PIRACY ON THE HIGH C'S: MUSIC DOWNLOADING, SALES DISPLACEMENT, AND SOCIAL WELFARE IN A SAMPLE OF COLLEGE STUDENTS*

RAFAEL ROB
University of Pennsylvania

and

JOEL WALDFOGEL
University of Pennsylvania

ABSTRACT

Recording industry revenue has fallen sharply in the last 3 years, and some—but not all—observers attribute this to file sharing. We collect new data on albums obtained via purchase and downloading, as well as consumers' valuations of these albums, among a sample of U.S. college students in 2003. We provide new estimates of sales displacement induced by downloading, using both ordinary least squares and an instrumental variables approach with access to broadband as a source of exogenous variation in downloading. We find that each album download reduces purchases by about .2 in our sample, although possibly by much more. Our valuation data allow us to measure the effects of downloading on welfare as well as expenditure in a subsample of University of Pennsylvania undergraduates, and we find that downloading reduces their per capita expenditure (on hit albums released 1999–2003) from $126 to $101 but raises per capita consumers' surplus by $70.

> No black flags with skull and crossbones, no cutlasses, cannons, or daggers identify today's pirates. You can't see them coming; there's no warning shot across your bow. Yet rest assured the pirates are out there because today there is plenty of gold (and platinum and diamonds) to be had. [Recording Industry Association of America 2003]

I. INTRODUCTION

AFTER growing an average of 10 percent per year over the previous 7 years, U.S. music industry revenue has shrunk by 16 percent in the last 3 years (see Figure 1). Although there is considerable debate about causes, the advent and spread of Internet technology, and the file sharing that it enables,

* We are grateful to Betsy Bailey, Mary Benner, Lisa George, Jeff Milyo, and Yochanan Shachmurove for allowing us to administer surveys to their students and to Hannah and Sarah Waldfogel for help administering the surveys. Lu Chen, David Sokoler, and Duygu Baydur provided diligent research assistance. We thank Felix Oberholzer-Gee and Koleman Strumpf, as well as seminar participants at the University of Pennsylvania Law School, the University of Toronto, the Wharton School, the Kellogg School of Management, and Yale University, for

[*Journal of Law and Economics*, vol. XLIX (April 2006)]
 0022-2186/2006/4901-0002$01.50

have emerged as potentially important culprits. This study examines the effect of downloading on sales and welfare. In particular, we address the following three questions. (1) How many purchased albums are displaced by each album that is illegally downloaded? (2) Is the subjective valuation of downloaded albums high or low? (3) What are the revenue and welfare consequences of downloading?

To answer these questions we administered a survey at four U.S. colleges. Among other things, the survey asked students to list which and how many albums they purchased and downloaded, whether they had an Internet connection, and, if so, whether the connection was broadband. We surveyed 412 students, eliciting information on their purchases and downloads of 8,200 albums. Using both ordinary least squares (OLS) and an instrumental variable approach with access to broadband as a source of exogenous variation in downloading, we show that one downloaded album reduces music purchases by roughly one-fifth of an album, and possibly by much more. Using this conservative estimate and information on the average number of albums purchased and downloaded, we infer that downloading reduced purchases by individuals in the sample by about 10 percent during 2003.

We then turn to the question, which albums tend to be downloaded, those to which individuals assign high or low subjective valuation? To address that question, the survey asked respondents not just what they purchased and downloaded, but also how highly they value the music in dollar terms. We document that, on average, our respondents downloaded music that they value a third to a half less than their purchased music. This result is consistent with our incomplete-displacement finding (that one downloaded album displaces less than one purchased album), which indicates that at least some of the music that is downloaded would not have otherwise been purchased. It also suggests that some of the surplus enjoyed by downloaders would have otherwise been the deadweight loss associated with foregone, but socially beneficial, transactions.

This last result led us to revenue and welfare analysis. Two features of the CD market make the welfare analysis of downloading interesting. First, substantial price discrimination was impracticable and was not exercised. As a result, firms priced albums as single-price monopolists, leaving some socially beneficial transactions (with buyer valuation above marginal cost but below the monopoly price) unconsummated. This problem was exacerbated by CDs' second feature: the product has a low marginal cost, so the market without illegal downloading has the potential for substantial deadweight loss.

While perhaps paradoxical to the law-abiding citizen, illegal downloading may actually alleviate the monopoly deadweight-loss problem. Indeed, down-

helpful comments. Rob thanks the National Science Foundation and Waldfogel thanks the Wharton Electronic Business Initiative and the Mack Center for Technological Innovation at Wharton for financial support. All errors are our own.

