Offer Music Online." June 25. [http://www.riaa.com/news/newsletter/062503\\_a.asp](http://www.riaa.com/news/newsletter/062503_a.asp).
- Rivers, Douglas, and Quang Vuong. 1988. "Limited Information Estimators

- and Exogeneity Tests for Simultaneous Probit Models.” *Journal of Econometrics* 39:347–66.
- Rob, Rafael, and Joel Waldfogel. 2004. “Piracy on the High C’s: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students.” Unpublished manuscript. University of Pennsylvania, Wharton School, Philadelphia.
- Romer, Paul. 2002. “When Should We Use Intellectual Property Rights?” *American Economic Review: Papers and Proceedings* 92:213–16.
- Rowan, David. 2002. “Inside the Web of Thieves.” *Observer*, April 21. <http://www.guardian.co.uk/internetnews/story/0,,688022,00.html>.
- Staiger, Douglas, and James Stock. 1997. “Instrumental Variable Regression with Weak Instruments.” *Econometrica* 65:557–86.
- Standard & Poor’s. 2002. *Movies and Home Entertainment. Industry Survey*. New York: Standard & Poor’s.
- Tannock, Charles. 2003. Letter to the Commission. Subject: CD Prices in Europe. E-2232/02. *Official Journal of the European Union*, June 12.
- Varian, Hal. 2000. “The Internet Carries Profound Implications for Providers of Information.” *New York Times*, July 27.
- Zentner, Alejandro. 2005a. “File Sharing and International Sales of Copyrighted Music. An Empirical Analysis with a Panel of Countries.” *Topics in Economic Analysis and Policy* 5, art. 21, pp. 1–17. <http://www.bepress.com/bejeap/topics/vol5/iss1/art21>.
- . 2005b. “Measuring the Effect of Music Downloads on Music Purchases.” Ph.D. dissertation. University of Chicago, Department of Economics.