



Working Papers in Economics

Working Paper EC-05-01

**Pirates or Explorers? Analysis of Music
Consumption in French Graduate Schools**

David Bounie
Marc Bourreau
Patrick Waelbroeck

June 2005

Department EGSHE ■ Telecom Paris ■ 46 rue Barrault ■ 75013 Paris ■ France
Email: wpecon@telecom-paris.fr

Pirates or Explorers?

Analysis of Music Consumption in French Graduate Schools^{*}

David Bounie, *ENST*

Marc Bourreau, *ENST and CREST-LEI*

Patrick Waelbroeck, *ECARES and FNRS*

June 2005

Abstract

This paper analyzes the impact of music file sharing on CD purchases. Traditionally, two arguments are opposed concerning the impact of music file sharing on CD sales. On the one hand, MP3 downloads only reduces sales of legitimate CDs (the "competition effect"). On the other hand, consumption of free MP3s could lead people to buy CD which they would never have bought otherwise (the "sampling effect"). Because the court in the Napster case and some academic researchers have dismissed this sampling effect, this article seeks to assess whether sampling does indeed occur and in the affirmative what are the relative contributions of the positive sampling and the negative competition effects of MP3 files on CD purchases. To do so, we administered an anonymous online survey in two French graduate schools from May 26 to June 3, 2004. We find that there are two populations: explorers and pirates. For the explorers, MP3 sharing through physical contacts has a positive impact on CD consumption. For the pirates, anonymous file-sharing on the internet and a large number of MP3 files have a negative effect on CD consumption.

JEL Classification: L82, L86

Keywords: Internet, Peer-to-Peer, File sharing, Piracy, Music, Students

* Corresponding author: Patrick Waelbroeck, ECARES/ULB, CP 114, Université Libre de Bruxelles, 50 av. Roosevelt, 1050 Brussels, Belgium; pwaelbro@ulb.ac.be. We would like to thank Michel Gensollen, Joel Waldfogel, Bernard Salanié and seminar participants at CREST-LEI, FORUM - Université Paris 10 and ENST for helpful suggestions. Amandine Ferré provided valuable research assistance.

1 Introduction

Lawsuits against music uploaders on file-sharing or Peer-to-Peer (P2P) networks have made headline news. These lawsuits are motivated by the perception that music shared on P2P networks substantially harms record companies and even threatens the existence of the music industry. The vision of the record companies that digital copies available on P2P networks only displace sales of legitimate CDs has also led the industry to increase technological protection, such as new Digital Rights Management protection, and lobbied for increased copyright protection. This argument is related to the “competition effect” according to which digital copies that offer a similar quality at zero cost directly and unfairly compete with original CDs.

While there has been some evidence that music downloads have decreased music sales in the early days of MP3 downloads first from Napster and then from Kazaa, this vision is shortsighted. First, other factors have played an important role in the downturns in CD sales during the period 2000-2003 (see Peitz and Waelbroeck, 2004a; PW hereafter). Secondly, although MP3 could be downloaded quite fast with the first two generations of P2P networks (Napster and Kazaa), MP3 files take substantially more time to download from recent file-sharing networks such as eMule and bitTorrent. Indeed, priority rules limit the speed at which a file can be downloaded by an internet user as a function of the number and size of files that he shares/uploads. So, there is at least an opportunity cost of downloading music files. Third, CDs and MP3 files do not offer the same value to music fans who appreciate uncompressed sound and booklets with lyrics and pictures. Finally and most importantly, digital copies can provide information on the value of an album. This is an important feature of music for which consumers have different tastes and for which there is uncertainty on the genres, style and quality of new releases that represent more than half of yearly sales. This so-called “sampling” or “matching” property of digital copies can potentially lead an internet user to purchase a CD that he or she would never have purchased without this information.

Because this sampling effect has been sometimes dismissed, for instance during the Napster case, this article seeks to assess whether sampling does indeed occur and in the affirmative what are the relative contributions of the positive sampling and the negative competition effects of MP3 files on CD purchases. For this purpose, we analyze the factors that influence individual CD consumption, including variables related to MP3 uses.

We administered surveys in two French "Grandes Ecoles" (graduate schools). College and Graduate school students are a good group to focus on for several reasons. First, college students are technophiles who are more likely to possess a broadband connection and to have downloaded and shared files on the Internet (see surveys documented in PW). Secondly, college students have a strong taste for music and frequently go to live concerts. Finally, college students have an important discretionary income that they spend on entertainment goods, including music. We find that there are two populations: "explorers" and "pirates". For the "explorers," MP3 sharing through physical contacts has a positive impact on CD consumption. For the "pirates," anonymous file-sharing on the internet and a large number of MP3 files have a negative effect on CD consumption.

There are several good surveys of the empirical studies that analyze the effect of internet piracy on music sales (see for instance PW or Liebowitz, 2004). Most of these studies use aggregate data and proxy variables for internet piracy or music downloads that make results difficult to interpret. Aggregated data are subject to much more criticisms than individual data, as argued by Liebowitz (2004). The two closest articles to our research are Zentner (2004) and Rob and Waldfogel (2004). Zentner (2004) uses individual survey data from October 2001 in large European countries. After controlling for unobserved heterogeneity in music taste, Zentner finds that music downloads reduce the probability to purchase music by 30%. Assuming that people who download music purchase as much as people who do not, Zentner finds that internet piracy could have decreased CD sales in unit by 7% in the countries considered. Rob and Waldfogel (2004) use a survey of college students to determine which albums have been downloaded most. Using a list of

hit albums and a list of albums acquired by the respondents during the past year, they explain variation in individual CD consumption by the number of albums downloaded from the corresponding list. They find a statistically negative effect of downloaded albums on purchased albums. Next they use answers to valuation questions to determine if students download high- or low-value albums. Data suggest that depreciation and the nature of music as an experience good can explain the difference and the correlation between ex-ante and ex-post valuations and that students download low-valuation albums.

Our research contributes to the existing literature on several points. First, we have several variables that characterize sampling, taste for music and the channels of MP3 acquisitions, including file-sharing on the internet but also file exchanged over intranets as well as direct physical exchanges through USB keys and CD-Rs. These channels will turn out to be important in the econometric analysis. Secondly, our survey provides direct evidence on the number of files downloaded rather than an indicator of downloading (Zentner, 2004) or the number of albums downloaded (RW, 2004). Thirdly, we have direct measures of the extent of sampling among our respondents that we use to interpret our results and we also have information to distinguish between different behaviors related to MP3 uses.