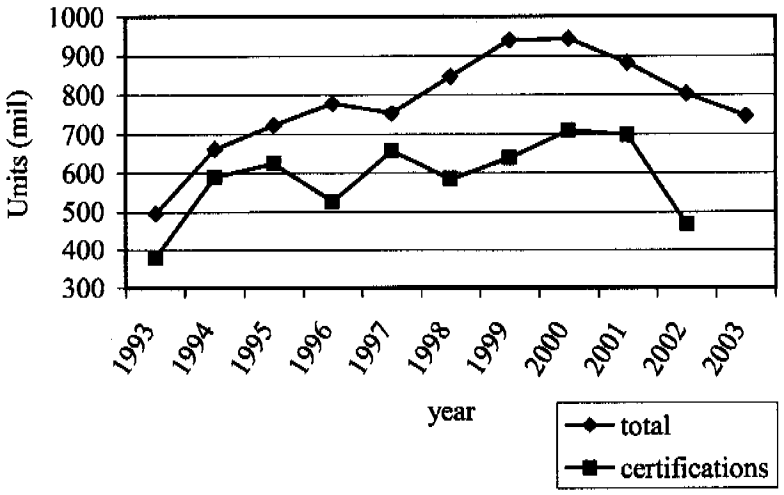

FIGURE 1.—Albums sold, 1993–2003

loading allows consumers to engage in a crude "do-it-yourself" form of price discrimination. On the basis of access to the Internet and willingness to engage in file sharing, consumers segment themselves into downloaders and buyers. Downloaders pay zero and are clearly (by revealed preference) better off. Whether these gains are offset by harm done to firms and their suppliers depends on which albums consumers download. If downloading tends to occur for albums that consumers value highly and would otherwise have purchased, then revenues are reduced and an offsetting harm indeed occurs. However, if downloading tends to occur for low-valued albums, downloaded albums are not candidates for being purchased in the first place, and, hence, no harm is done. As stated above, our empirical results indicate that downloaded albums tend to be low valued, which suggests that the harm done by downloading is limited.

To further investigate this question and to quantify welfare losses or gains, we made a welfare comparison between a (counterfactual) regime in which illegal downloading is feasible and a regime in which it is not. To carry out this comparison, we administered an additional survey to 92 students, eliciting both ex ante and ex post valuation information for 1,209 purchased and downloaded music albums (this information is elicited because music is an experience good, so the purchase versus no-purchase decision is based on ex ante valuations). Using the information on ex post valuations, we created an empirical demand curve for music, which we apply to both the downloading and the no-downloading regimes. Using the information on ex ante valuations and assuming that consumers would have paid $15 for each album

they valued at or above $15, we partitioned individual-album pairs into situations that would and would not have resulted in purchases if downloading had not been feasible. We then compute welfare at the (counterfactual) outcome that would have occurred in the no-downloading regime as well as welfare at the (factual) outcome in the downloading regime and compare the two. We found that downloading reduces per capita expenditures by individuals in this subsample by $25, from $126 to $101 per capita, but raises consumers' surplus by $70 per capita (for purchases of hit albums 1999–2003). Thus, illegal downloading effected a per capita decrease of $45 in deadweight loss, which is nearly double the reduction in industry revenue.

The paper proceeds in five sections. Section II provides industry background and links to relevant literatures. Section III explores possible effects of downloading on consumption and welfare in theory. Section IV describes the data used in the study. Section V presents empirical results on sales displacement and whether high- or low-valuation albums are downloaded. Section V also presents estimates of the effects of downloading on welfare. Section VI concludes.

II. Background

A. Industry Background

The U.S. music industry is highly concentrated. Although there are hundreds of labels, five owners account for 80 percent of music sales. The major firms include Sony, Vivendi-Universal, EMI, Bertelsman, and Time Warner (Graves 2004). According to some reports, there are nearly 30,000 CDs released per year. The major labels collectively release about 7,000.[1] Sales are very highly concentrated, however. In 2000, the top 250 (100, 50, 10) albums accounted for 48 (32, 22, 8) percent of total sales. Sales in other years are similarly skewed.

After increasing about 10 percent per year during the 1993–99 period, U.S. album sales have declined 16 percent since that time (see Figure 1). The recording industry, represented by the Recording Industry Association of America (RIAA), blames illegal downloading, which it terms "piracy," for sales displacement. "Each year, *the industry loses about $4.2 billion to piracy worldwide*—'we estimate we lose millions of dollars a day to all forms of piracy'" (RIAA 2003). The decline in demand is substantial enough to encourage consolidation. Sony and Bertelsman have had discussions, and EMI and Time Warner have been rumored to be in discussions (Helmore 2003).

Other developments are suggestive of downloading. Internet penetration has increased from 16 percent in 1997 to 47 percent in 2000. And broadband

[1] These figures are attributed to the RIAA (Ziemann 2002).

connection, allowing convenient downloading of large files, stood in early 2003 at 40 percent of connected households. Moreover, the sales of important complementary goods, MP3 players and blank CDs, have increased substantially over the past 5 years. Sales of MP3 players grew to an estimated 3.8 million in 2003 (Consumer Electronics Association 2004). Of course, one can use such goods for purchased as well as downloaded music, so these developments are merely suggestive of illegal downloading.

Concerned with reductions in revenue—and convinced that music downloading is the cause—the industry responded in three ways. First, the industry initiated hundreds of lawsuits against illegal downloaders as a means of deterring downloading activity. There is some evidence that the lawsuits are having an effect (Schwartz 2004). Second, the industry slashed prices, making it more attractive to purchase albums rather than illegally download them. For instance, Universal reduced its list prices by 30 percent in October 2003 (McCarthy and Grant 2003).

The third response is more recent. The music industry, initially slow to offer sales over the Web, has moved recently to make music available online. Leading this trend is Apple Computer, maker of the dominant MP3 player, which launched its iTunes music site in 2002. There consumers can download music for $.99 per song. As of March 2004, customers had downloaded 50 million songs from iTunes (Apple Computer 2004).