The article is organized as follows. First, we describe the dataset and we assess the importance of sampling activities among survey respondents. Next, we determine the effect of MP3 consumption on CD purchases and discuss the implications on aggregate CD sales. Some conclusions follow.

2 Data

We administered an anonymous online survey in two French graduate schools from May 26 to June 3, 2004. The first school is located in a large metropolitan area, while the second one is located in a campus near a much smaller metropolitan area. We include in our analysis a binary variable that equals 1 for the campus school and 0 for the school located in the larger metropolitan area. A web site was set up to host the survey, which we directly coded in HTML. We then requested by mailing lists all undergraduate and graduate students as well as all administrative staff and professors to fill in the online survey. The questionnaire included thirty six closed questions divided in six main parts: information on the respondent (school, sex and age), Internet access, music consumption, MP3 exchanges, MP3 uses and opinion on downloading. At the end of the questionnaire, the respondent was asked to submit the form (by pressing a button). In doing so, a script automatically gathered the answers and sent them to us by email. Overall, 589 people answered the questionnaire, i.e. a 30% response rate. We use answers of 456 of students (Graduates and PhDs) who have acquired MP3 files in our analysis.¹ Summary statistics are reported in Appendix 1.

2.1 Demographics

The respondents are mainly male (90%). 93% are enrolled in an undergraduate or graduate program and 7% in a PhD program. 85% of the students are less than 24. 53% of students study in the school located in the large metropolitan area. While we do not directly control for income effects, most of the students come from the same social background. Differences in willingness to pay for music will be captured by the number of CDs in their music collection (see section 2.3).

¹ We removed the administrative staff and professors to have a more homogenous sample. Professors typically do not download music from the internet and have a much higher income.

2.2 Internet connection

Almost all the respondents have Internet access at home (89%) and most of them (85%) have a broadband connection. 88% of the respondents spend at least 5 hour per week online and 28% spend more than 30 hours per week.

2.3 Taste for music

The respondents have a strong taste for music as 83% listen to more than 5 hours of music per week and 53% of the respondents listen to music more than 10 hours per week. Moreover, among the music-related activities that the respondents do on a regular basis, 44% visit a record store to get information or to buy music, 36% go to live concerts and 34% play an instrument.

As expected, people in the survey purchase more CDs than in the average French population. More than 34% of the respondents purchase 5 CDs or more each year, with an average of about 4.8 per year. This contrasts with the average CD consumption of 2.6 per capita in France among adults of 20 or more.² However, 17% of the respondents claim they do not purchase any CD. Students between 25 and 29 purchase proportionally more "5 CDs or more" (see Table 1).

Table 1. CD purchases by age

Age	0	1-4	5-10	> 10	Total
	69	192	78	50	389
≤ 24	17.74	49.36	20.05	12.85	100
	87.34	86.88	84.78	78.13	85.31

² According to IFPI, 118 million CDs were sold in France in 2003, and there were 45 million adults of 20 or more according to INSEE (75% of the French population).

	9	23	13	12	57
25-29	15.79	40.35	22.81	21.05	100
	11.39	10.41	14.13	18.75	12.5
	1	6	1	2	10
≥ 30	10	60	10	20	100
	1.27	2.71	1.09	3.13	2.19
Total	79	221	92	64	456

Large volumes of annual CD purchases translate into large collections of music CDs. The average number of CDs owned is around 80. 26% of the respondents own more than 100 CDs (See Table 2).

Table 2. CD purchases by number of CDs in music collection

CD collection	0	1-4	5-10	> 10	Total
< 10	37	28	0	0	65
	56.92	43.08	0	0	100
	46.84	12.67	0	0	14.25
10-30	17	63	3	2	85
	20	74.12	3.53	2.35	100
	21.52	28.51	3.26	3.13	18.64
30-50	10	60	26	1	97
	10.31	61.86	26.8	1.03	100
	12.66	27.15	28.26	1.56	21.27
51-100	7	46	32	5	90
	7.78	51.11	35.56	5.56	100
	8.86	20.81	34.78	7.81	19.74
100-200	3	17	26	22	68
	4.41	25	38.24	32.35	100
	3.8	7.69	28.26	34.38	14.91
> 200	5	7	5	34	51
	9.8	13.73	9.8	66.67	100
	6.33	3.17	5.43	53.13	11.18
Total	79	221	92	64	456

2.4 Consumption of MP3s

2.4.1 Means of MP3 acquisition

In this study, we are interested in the effect of total MP3 consumption on music CD sales. This includes MP3 files downloaded from the Peer-to-Peer networks, from intranet networks, or physically exchanged, such as MP3 files on a CD-R burned by friends or family members. 97% of the respondents obtained free MP3 files by any of these means.³ Overall, 65% downloaded files from P2P networks, 75% downloaded files from intranet networks, and 53% obtained files by physical exchanges (CD-Rs, USB memory keys, etc.). These percentages have two implications. First, the large percentage of respondents acquiring MP3 files by other means than public file-sharing networks suggests that even if P2P networks were shot down, music file sharing would still prevail. Secondly, intranet and physical exchanges suggest the existence of online and offline communities. When we break down CD purchases by means of MP3 acquisition, the role of these communities seems to have two opposite effects: intranet sharing are associated with low CD consumption, while physical sharing with high CD consumption (see Table 3). On the one hand, people who exchange files physically share music through more personal relationships as opposed to anonymous file sharing on P2P networks. It is also important to stress that existing file-sharing software such as Kazaa or eDonkey are not well designed to explore new music since they do not include search by genres and do not recommend songs according to past downloading behavior. Internet users need to obtain this information from external sources such as internet sites, forums or friends. On the other hand, students exchanging files over an intranet build a huge collective music library from which any user can retrieve songs and albums. Therefore, the willingness to pay for prerecorded music of intranet users is relatively low. This effect seems to

³ The other 3% have legitimately transformed CDs that they possess into MP3 files.

dominate any positive informational effect related to proximity inside the student community.