B. Relevant Literature

1. Theoretical Literature

This study is related to several strands of theoretical research. The first strand is research on differentiated product oligopoly, which includes the literature on monopolistic competition, found in Spence (1976a, 1976b) and Dixit and Stiglitz (1977). This literature considers the endogenous offering of product variety (in our case, different genres of music and different varieties within one genre) and how that relates to the firm revenues. The revenues that firms harvest, in turn, are determined by the competition and the demand conditions they face. Obviously, downloading reduces demand and thereby undermines the ability of firms to harvest revenues. Our paper speaks to this revenue effect but not to whether this induces a reduction in the varieties of music offered.

More relevant to what we do here is the literature on the effect of reproduction technology on markets for information goods, beginning with work by Liebowitz (1985), who argues that reproduction of journal articles through photocopying need not reduce the demand for and, hence, the profit from selling academic journals. He argues that prices that publishers charge libraries can be raised to reflect the value to library patrons from photocopying. Then, although publishers lose much of the individual market to photocopiers,

they recoup some of these losses from higher library prices. The issues addressed and the line of reasoning in Liebowitz are echoed in the software piracy literature (Besen 1986; Conner and Rumelt 1991; Takeyama 1997).

More recent literature extends these ideas to a context in which file sharing takes place in small groups—for example, within families or among friends. A number of papers make the point that file sharing need not reduce seller revenue (Bakos, Brynjolfsson, and Lichtman 1999; Varian 2000). To see this, imagine two individuals each valuing a good below its price. If they share cost and content, they may together be willing to purchase the good that they individually were not willing to purchase, which implies that seller revenue may actually increase rather than decrease. Our work relates to this strand of literature inasmuch as we document the effect of reproduction or downloading on demand and seller revenue. On the other hand, our work does not relate to the argument that file sharing, which takes place in an anonymous setting and without a middleman who stands to benefit from it, may benefit sellers because cost sharing is not feasible. Another theoretical argument relevant to our study is that file sharing is like free samples and, as such, it may actually stimulate sales (Shapiro and Varian 1999).

2. Empirical Literature

Our paper is one of a number of recent empirical studies of the effect of music downloading on album sales. Liebowitz (2003) examines a variety of possible explanations for the recent reductions in album sales and, finding them all wanting, concludes that downloading must be responsible.

Zentner (2003) uses international time-series aggregate data, in conjunction with Internet connectedness, to document that places with more Internet—and broadband—connections have experienced sharper reductions in album sales. He also uses European individual-level data to show that persons who self-report downloading music—instrumented with measures of technical sophistication—are also less likely to have purchased music recently. Zentner's micro data include only binary measures of downloading and music purchase (Do you download music? Have you purchased music recently?), which prevents him from measuring the size of the downloading-induced sales displacement. By contrast, we collected data on the numbers of albums purchased and downloaded, which enables us to quantify this displacement.

Hui and Png (2003), using international panel data for 1994–98, estimate that each download reduces sales by .42. The time period they study predates the growth of broadband and widespread file sharing. Moreover, their measure of piracy has the shortcoming that it is an estimate based in part on the level of legitimate sales in a country.

On the other hand, Oberholzer and Strumpf (2004) examine weekly time-series data on downloading and sales of major hit albums and find little relationship. And in expert testimony on the Napster case, Fader (2000)

argues that file sharing stimulates sales of recorded music, although his analysis covers a period prior to the sales decline depicted in Figure 1.

Oberholzer and Strumpf (2004) combine weekly album sales data by album with novel data on the weekly volume of downloads. Their approach identifies sales displacement using within-album weekly variation in downloading and sales. Using this data and estimation approach, they ask whether an album sells fewer copies in a week that it is downloaded frequently. This approach faces three obstacles, however. The first obstacle is that CDs are durable goods and, hence, if substitution or displacement occurs, it need not occur within 1 week. If an individual decides to download instead of purchase, her decision may very well be reflected in her future purchasing behavior; the absence of contemporaneous substitution does not rule out substitution more generally.

Their contemporaneous-album-sales approach also faces the handicap that the variation in a particular album's popularity over time would tend to induce a positive relationship between purchases and downloads. Purchase and download are simply two ways of obtaining an album. To see this clearly, consider a different pair of channels for obtaining an album, from Sam Goody or from HMV. Virtually no one purchases a particular album from both outlets. Thus, the displacement is one for one. Yet the weekly sales of a particular album at Sam Goody are surely highly positively correlated with the weekly sales of the same album at HMV (as a result of aggregate temporal shocks that affect both). Do HMV sales stimulate Sam Goody sales? Probably not. Looking at downloading and purchases has the same problem. An album's popularity has a time component to it, and when it is popular through one distribution channel, it is popular through other channels as well. But this does not mean that availability through one of the channels stimulates demand through other channels.

A third shortcoming is that this estimation approach does not necessarily inform the sales versus displacement question. Suppose there are two types of consumers, buyers and downloaders. Suppose that buyers never download and downloaders never buy. Then a negative correlation between sales and downloads across albums just means that buyers and downloaders have different tastes. What we really want to know is whether a person's downloading of an album reduces the probability that he will purchase the album. Hence, it is not enough to have sales and downloads by album. Rather, one needs data on sales and downloads by person, or groups of persons.