Table 3. CD purchases and means of MP3 acquisition

	0	1-4	5-10	> 10	Total
Peer-to-peer	51	140	63	43	297
	<i>17.17</i>	<i>47.14</i>	<i>21.21</i>	<i>14.48</i>	<i>100</i>
	<i>64.56</i>	<i>63.35</i>	<i>68.48</i>	<i>67.19</i>	<i>65.13</i>
Intranet	64	168	65	45	342
	<i>18.71</i>	<i>49.12</i>	<i>19.01</i>	<i>13.16</i>	<i>100</i>
	<i>81.01</i>	<i>76.02</i>	<i>70.65</i>	<i>70.31</i>	<i>75</i>
Physical exchanges	32	111	53	47	243
	<i>13.17</i>	<i>45.68</i>	<i>21.81</i>	<i>19.34</i>	<i>100</i>
	<i>40.51</i>	<i>50.23</i>	<i>57.61</i>	<i>73.44</i>	<i>53.29</i>

2.4.2 Number of MP3 files

More than 60% of the respondents reported that they had more than 500 MP3 files on their computer. Table 4 indicates that younger students have proportionally more MP3s on their hard drives than older students.

Table 4. Number of MP3s by age

Age	1-100	101-200	201-500	> 500	Total
≤ 24	48	46	44	251	389
	<i>12.34</i>	<i>11.83</i>	<i>11.31</i>	<i>64.52</i>	<i>100</i>
	<i>75</i>	<i>82.14</i>	<i>80</i>	<i>89.32</i>	<i>85.31</i>
25-29	11	8	10	28	57
	<i>19.3</i>	<i>14.04</i>	<i>17.54</i>	<i>49.12</i>	<i>100</i>
	<i>17.19</i>	<i>14.29</i>	<i>18.18</i>	<i>9.96</i>	<i>12.5</i>

≥ 30	5 <i>50</i> <i>7.81</i>	2 <i>20</i> <i>3.57</i>	1 <i>10</i> <i>1.82</i>	2 <i>20</i> <i>0.71</i>	10 <i>100</i> <i>2.19</i>
Total	64	56	55	281	456

Breaking down the number of MP3s by the number of CDs in a student's collection (Table 5) reveals that students with less than 10 CDs in their collection and therefore a low willingness to pay for music have proportionally many MP3s (63% have more than 500 files). Assuming that 10 MP3 files correspond to an album on CD and that students who have more than 500 files have on average 750 files leads to a ratio of purchased albums to downloaded albums of $5/75= 1:15$ for those who have less than 10 CDs in their collection (assuming a mean of 5) and more than 500 MP3 files. This supports the argument of sales displacement by MP3 files, at least for that subpopulation. Students who have more than 200 CDs in the collection (assuming a mean of 300) and who have more than 500 MP3 files have a purchase to download ratio of $300/75 = 4:1$, which greatly contrasts with the previous population and suggests possible heterogeneity of the effect of MP3 consumption on CD purchases.

Table 5. Number of MP3 files and CD collection

CD collection	1-100	101-200	201-500	> 500	Total
< 10	12	4	8	41	65
	<i>18.46</i>	<i>6.15</i>	<i>12.31</i>	<i>63.08</i>	<i>100</i>
	<i>18.75</i>	<i>7.14</i>	<i>14.55</i>	<i>14.59</i>	<i>14.25</i>
10-30	17	11	18	39	85
	<i>20</i>	<i>12.94</i>	<i>21.18</i>	<i>45.88</i>	<i>100</i>
	<i>26.56</i>	<i>19.64</i>	<i>32.73</i>	<i>13.88</i>	<i>18.64</i>
30-50	14	17	13	53	97
	<i>14.43</i>	<i>17.53</i>	<i>13.4</i>	<i>54.64</i>	<i>100</i>
	<i>21.88</i>	<i>30.36</i>	<i>23.64</i>	<i>18.86</i>	<i>21.27</i>
51-100	9	10	8	63	90
	<i>10</i>	<i>11.11</i>	<i>8.89</i>	<i>70</i>	<i>100</i>
	<i>14.06</i>	<i>17.86</i>	<i>14.55</i>	<i>22.42</i>	<i>19.74</i>

100-200	7 <i>10.29</i> <i>10.94</i>	9 <i>13.24</i> <i>16.07</i>	6 <i>8.82</i> <i>10.91</i>	46 <i>67.65</i> <i>16.37</i>	68 <i>100</i> <i>14.91</i>
> 200	5 <i>9.8</i> <i>7.81</i>	5 <i>9.8</i> <i>8.93</i>	2 <i>3.92</i> <i>3.64</i>	39 <i>76.47</i> <i>13.88</i>	51 <i>100</i> <i>11.18</i>
Total	64	56	55	281	456

We can also tabulate annual CD purchases by the number of MP3 files on the computer. This is done in Table 6. We can nuance our previous finding. On the one hand, we again find that 67.1% of those who do not purchase any CDs have more than 500 MP3 files on their computer compared to 61.6% on average, which again suggests sales displacement. On the other hand, we also find 65.6% of those who purchase more than 10 CDs per year, there are proportionally more who also have a large stock of MP3 files. Thus Table 6 suggests that MP3 consumption can lead to an amplification of consumption pattern in the sense that there are consumers who download a large number of MP3 files and stop purchasing new CDs, while consumers with a high willingness to pay for music seem to listen to MP3 files for other reasons.

Table 6. CD purchases and MP3 files

Number of MP3s	0	1-4	5-10	> 10	Total
1-100	14 <i>21.88</i> <i>17.72</i>	28 <i>43.75</i> <i>12.67</i>	10 <i>15.63</i> <i>10.87</i>	12 <i>18.75</i> <i>18.75</i>	64 <i>100</i> <i>14.04</i>
101-200	2 <i>3.57</i> <i>2.53</i>	30 <i>53.57</i> <i>13.57</i>	18 <i>32.14</i> <i>19.57</i>	6 <i>10.71</i> <i>9.38</i>	56 <i>100</i> <i>12.28</i>
201-500	10 <i>18.18</i> <i>12.66</i>	34 <i>61.82</i> <i>15.38</i>	7 <i>12.73</i> <i>7.61</i>	4 <i>7.27</i> <i>6.25</i>	55 <i>100</i> <i>12.06</i>
> 500	53	129	57	42	281

	<i>18.86</i>	<i>45.91</i>	<i>20.28</i>	<i>14.95</i>	<i>100</i>
	<i>67.09</i>	<i>58.37</i>	<i>61.96</i>	<i>65.63</i>	<i>61.62</i>
Total	79	221	92	64	456

2.5 Sampling

In this section, we determine if there exists any positive aspects of music downloads. In particular, we assess the importance of sampling activities among survey respondents. We find that the large majority of MP3 users sample new music that they eventually purchase. Indeed, 93% of those who obtained free MP3s claimed that they have discovered new artists through listening to MP3s, and 67% reported that listening to digital music led them to purchase CDs that they would not have purchased otherwise. These numbers illustrate a strong “sampling effect” among the respondents of the survey. Both people with a large CD collection (high willingness to pay for music) and a large number of MP3 files are more likely to have discovered and purchased new music (Table 7).

Table 7. Did listening to MP3s lead you to purchase a CD that you would not have purchased otherwise?

CD stock	Sampling/purchase		Number of MP3s	Sampling/purchase	
	Yes	Total		Yes	Total
10-	25	65	1-100	38	64
	<i>38.46</i>	<i>100</i>		<i>59.38</i>	<i>100</i>
	<i>8.25</i>	<i>14.25</i>		<i>12.54</i>	<i>14.04</i>
10-30	55	85	101-200	38	56
	<i>64.71</i>	<i>100</i>		<i>67.86</i>	<i>100</i>
	<i>18.15</i>	<i>18.64</i>		<i>12.54</i>	<i>12.28</i>
30-50	62	97	201-500	35	55
	<i>63.92</i>	<i>100</i>		<i>63.64</i>	<i>100</i>
	<i>20.46</i>	<i>21.27</i>		<i>11.55</i>	<i>12.06</i>
51-100	61	90	> 500	192	281

	67.78	100	68.33	100
	20.13	19.74	63.37	61.62
100-200	57	68		
	83.82	100		
	18.81	14.91		
> 200	43	51		
	84.31	100		
	14.19	11.18		

2.6 Pirates and explorers

People who sample music do not necessarily purchase more CDs. Indeed, an internet user could discover and purchase one album but substitute four CDs by downloaded MP3 files. Peitz and Waelbroeck (2004b) point out two effects of file-sharing technologies on CD consumption. On the one hand, P2P networks improve the matching between products and buyers. They call this the matching effect. On the other hand, consumers receive a copy which, although it is an imperfect substitute to the original, may reduce their willingness-to-pay for the original. They call this the competition effect. Accordingly, we define an "explorer" as an internet user for whom the matching effect dominates the competition effect, and a "pirate" as an internet user for whom the competition effect dominates the matching effect.

As a starting point, we asked the following question to try and discriminate between these two types of students: "Since you started consuming free MP3 files, what is globally the effect of MP3 consumption on your purchases of CDs?" 52% said free downloading has not influenced CD purchases, 30% say it has decreased CD purchases and 18% say it has increased. Even though this question is highly subjective and answers to this question are probably biased, we can gain some information on the characteristics of the explorers and the pirates by cross-tabulating the effect of MP3 consumption by our set of exogenous variables. Tables 8 - 10 suggest that MP3 consumption has led to an amplification of

consumption patterns in the sense that the percentage of respondents saying that MP3 consumption has no influence decreases with internet time (Table 8), music taste (Table 9) and the number of MP3 files (Table 10).

Table 8. Effect of MP3 and Internet time

Internet time	Influence			Total
	Decrease	Same	Increased	
< 5	14	37	5	56
	<i>25</i>	<i>66.07</i>	<i>8.93</i>	<i>100</i>
	<i>10.29</i>	<i>15.55</i>	<i>6.1</i>	<i>12.28</i>
5-10	52	81	28	161
	<i>32.3</i>	<i>50.31</i>	<i>17.39</i>	<i>100</i>
	<i>38.24</i>	<i>34.03</i>	<i>34.15</i>	<i>35.31</i>
10-15	31	60	20	111
	<i>27.93</i>	<i>54.05</i>	<i>18.02</i>	<i>100</i>
	<i>22.79</i>	<i>25.21</i>	<i>24.39</i>	<i>24.34</i>
> 15	39	60	29	128
	<i>30.47</i>	<i>46.88</i>	<i>22.66</i>	<i>100</i>
	<i>28.68</i>	<i>25.21</i>	<i>35.37</i>	<i>28.07</i>

Table 9. Effect of MP3 and taste for music

Music time	Influence			Total
	Decrease	Same	Increased	
< 5	20	52	7	79
	<i>25.32</i>	<i>65.82</i>	<i>8.86</i>	<i>100</i>
	<i>14.71</i>	<i>21.85</i>	<i>8.54</i>	<i>17.32</i>
5-10	39	75	21	135
	<i>28.89</i>	<i>55.56</i>	<i>15.56</i>	<i>100</i>
	<i>28.68</i>	<i>31.51</i>	<i>25.61</i>	<i>29.61</i>
> 10	77	111	54	242
	<i>31.82</i>	<i>45.87</i>	<i>22.31</i>	<i>100</i>
	<i>56.62</i>	<i>46.64</i>	<i>65.85</i>	<i>53.07</i>

Table 10. Effect of MP3 and number of MP3 files

Number of MP3s	Influence			Total
	Decrease	Same	Increased	
1-100	10	43	11	64
	<i>15.63</i>	<i>67.19</i>	<i>17.19</i>	<i>100</i>
	<i>7.35</i>	<i>18.07</i>	<i>13.41</i>	<i>14.04</i>
101-200	18	29	9	56
	<i>32.14</i>	<i>51.79</i>	<i>16.07</i>	<i>100</i>
	<i>13.24</i>	<i>12.18</i>	<i>10.98</i>	<i>12.28</i>
201-500	13	37	5	55
	<i>23.64</i>	<i>67.27</i>	<i>9.09</i>	<i>100</i>
	<i>9.56</i>	<i>15.55</i>	<i>6.1</i>	<i>12.06</i>
> 500	95	129	57	281
	<i>33.81</i>	<i>45.91</i>	<i>20.28</i>	<i>100</i>
	<i>69.85</i>	<i>54.2</i>	<i>69.51</i>	<i>61.62</i>