The ideal data for studying sales displacement would be volumes of sales and downloads by individual rather than by album. If one could find exogenous variation in downloading across individuals, then displacement could be inferred from the relationship between downloads and sales (for example, do people who download more albums, or songs, purchase fewer albums?). To our knowledge, surveys are the only way to obtain this information.

A lack of alternative data sources led us to undertake our own surveys.

The benefit of this data collection is that it allowed us to get information on individuals' volumes of music downloads and purchases (and on willingness to pay). Because we surveyed the college student populations available to us, our surveys are necessarily not nationally representative. We cannot therefore draw inferences about the magnitude of the U.S.-wide effect of downloading on album sales. But we can ask whether the sales displacement phenomenon operates in our sample. That is, do people who download more purchase less? If so—and if we believe we have isolated exogenous variation in the volume of downloading—then we can at least draw an inference about whether downloading—again for our sample—tends to stimulate or cannibalize album sales. This is the spirit in which we proceed.

III. Theory

In this section we outline the theory underlying our welfare comparison exercise. It is the routine theory of monopoly pricing and its attendant deadweight loss, except that we have to specify the outcome that would have occurred in the (counterfactual) no-downloading regime. To this end, two features are relevant. The first is the distinction between ex post and ex ante valuation, which is necessary because music is an experience good. The second is the relationship between the decision to download and the subjective ex post valuation. To simplify the exposition we ignore for now the distinction between ex ante and ex post valuation and focus instead on the possibility that downloading may be correlated with valuations. We integrate ex ante valuation into the theory immediately afterward.

To further simplify the exposition we also assume that the marginal cost of CDs is zero. One justification for this is that digital distribution, which has virtually zero marginal cost, is technically feasible. If the marginal cost is not zero, then our computations are easily adapted, since the procedure that we introduce here (and on which these computations are based) is quite general.

A. Effects of Downloading on Revenues and Welfare

Consider a world without downloading and let $D(p)$ be a demand function for some album. Since each consumer buys one album at the most, $D(p)$ reflects the distribution of consumers' willingness to pay for this album. The seller is viewed as a single-price monopolist. Then, for a typical demand function, this will lead to monopoly pricing, restriction of output, and deadweight loss. The areas corresponding to revenue, consumers' surplus, and deadweight loss at the profit-maximizing outcome are illustrated in Figure 2. These areas add up to the total area under the demand curve. This provides a baseline with which the outcome under downloading is compared.

Consider now a world with downloading. Then the overall demand $D(p)$ is segmented into two subdemands, buyers' demand $D_B(p)$ and down-

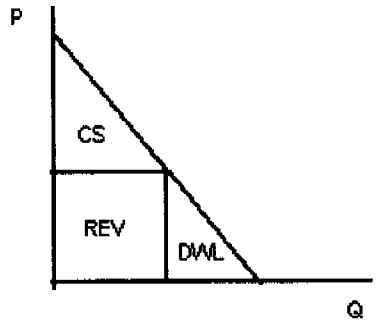

FIGURE 2.—Single-price monopoly: the market before downloading. CS = consumer surplus; DWL = deadweight loss; rev = revenue.

loaders' demand $D_D(p)$. The term $D_D(p)$ reflects hypothetical willingness to pay, since downloaders' demand does not (obviously) translate into revenues. By definition, these subdemands add up to the overall demand: $D_D(p) + D_B(p) = D(p)$, for each p.

To show how welfare comparisons (in terms of deadweight loss, revenues, and consumers' surplus) between a downloading and a no-downloading regime work, and what these comparisons hinge on, we work through three illustrative examples.

First, consider an example in which each consumer is a downloader with some probability, say q, and a buyer with the complementary probability, $1 - q$, where q is independent of the consumer's valuation. In this case downloaders are uniformly drawn from the overall distribution of valuations. The buyers' demand curve in this "neutral" example is simply a proportional inward rotation of the overall demand, $D_B(p) = (1 - q)D(p)$. In this example downloading raises consumers' surplus while reducing revenue and deadweight loss; the monopolist's profit-maximizing price remains the same.

Second consider two polar cases, illustrated in Figures 3 and 4, where q does depend on valuation. In one polar case, low-valuation consumers are downloaders and high-valuation consumers are buyers, so $q = 1$ for sufficiently low p and $q = 0$ for sufficiently high p. To further accentuate what happens in this case, let the critical p (the p at which q changes from zero to one) be the predownloading price. Then the monopolist's postdownloading price equals his predownloading price. Also, all consumers with valuations above p continue to purchase, while all consumers with valuations between p and zero, which is the marginal cost of downloading, download. In this extreme case, both revenue and consumers' surplus of persons who bought prior to downloading are unaffected by downloading. The deadweight loss existing prior to downloading is transformed into consumers' surplus (see

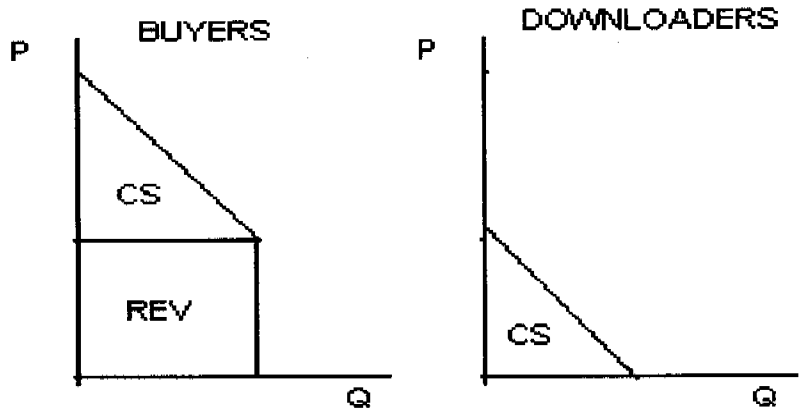

FIGURE 3.—Downloaders are low-valuation demanders. CS = consumer surplus; rev = revenue.