In addition to being highly subjective, answers to the question related to the influence of MP3 files on CD consumption are endogenous to the number of CD purchased per year. To formally test difference in behavior between the "explorers" and the "pirates", we use a less subjective question that is exogenous to CD purchases: "Do you keep less than half of the MP3 files that you acquire on your hard drive?" Approximately 75% of those who have MP3 files on their computer keep more than half of their files. We use this variable to discriminate between "pirates" and "explorers". Indeed, consider first a "pirate", that is, a user whose main motivation for getting MP3 files is to constitute a digital music library at no cost. Then, obviously, this user will keep most of his MP3 files. On the other hand, consider an "explorer", who uses MP3 files so as to discover new artists or new music. After listening to new MP3 files (unknown ex ante), this user has incentives to delete the files he has not interested in because he is sure that he will never listen to these songs again in the future. In addition, deleting unwanted files can save time organizing files and directories. Indeed, finding and entering accurate information (song title, album name, date of production, composers, genre, subgenre, mood, etc) on all downloaded files and creating folders and subfolders can be very time-consuming. To make sure that people who delete files do not do so only because of a low hard disk capacity, we tabulate answers to the "delete" question according to the number of MP3 files and to the

availability of a CD burner which is a good indicator of whether the computer is recent and has a large storage space or not. First, if disk space was a serious concern, we would see people with large MP3 stocks delete proportionally more. The opposite phenomenon is observed in Table 11. Similarly, people with a computer without a CD burner only marginally delete less. These two observations suggest that disk storage capacity is not the primary reason why people delete music files. Anyway, 200 songs of 4 Megabits represent 800 Megabits, which is a small size for computers with CD burners (87% of the sample).

Table 11. Do you keep less than half of MP3 files you acquire?

Number of MP3s	Delete?			Delete?	
	Yes	Total		Yes	Total
1-100	26	64	No CDR	17	61
	40.63	100		27.87	100
	23.21	14.04		15.18	13.38
101-200	24	56	CDR	95	395
	42.86	100		24.05	100
	21.43	12.28		84.82	86.62
201-500	14	55			
	25.45	100			
	12.5	12.06			
> 500	48	281			
	17.08	100			
	42.86	61.62			

To check whether the "delete" variable and the "influence" variable are related, we tabulate answers to both questions in Table 12. We indeed observe that those who claim to have increased their CD purchases are one average more likely to delete MP3 files than those who said that they decreased their CD consumption after downloading MP3 files (35.15% against 15.44%), which justifies the use of this proxy variable to

distinguish between the two populations. Summary statistics for the "pirates" and the "explorers" are presented in the appendix.

Table 12. "Delete" behavior and MP3 influence

Influence	Delete?	
	Yes	Total
Decreased	21	136
	<i>15.44</i>	<i>100</i>
	<i>18.75</i>	<i>29.82</i>
Same	63	238
	<i>26.47</i>	<i>100</i>
	<i>56.25</i>	<i>52.19</i>
Increased	28	82
	<i>34.15</i>	<i>100</i>
	<i>25</i>	<i>17.98</i>
Total	112	456
	24.56	100

3 Estimation method and results

We determine the net contribution of Internet piracy on CD purchases by estimating the equation

$$y_i = x_i' \beta + w_i' \alpha + \varepsilon_i \quad (*)$$

where y_i is the number of CDs purchased per year for individual i ($= 1, \dots, n$), x_i are individual control variables (sex, age, school, internet time, and music CD collection), w_i are internet piracy variables (means of MP3 acquisition and number of MP3 files) and ε_i represents unobserved individual characteristics that we assume uncorrelated with the other variables in the equation. Parameters α and β need to be estimated.

We estimated equation (*) by ordinary least squares in section 3.1 and by applying an ordered probit model in section 3.2.⁴ The latter estimation method takes account of the fact that the dependent variable is discrete.

3.1 Ordinary least squares

We estimate equation (*) by using the means of MP3 acquisition (P2P, intranet, physical exchanges) as internet piracy variables, w_i , in section 3.1. We use the number of MP3 files in section 3.2.⁵

3.1.1 Means of MP3 acquisition

Estimation results are respectively given in columns (1) - (3) of Table 13. Coefficients in columns (1) are estimated on the full sample; coefficients in column (2) are estimated on the subpopulation of people who delete more than half of their music files, while coefficients in columns (3) are estimated on the complementary subpopulation. Most control variables have the expected signs. Taste for music measured by the time spent listening to music has a positive and significant effect on CD consumption. Offline information sources obtained by visiting record stores and reading music magazines also increase music consumption. Finally, the willingness to pay for music captured by the number of CDs in a music collection also increases music consumption.

An interesting result concerns the effect of live concerts on CD consumption. Overall, we find that people who regularly go to live concerts have a stronger demand for pre-recorded music (column 1). However, when we decompose the effect for pirates and

⁴ To apply ordinary least squares, we have constructed a continuous dependent variable that is equal to the mean of the category. If the respondent purchases 5-10 CD per year, the variable is set to 7.5. It was set to 15 for the category "10+" and to 2.5 for the category "1-4". Estimation by OLS does not require distributional assumption for the unobservable variable. The ordered probit model assumes normality.

⁵ We also used as specification in which we interacted the channels of MP3 acquisition and the number of MP3 files. Results were qualitatively identical to those presented in this section.

explorers, our estimation results indicate that live concerts also have a positive effect for the population of pirates (column 3), but a negative effect for the explorers (column 2). This finding needs to be confirmed by other studies, but it implies that people who save on the cost of acquiring pre-recorded music by downloading files from the internet (the pirates) can spend more money on live concerts, while live concerts more directly compete for the money of the explorers. Thus business models based on sharing revenues between CD sales and sales of ancillary products such as live concerts as proposed by Moreau and Curien (2005) could be profitable to some extent. Another interesting result concerns the variable "play instrument". The associated coefficient is negative in column (2) and positive in column (3), which suggests that there is a time constraint for the "explorer" that seems less relevant for the "pirates".

Downloading files from Peer-to-peer networks or from an intranet has a negative impact of CD purchases overall and on the subpopulation of pirates. The effect of sharing by physical exchanges has a positive effect on CD purchases overall and on the subpopulation of explorers. This result is partly compatible with an argument based on sharing of information goods developed by Bakos et al. (1999), which states that sharing in clubs can increase individual demand. The argument is not directly applicable in our context because MP3 files burned on a CD-R have a lower quality than a CD with uncompressed sound and CD case, pictures, lyrics, etc. However, more importantly for this article, the finding supports the view that physical exchanges of MP3 files also transmit meta-information on the music and genres that a community member will most likely enjoy. This seems to be the situation in which the sampling/exposure/discovery effect is the strongest.

Table 13. Estimation results by OLS using means of MP3 acquisition

	(1)		(2)		(3)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Female	0.3800	0.5252	-0.8670	1.0084	0.8841	0.6102
Age 20-29	0.7845	0.4890*	1.7427	1.1595	0.5011	0.5338
Age 30+	1.5235	1.1116	3.2550	1.7158*	0.4249	1.4312

School	0.1613	0.3214	0.4762	0.6678	0.1557	0.3641
Internet 6-15	0.4217	0.5186	1.0719	1.0542	0.1202	0.5978
Internet 15-30	0.2862	0.5544	0.6870	1.1945	0.2735	0.6258
Internet 30+	0.2048	0.5459	-0.6413	1.0735	0.4440	0.6338
Music 5-10	-0.1036	0.4866	-0.0786	0.9597	-0.3580	0.5609
Music 10+	0.8757	0.4633*	-0.3352	0.9210	0.9892	0.5374*
Record stores	2.8498	0.3568***	3.6142	0.6715***	2.6472	0.4200***
Music magazines	1.2705	0.5888**	2.2347	1.2926*	1.2039	0.6555*
Live concerts	0.8384	0.3738**	-1.5885	0.8035**	1.3998	0.4209**
Play instrument	0.3769	0.3466	-1.4163	0.7109**	0.6741	0.3917*
CDs 30-50	1.0395	0.4445**	1.9524	0.8960**	1.0939	0.5105**
CDs 50-100	1.6727	0.4664***	3.3638	1.0168***	1.5770	0.5261***
CDs 100-200	4.3505	0.5383***	5.3209	1.1181***	4.6769	0.6247***
CDs 200+	5.8891	0.6322***	8.9332	1.3010***	5.4772	0.7294***
Peer-to-peer	-0.5589	0.3390*	0.2871	0.7612	-1.0885	0.3891***
Intranet	-0.7811	0.3823**	-0.3725	0.7417	-0.8325	0.4373**
Physical exchanges	0.7047	0.3213**	1.4788	0.6555**	0.4328	0.3695
Constant	0.7662	0.6816	0.0874	1.2145	1.1346	0.8160
R-squared	0.5457		0.6409		0.5763	
Number of obs.	456		112		344	

We can test the equality of coefficients between the populations (2) and (3). Let $\theta_1 = (\beta_1', \alpha_1)'$ denote the vector of coefficients estimated on the population of explorers (column 2) and $\theta_0 = (\beta_0', \alpha_0)'$ denote the vector of coefficients estimated on the complementary subpopulation. We tested the hypothesis $H_0: \theta_0 = \theta_1$ against $H_1: \theta_0 \neq \theta_1$. The test statistics is 2.213 while the tabulated percentile is $F(20, 415, 2.213) > .99$. Thus we reject H_0 at the 1% confidence level. We also tested from the unrestricted model, the hypothesis $H_0': \alpha_0 = \alpha_1$ against $H_1': \alpha_0 \neq \alpha_1$. The test statistics is 2.15. The tabulated $F(3, 4215)$ value at the 10% level is 2.085. Thus we can reject H_0' at the 10% level.

3.1.2 Number of MP3 files

We also estimated equation (*) using the number of MP3 files to explain CD purchases. Results are reported in Table 14. Overall, the effect of MP3 files is negative for the full sample and on the population of pirates. It is not significantly different from zero for the population of explorers. We again tested $H_0: \theta_0 = \theta_1$ against $H_1: \theta_0 \neq \theta_1$. The F-statistics is evaluated at 2.09, which makes us reject H_0 at the 1% confidence level. We also tested

H_0' : $\alpha_0 = \alpha_1$ against H_1' : $\alpha_0 \neq \alpha_1$. The test statistics is 1.17, while the corresponding percentile is $F(3, 415, 1.17) = .68$. Thus we can not reject H_0 at the 10% level. This was expected since the effect of the number of MP3 files on CD consumption is insignificant for the explorers.

Table 14. Estimation results by OLS using number of MP3s

	0+1		0		1	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Female	0.1986	0.5383	-0.8504	1.0081	0.3738	0.6380
Age 20-29	0.8720	0.4810*	1.8871	1.1320*	0.5253	0.5280
Age 30+	1.1544	1.1127	3.3684	1.7475*	0.1535	1.4422
School	0.0938	0.3160	0.3668	0.6829	0.0587	0.3587
Internet 6-15	0.4799	0.5155	1.1133	1.0526	0.1481	0.5972
Internet 15-30	0.5528	0.5579	1.2294	1.2532	0.4188	0.6284
Internet 30+	0.3696	0.5434	-0.1854	1.0794	0.5422	0.6334
Music 5-10	-0.1073	0.4865	-0.1975	0.9634	-0.3180	0.5625
Music 10+	1.1046	0.4669**	-0.2958	0.9233	1.3240	0.5387**
Record stores	2.7047	0.3562***	3.4568	0.6906***	2.3190	0.4143***
Music magazines	1.4432	0.5879**	2.2696	1.3180*	1.3724	0.6555**
Live concerts	0.8577	0.3750**	-1.6287	0.8254**	1.4461	0.4244***
Play instrument	0.3726	0.3482	-1.2820	0.7058*	0.6534	0.3947*
CDs 30-50	1.1751	0.4417***	2.5190	0.9176	1.0421	0.5057**
CDs 50-100	1.9051	0.4694***	3.6513	1.0209***	1.7887	0.5321***
CDs 100-200	4.5395	0.5396***	5.7327	1.1345***	4.8734	0.6274***
CDs 200+	6.2305	0.6320***	9.2518	1.2879***	5.9363	0.7282***
MP3s 1-100	1.5073	0.4916***	1.1440	0.8871	1.6140	0.6077***
MP3s 101-200	1.1050	0.5113**	-0.1312	0.8652	1.6765	0.6658**
MP3s 201-500	0.5489	0.5169	-0.2526	1.0063	1.2023	0.6008*
Constant	-0.5268	0.6459	0.0624	1.2638	-0.6272	0.7483
R-squared	0.5471		0.6267		0.5769	
Number of obs.	456		112		344	