Figure 3). This is a case, therefore, in which there are only welfare gains and no welfare losses from downloading.

In the other polar case, high-valuation consumers are downloaders and low-valuation consumers are buyers, so $q = 1$ for sufficiently high p and $q = 0$ for sufficiently low p. Again, let the critical p be the predownloading price. Then, given that high-valuation consumers no longer buy, the monopolist lowers his price to attract some low-valuation consumers. Consequently, regions representing consumers' surplus and revenue prior to downloading become downloader surplus. And the region formerly representing deadweight loss now contains revenue, additional consumers' surplus, and some deadweight loss. Thus, overall consumers' surplus in this case increases and deadweight loss decreases. Seller's revenue, however, decreases, so there is harm along with reduced deadweight loss (see Figure 4).

As stated earlier, these special cases are concocted to illustrate how downloading works and its welfare consequences. A realistic analysis of the welfare effects of downloading requires information about the actual distribution of valuations for buyers and downloaders. Then we can measure areas under the empirical demand curve and quantify welfare changes. For now let us just note that the key question, for both the welfare analysis of downloading and the size of the sales displacement, is whether individuals download high- or low-valuation albums. The more that consumers obtain low-valuation music through downloading, the smaller its negative impact on revenue and the greater its beneficial effect on welfare.

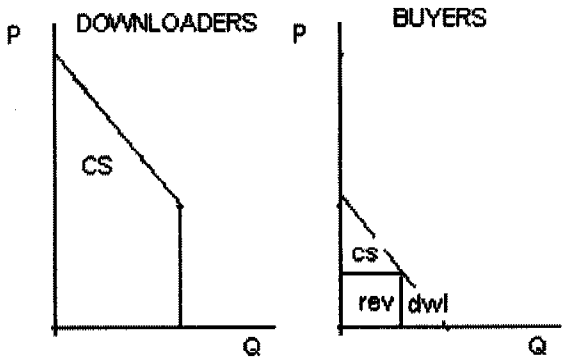

FIGURE 4.—Downloaders are high-valuation demanders. CS = consumer surplus; DWL = deadweight loss; rev = revenue.

B. Procedure and Incorporation of Ex Ante Valuation

Because music is an experience good, the ex ante valuation determining purchase is not the same as the ex post valuation, which becomes known only after purchase. As a result, actual welfare is slightly more complicated than the foregoing section would suggest. At the time of first obtaining the album, the individual's willingness to pay is based only on a guess about how much she will like an album. Call this the ex ante valuation, v_i^a, where superscript a denotes ex ante and subscript i indexes a consumer-album pair. A consumer's ex post valuation of the album, once she has listened to it for a while, can deviate from v_i^a. Term the ex post valuation v_i^p, where the superscript p denotes ex post. Define p_i as the purchase price the consumer faces. Finally, each consumer-album pair has a binary willingness to download: δ_i equals one if yes, zero if no.

As before, there are two regimes: a downloading regime (when illegal downloading is feasible) and a no-downloading regime (when it is not). When downloading is feasible, each consumer-album pair requires first a decision about willingness to download. If δ_i equals one (yes), then downloading occurs as long as $v_i^a > 0$ (regardless of whether $v_i^a \geq p_i$). If δ_i equals zero (no), then the consumer purchases only if $v_i^a \geq p_i$. If δ_i equals zero and $v_i^a < p_i$, she goes without, even though she has valuation in excess of downloading's zero marginal cost. In the no-downloading regime, individuals purchase only when $v_i^a \geq p_i$.

While individuals decide whether and how to obtain albums on the basis of (v_i^a, δ_i, p_i), their ultimate satisfaction from the album depends on v_i^p, less a price paid, if any. When this theory is implemented, the procedure for computing welfare in the two regimes is as follows. Consumers' surplus in the downloading regime is CS = v_i^p for downloaded albums, CS = v_i^p −

TABLE 1

SUMMARY OF BEHAVIOR AND WELFARE IN DOWNLOADING AND NO-DOWNLOADING REGIMES

Ex Ante Valuations Relative to Price	Willing to Download ($\delta_i = 1$)	Not Willing to Download ($\delta_i = 0$)
Downloading regime:		
$0 < v_i^a < p_i$:	v_i^p = CS rev = 0 downloading occurs	v_i^p = DWL rev = 0 album not obtained not observed in our data
$v_i^a \geq p_i$:	v_i^p = CS rev = 0 downloading occurs	$v_i^p - p_i$ = CS p_i = rev purchase occurs
No-downloading regime:		
$0 < v_i^a < p_i$:	v_i^p = DWL rev = 0 album not obtained	v_i^p = DWL rev = 0 album not obtained not observed in our data
$v_i^a \geq p_i$:	$v_i^p - p_i$ = CS p_i = rev purchase occurs	$v_i^p - p_i$ = CS p_i = rev purchase occurs