3.2 Ordered probit

To take into account the fact that the dependent variables takes only a finite number of values, we estimated equation (*) using an ordered probit model. Table 15 reports estimation results using the means of MP3 acquisition, Table 16 uses the number of MP3 files. Estimation results are qualitatively similar to those presented in the previous section. We again tested the hypothesis $H_0: \theta_0 = \theta_1$ against $H_1: \theta_0 \neq \theta_1$ using both specifications. The likelihood ratio test statistics are respectively 36.34 and 33.62, while the corresponding tabulated χ_2 value is 31.41 at the 5% level. So we reject H_0 in both cases at the 5% level.

Table 15. Estimation by ordered probit using means of MP3 acquisition

	(1)		(2)		(3)	
	Coef.	Std. err.	Coef.	Std. err.	Coef.	Std. err.
Female	0.2524	0.1839	-0.3922	0.4137	0.4583	0.2193
Age 20-29	0.2405	0.1740	0.4641	0.4809	0.1720	0.1932
Age 30+	0.5200	0.3918	0.7818	0.6950	0.3966	0.5064
School	0.0125	0.1134	0.0985	0.2733	0.0093	0.1311
Internet 6-15	0.1606	0.1843	0.5779	0.4453	0.0087	0.2150
Internet 15-30	0.1664	0.1974	0.5579	0.5003	0.1187	0.2257
Internet 30+	0.0248	0.1935	-0.3356	0.4483	0.0742	0.2273
Music 5-10	-0.0168	0.1716	-0.1808	0.3850	-0.0531	0.2023
Music 10+	0.2461	0.1634	-0.3592	0.3781	0.3313	0.1939*
Record stores	1.0128	0.1308***	1.6893	0.3256***	0.9480	0.1544***
Music magazines	0.3938	0.2126*	0.9118	0.5240*	0.3601	0.2411
Live concerts	0.3676	0.1303***	-0.5753	0.3406*	0.5375	0.1499***
Play instrument	0.1355	0.1224	-0.4476	0.2960	0.2136	0.1409
CDs 30-50	0.6801	0.1606***	1.4834	0.4328***	0.6510	0.1862***
CDs 50-100	0.8675	0.1688***	1.9042	0.4777***	0.8079	0.1928***
CDs 100-200	1.4629	0.1961***	2.6653	0.5488***	1.5351	0.2325***
CDs 200+	1.6490	0.2311***	3.2810	0.6130***	1.5813	0.2714***
Peer-to-peer	-0.2206	0.1203*	0.1585	0.3095	-0.3956	0.1420***
Intranet	-0.3011	0.1356**	-0.3450	0.3040	-0.3034	0.1588**
Physical exchanges	0.2560	0.1133**	0.6684	0.2722***	0.1570	0.1322
Cut point 1	0.0207	0.2427	0.0755	0.5012	-0.0764	0.2946
Cut point 2	2.0134	0.2649	3.0044	0.6079	1.8491	0.3186

Cut point 3	3.1578	0.2848	4.3534	0.6715	3.0669	0.3425
Log likelihood	-414.70		-80.88		-311.72	
Number of obs.	456		112		344	

Table 16. Estimation by ordered probit using number of MP3 files

	(1)		(2)		(3)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Female	0.1603	0.1888	-0.4277	0.3971	0.2418	0.2295
Age 20-29	0.2874	0.1712	0.5800	0.4510	0.1963	0.1909
Age 30+	0.4283	0.3989	1.0090	0.6896	0.3303	0.5135
School	-0.0204	0.1118	0.0978	0.2716	-0.0300	0.1293
Internet 6-15	0.1756	0.1836	0.6094	0.4356	0.0224	0.2158
Internet 15-30	0.2455	0.1988	0.6835	0.5124	0.1582	0.2268
Internet 30+	0.0845	0.1929	-0.1229	0.4414	0.1170	0.2279
Music 5-10	-0.0360	0.1721	-0.2155	0.3722	-0.0490	0.2038
Music 10+	0.3201	0.1655**	-0.3792	0.3642	0.4689	0.1957**
Record stores	0.9568	0.1302***	1.5821	0.3202***	0.8298	0.1514***
Music magazines	0.4662	0.2127*	0.8260	0.5163	0.4456	0.2418*
Live concerts	0.3856	0.1312***	-0.5297	0.3385	0.5570	0.1517***
Play instrument	0.1328	0.1231	-0.3956	0.2876	0.2053	0.1423
CDs 30-50	0.7171	0.1604***	1.5802	0.4243***	0.6308	0.1850***
CDs 50-100	0.9413	0.1718***	1.8828	0.4619***	0.8894	0.1972***
CDs 100-200	1.5222	0.1985***	2.6462	0.5391***	1.6132	0.2360***
CDs 200+	1.7811	0.2339***	3.2487	0.6000***	1.7601	0.2741***
MP3s 1-100	0.4431	0.1777**	0.1031	0.3483	0.5625	0.2242***
MP3s 101-200	0.5451	0.1818***	0.0396	0.3377	0.7768	0.2410***
MP3s 201-500	0.2667	0.1850	0.1308	0.4253	0.4665	0.2191**
Cutpoint 1	0.4945	0.2308	0.1150	0.5069	0.5696	0.2713
Cutpoint 2	2.4997	0.2620	2.9154	0.6181	2.5316	0.3080
Cutpoint 3	3.6445	0.2855	4.1903	0.6715	3.7490	0.3385
Log likelihood	-414.28		-84.67		-310.03	
Number of obs.	456		112		344	

4 Discussion

Although the survey was not designed to measure the overall quantitative effect of MP3 files on CD sales, we can tabulate answers to the effect of MP3s on CD purchases of a

respondent with respect to his answers to the number of CDs he purchases on average per year. This is done in Table 17.

Table 17. CD purchases and MP3 influence

Annual CD purchases	Influence			Total
	Decrease	Same	Increased	
0	28	47	4	79
1-4	83	110	28	221
5-10	18	51	23	92
>10	7	30	27	64
Total	136	238	82	456

We can use these numbers to create different scenarios about the effect of MP3 files on CD sales among survey respondents. Table 8 was constructed by taking the mean of each CD purchase category and by assuming that respondents increased or decreased CD purchases by 10%, 20%, 30% and 40%. We divided the results by the total number of CD purchased in the population (including those who claimed MP3s have no influence on CD consumption). Assuming that respondents who increased their music purchases increased them by the same percentage as those who decreased CD consumption would lead to an overall positive effect of MP3 files on CD sales (numbers in the main diagonal of Table 18). Table 18 also indicates that while record companies could loose sales by at most 5.2% per year due to MP3 downloads, they could also increase sales by more than 9.7% per year by offering music fans attractive options to sample new music. Measuring accurately the substitution and the complementarity between MP3 files and CDs is an interesting topic for further research. We would like again to stress that results from this exercise should be taken with caution since we work with a specific sample and that answers are highly subjective and self-reported. However, numbers reported in Table 17 (30% increase music spending, while 18% decrease it) are strikingly close to those found in a survey carried out by Peter Hart commissioned by the RIAA in 2002 that found that 15% of music downloaders spent more on music purchases while 27% spent less on music since they started downloading. Other studies reported by the IFPI (2004) find that

15% of file-sharers have increased music spending, while "for every one of these *samplers*, there are two *substituters* who (...) use P2P networks to avoid the cost of buying music." (p. 16)

Table 18. Effect of MP3 on CD sales

Decrease by	Increase by			
	10%	20%	30%	40%
10%	0.91%	3.85%	6.79%	9.73%
20%	-1.12%	1.82%	4.76%	7.70%
30%	-3.16%	-0.22%	2.72%	5.66%
40%	-5.19%	-2.25%	0.69%	3.63%

5 Conclusion

The main contribution of this article is to show that there are two populations of MP3 users: people who use MP3s to sample and discover new music through physical exchanges, which leads them to purchase more CDs (the explorers) and people who mainly use file-sharing networks such as Peer-to-Peer networks and Intranets as substitutes to regular CDs (the pirates). Thus new file-sharing technologies have amplified consumption patterns in the sense that music fans have used MP3 to discover new music and increase their consumption of pre-recorded music while people with low willingness to pay for music have used MP3 files as direct substitutes to legal purchases. Our estimation results have two implications. First, new business models should try and better discriminate between these two types of users of digital music in order to extract more surplus from the true music fans. There are already several interesting initiatives. Napster 3.0 allow internet users to download as many music files as they want for a monthly fee, while Altnet uses Peer-to-Peer networks to match file-sharing users to new artists and allow the former to listen to DRM-protected files for several times before a

protection is triggered. Secondly, new business models should also encourage the formation of online communities as our results suggest that people who share music through personal contacts among a community have a higher probability to purchase more CDs.

6 References

- Bakos, Y., Brynjolfsson, E. and Lichtman, D. (1999) "Shared Information Goods," *Journal of Law and Economics*, 117, 23-25.
- IFPI (2004), *The Recording Industry in Numbers 2003*.
- Liebowitz, S. (2004), "Pitfalls in Measuring the Impact of File-sharing," forthcoming in *CESifo Economic Studies*.
- Moreau, F. and N. Curien (2005), "The Music Industry in the Digital Era: Towards New Business Frontiers," Mimeo
- Peitz, M. and P. Waelbroeck (2004a), "An Economist's Guide to Digital Music," *CESifo Economic Studies*, forthcoming.
- Peitz, M. and P. Waelbroeck (2004b), "Files-sharing, Sampling and Music Distribution," *International University in Germany Working paper 26/2004*
- Rob, R. and J. Waldfogel (2004), "Piracy in the High C's: Music Downloading, Sales Displacements, and Social Welfare in a Sample of College Students," *NBER Working Paper*.
- Zentner, A. (2004), "Measuring the Effect of Online Piracy on Music Sales," *mimeo*, University of Chicago.

7 Appendix 1

Table A1. Summary statistics

Variable	Mean	Std. Dev.
CDs	4.8300	4.7620
Female	0.1031	-
Age ≤ 24	0.8531	-
Age 25-29	0.1250	-
Age ≥ 30	0.0219	-
School	0.4693	-
Internet < 5	0.1228	-
Internet 6-15	0.3531	-
Internet 15-30	0.2434	-
Internet > 30	0.2807	-
Music < 5	0.1732	-
Music 5-10	0.2961	-
Music > 10	0.5307	-
Record stores	0.4408	-
Music magazines	0.1031	-
Live concerts	0.3662	-
Play instrument	0.3421	-
CDs < 30	0.3289	-
CDs 30-50	0.2127	-
CDs 50-100	0.1974	-
CDs 100-200	0.1491	-
CDs > 200	0.1118	-
Peer-to-peer	0.6513	-
Intranet	0.7500	-
Physical exchanges	0.5329	-
MP3s 1-100	0.1404	-
MP3s 101-200	0.1228	-
MP3s 201-500	0.1206	-
MP3s > 500	0.6162	-

Table A2. "Pirates" and "explorers"

	Delete?	
	Yes	No
Female	0.116	0.099
Age ≤ 24	0.875	0.846
Age 20-29	0.089	0.137
Age ≥ 30	0.036	0.017
School	0.393	0.494
Internet < 5	0.134	0.119
Internet 6-15	0.357	0.352
Internet 15-30	0.161	0.270
Internet > 30	0.348	0.259
Music < 5	0.179	0.172
Music 5-10	0.339	0.282
Music > 10	0.482	0.547
Record stores	0.438	0.442
Music magazines	0.071	0.113
Live concerts	0.330	0.378
Play instrument	0.313	0.352
CDs < 30	0.348	0.323
CDs 30-50	0.241	0.203
CDs 50-100	0.143	0.215
CDs 100-200	0.152	0.148
CDs > 200	0.116	0.110
Peer-to-peer	0.634	0.657
Intranet	0.696	0.767
Physical exchanges	0.482	0.549
MP3s 1-100	0.232	0.110
MP3s 101-200	0.214	0.093
MP3s 201-500	0.125	0.119
MP3s > 500	0.429	0.677