NOTE.—Assume that zero marginal cost distribution is available when illegal downloading is not feasible. CS = consumers' surplus; DWL = deadweight loss; rev = revenue.

p_i for purchased albums, and zero otherwise. Sellers' revenue in the downloading regime is p_i for each purchased album and zero otherwise. Consumers' surplus in the no-downloading regime is CS $= v_i^p - p_i$ (if they purchase), and seller's revenue is p_i. Table 1 summarizes this information in the downloading and no-downloading regimes and according to whether δ_i equals one and whether $v_i^a \geq p_i$. To deliver quantitative information we let $p_i = 15$ and let δ_i take its empirical value (one if the individual bought and zero otherwise).

One point bears discussion here. Because our survey—discussed below—elicits information only about albums that individuals have in their possession (either via purchase or downloading), we will have data on v^a, v^p, and δ for only three of the four cells in the tables. We have no valuation information about albums that people are neither willing to download or purchase (with $\delta = 0$ and $v_i^a < p_i$). Note, however, that the deadweight loss in this cell is identical across downloading and no-downloading regimes, so our inability to measure this will not inhibit our ability to measure the change in welfare from downloading.

IV. DATA

The basic data for this study are derived from two surveys administered to college students (at the University of Pennsylvania, Hunter College, the University of Chicago's Master of Arts [M.A.] program in public policy, and City College of New York [CCNY]) between December 2003 and February

2004. The first survey, administered to 412 students, elicited information on the respondent (age, race, gender, family income, number of CDs owned, speed of Internet access, and interest in music), along with ex post valuation on two groups of albums purchased or downloaded: those obtained in the last year and those among a common list of 261 hit albums released since 1999.[2]

Table 2 presents some characteristics of the individuals in the survey 1 sample. Blacks make up 9 percent of the sample; Hispanics, 10 percent; and Asians, 32 percent. The remainder are white. Fourteen percent of the sample has family income below $25,000; 20 percent has family income between $25,000 and $50,000; 28 percent has family income between $50,000 and $100,000; 22 percent has family income between $100,000 and $250,000; and the remaining 16 percent has family income in excess of $250,000. The mean age is 21.9.

The respondents report a high level of interest in music. Only 15 percent report being less of a music fan than others they know. Nearly 40 percent claim to be about the same, 30 percent are somewhat more, and 17 percent are a lot more interested in music than others they know. The mean CD collection is 103.

In a second wave of the first survey we asked about current and past Internet access. We obtained this information on 260 of the 412 respondents. Seventy percent report having dial-up access in 1999, while 21 percent had broadband (digital subscriber line [DSL], cable modem, or Ethernet) access in 1999. By 2003, roughly 85 percent had high-speed access, while the remainder had dial-up access.

Respondent characteristics vary substantially across the universities sampled. Chicago and Penn are heavily white and Asian, while Hunter and especially CCNY have larger black and Hispanic populations. Chicago income is low because the M.A. students tend to report their own, as opposed to parents', income (note their comparatively high ages). Penn income is substantially higher than elsewhere. Reported interest in music is somewhat higher at Penn and Hunter than elsewhere. Finally, patterns of Internet access differ substantially across schools. Penn students, most living in wired dorms, have virtually 100 percent broadband access. Rates elsewhere are much lower.

The second part of survey 1 asks respondents to attach a dollar value to two groups of albums they have. First, we present them with a list of all 261 albums certified by the RIAA as having sold 2 million or more copies since 1999. For each album, respondents indicate whether they own it, how they got it, and how highly they value it in dollar terms. We term this the "hit"

[2] Our survey responses are unlikely to be much affected by legal song downloading sites such as iTunes for two reasons. First, many of the albums we ask about were released before iTunes was introduced in April 2003. Second, iTunes' cumulative sales reached 50 million songs after we completed our surveys. At $.99 per song, iTunes' online song sales amount to less than half a percent of annual CD sales.

TABLE 2

RESPONDENT CHARACTERISTICS

	Overall	City College of New York	University of Chicago	Hunter College	University of Pennsylvania	
					Survey 1	Survey 2
Male	60	65	29.7	45.4	67.7	58
N	414	66	64	33	251	98
Black	8.9	31.8	10.9	6.3	4.0	3.0
Hispanic	9.8	33.3	3.1	25.0	3.2	1.0
Asian	31.8	21.2	26.6	12.5	38.5	43.4
White	48.2	13.6	59.4	53.1	53.8	52.5
N	409	66	64	32	247	99
Age	21.9/400	24.6	27.4	20.5	20.0	20.2
Annual income:						
<$25K	14.3	23	41.4	12.9	5.2	2.2
$25K–$50K	19.8	41.5	25.9	38.7	19.8	8.8
$50K–$100K	27.6	29.2	17.2	29.3	27.6	36.3
$100K–$250K	21.9	4.6	12.1	12.9	30.4	34.1
>$250K	16.4	1.5	3.5	6.5	24.8	18.7
N	384	65	58	31	230	91
Music fan:						
1 (low)	6.3	6.2	6.3	3.0	6.8	11.1
2	9.2	9.2	10.9	6.1	9.2	10.1
3	38.4	47.7	43.8	30.3	35.6	33.3
4	29.6	18.5	12.3	36.4	33.2	29.3
5	16.5	18.5	15.6	24.2	15.2	16.2
N	412	65	64	23	250	
Dial-up:						
1999	70.4	85.2	56.5	N.A.	73.2	76.0
2000	63.1	82.1	60.3	N.A.	60.9	59.8
2001	49.8	77.8	58.7	N.A.	42.0	29.9
2002	31.5	63.0	47.6	N.A.	20.6	20.6
2003	16.2	53.6	32.3	N.A.	4.1	4.1
Present	15.2	57.1	30.6	N.A.	2.4	3.2
High-Speed:						
1999	20.6	.0	29.0	N.A.	20.8	17.7
2000	31.9	7.1	28.6	N.A.	37.3	38.1
2001	47.1	18.5	31.7	N.A.	57.4	69.1
2002	66.5	33.3	46.0	N.A.	79.4	79.4
2003	83.8	46.4	67.7	N.A.	95.9	95.9
Present	86.0	42.9	69.4	N.A.	97.6	96.8
N	260	28	63		170	97
Number of CDs in collection	103.3	157.9	147.5	96.4	79.0	83.6

NOTE.—Data are percents unless otherwise specified. N.A. = not available because the survey administered at Hunter College did not elicit information about connection speed.

TABLE 3

AVERAGE VALUE OF HITS AND CURRENT PURCHASES

	HIT			CURRENT		
	Mean ($)	Median ($)	*N*	Mean ($)	Median ($)	*N*
Survey 1 (buy), ex post:						
Bought	12.70	12	2,726	15.25	15	695
Downloaded	8.81	8.5	1,512	11.15	10	337
Survey 1 (sell), ex post:						
Bought	56,344	100	702	1,831	100	173
Downloaded	2,534	30	264	710	50	64
Survey 2, ex ante:						
Bought	15.91	15	617			
Downloaded	10.66	10	592			
Survey 2, ex post:						
Bought	13.39	13	617			
Downloaded	10.47	10	592			

module, and responses to this module provide valuations on a common set of albums across all respondents. Second, we ask respondents to list all albums they have obtained in the past year. For each album they list, we ask them how they obtained it (via purchase, downloading, or as a gift) as well as their valuation. We term this the "current" module.

Two different versions of the valuation question were used initially. The "buy" set asked respondents, "Imagine you no longer have the album and must pay to get it. What is the maximum you are willing to pay to get the album?" In the alternative "sell" treatment, the question is worded, "Imagine someone offered to pay you money to permanently give up the album (and never again hear any of the songs on it). For each album, how much compensation would you require never to hear it again?" We employed the buy formulation for 337 of 412 surveys.

We use the survey data in two ways. First, we can aggregate them to the respondent level to create a cross-respondent data set on the volume of CDs purchased and downloaded. Although we do not know when respondents purchased each CD, we do know the CDs' "vintages," that is, the years in which they were first certified. This allows us to create a second panel data set on the numbers of albums downloaded and purchased, among those of each vintage. Third, we can use the data at the album level.

Table 3 reports mean and median valuations of purchased (including gifts) and downloaded hit and current albums using both the buy and sell valuation methods. The first point to note is that sell-based valuations are enormous. The average purchased hit is valued at over $50,000. While the means are sensitive to outliers, even the median sell valuations—$100 per album for hits and current albums—are rather high. The buy-based valuations appear to be more reasonable. Of 2,726 purchased hits, the mean valuation is $12.70,

while the mean valuation of a purchased current album is $15.25. Second, downloaded albums are valued substantially below purchased albums.

Although more reasonable in the sense of being within an order of magnitude of purchase prices, the survey 1 buy valuation data are still, at first blush, puzzlingly low. Of 2,726 purchased albums valued via the buy method, the mean valuation was $12.70. On its face this seems to contradict logic, since a person should buy a product only if the value equals or exceeds the price. And, indeed, 58 percent were valued below $15. Moreover, 28 percent were valued below $10. These results raised concerns for us that music may be (1) an experience good, (2) subject to depreciation as listeners grow tired of music, or (3) both.

Low ex post valuations are not just a theoretical curiosity; they also prevent using survey 1 valuation data to determine which downloaded albums would have been purchased absent downloading. It seems natural to assume that individuals would have purchased albums when $v > p$. Yet it is not true that v exceeds p for all, or even most, purchased albums.

Our concern about music as an experience good led us to administer survey 2 to 92 Penn students, asking about both ex ante and ex post valuations of the 261 albums in the hit sample, as well as the individual characteristics covered in survey 1. Our valuation instructions on this survey were rather explicitly designed to elicit ex ante valuations at or above the prices paid.

> *Initial Valuation.* *At the time you obtained the album,* what is the maximum you would have been willing to pay for it? *If you paid for it, you must have valued it at least as much as its price initially. For this question, suppose that there is only one possible source. That is, your answer should not be based on an alternative price or sharing opportunity.* (Don't say you're not willing to pay since you know you can obtain it via sharing for free or that you are not willing to pay more than some particular amount since you know you can buy from another seller for some particular price. Here, you are asked to assume no such alternative sources exist.)
>
> Finally, if you would have been willing to purchase it at the time you downloaded it or received it as a gift, report a valuation at or above the going price at that time.
>
> *Current Valuation.* *Now that you've had the music for a while,* what is the maximum you would be willing to pay for it? *Here, again, suppose that there is only one possible source so your answer, as above, should not be based on an alternative price or sharing opportunity.*

The last column of Table 2 reports characteristics of the individuals in the survey 2 sample. The lower part of Table 3 reports ex ante and ex post valuations of hits from survey 2. Here, the mean ex ante valuation is nearly $16, while the mean ex post valuation, at $13.39, is very similar to the (ex post) valuation in survey 1. Two points merit discussion. First, the ex post valuation is below the ex ante, which indicates depreciation. Second, the correlation of ex ante and ex post value is only about .6, which indicates that most of the variation in ex post value is realized after purchase, so the experience good aspect of music is important.

TABLE 4

PURCHASES AND DOWNLOADS

	HIT			CURRENT		
	Purchase	Download	*N*	Purchase	Download	*N*
All	8.24	5.36	365	2.61	1.24	349
Asian	6.74	7.98	108	1.98	1.16	101
Black	9.53	2.34	32	3.34	1.31	32
Hispanic	11.62	4.83	37	2.85	1.03	33
White	8.31	4.55	184	2.75	1.35	179
City College of New York	10.96	3.24	54	2.90	.63	52
University of Chicago	8.21	2.25	56	2.82	.54	57
Hunter College	11.13	1.72	32	3.96	.56	27
University of Pennsylvania	7.17	7.17	365	2.31	1.67	213

We supplement the survey data with information on aggregate sales data on all of our hits—as well as roughly 8,000 other hit albums certified by the RIAA—from the RIAA Web site.[3] The source reports individual certifications. An album is certified "gold" when it has sold .5 million units and "platinum" when it sells 1 million, and it receives a multiple platinum certification for each additional million copies sold. An album released in, say, 1985, can continue to sell in subsequent years. We attribute the additional sales leading to most recent certification to the year of the current certification. This makes the certification-based data compatible with the RIAA aggregate sales data (except that the aggregate data also include albums selling fewer than .5 million copies). The RIAA also provides data on annual units sold and revenue. Figure 1 shows the overall and certified album sales. Albums certified by the RIAA account for roughly three-quarters of overall sales.

Table 4 presents the average number of purchases and downloads, in the current and hit samples, overall and by race and school.[4] Respondents report an average of 7.8 hit purchases and 5.1 hit downloads (for albums first certified in 1999 and by late 2003 selling at least 2 million copies). Blacks download fewer albums than others in the sample. Penn students in our sample download substantially more albums than Hunter, CCNY, or Chicago students (7.2 hit albums as opposed to 2–3). Interestingly, Penn students purchase fewer albums than the others, which suggests sales displacement.

We can compare our self-reported purchase information with aggregate data on album sales. According to the RIAA, U.S. album sales in 2002 totaled $11.233 billion. Of this total, 11.5 percent, or $1.292 billion, was sold to persons aged 20–24. According to the 2000 census, there were 18.964 million persons in this age group in 2000. Dividing, we get average annual expen-

[3] RIAA, Gold and Platinum (http://www.RIAA.com/gp/default.asp).

[4] In this paper we group albums received as gifts with purchases. All exercises reported in the paper were also performed excluding gifts from purchases. No substantive results change.

diture of $68 per person aged 20–24. Given the price of CDs, this translates to just under five albums per person. Our respondents, by contrast, report purchasing an average of 2.6 albums in the past year.

We can also examine the share of albums obtained via download by date of first certification, 1999–2003. The number obtained declines over time because albums released later have fewer years of sales exposure. Roughly a third of hit albums released in 1999 are obtained by downloading. The share rises to nearly half for albums released during 2002–3.

V. Sales Displacement

A. *Does Downloading Reduce Album Sales?*

Aggregating our survey data to the respondent level creates data on album purchases and downloads for 412 individuals. Of these, 364 have valid data on all variables needed for basic analysis of the hits. Our first pass at the sales displacement question is through cross-sectional regressions of the number of albums purchased on the number of albums downloaded and a host of controls. In particular, we estimate the following equation:

$$P_i = X_i\beta + \alpha D_i + \varepsilon_i, \tag{1}$$

where P_i is individual i's purchases of albums from some set (either the hit or current), D_i is individual i's downloads of albums (again from some set), X_i includes characteristics of the individual such as his or her level of interest in music, race, and income, ε_i is unobserved characteristics, and α and β are coefficients to be estimated.

Columns 1–3 of Table 5 report estimates of equation (1) using the entire hit album sample. In this equation, P is the total number of albums on the hit list (those albums selling 2 million or more between 1999 and 2003) that the individual has purchased, while D is the number of those albums that the person has obtained via downloading. The first column shows that the estimated sales displacement coefficient α using the full time period is insignificantly different from zero. The result does not change if we control for school (column 2) or estimate a Tobit rather than an OLS model (column 3). Columns 4–6 present estimates of equation (1) using the current sample. The OLS regressions in columns 4 and 5 give displacement estimates of $-.17$ to $-.19$, changing little when we control for school. Column 6 reports a Tobit estimate on the current sample, yielding a negative and significant displacement estimate.[5]

In interpreting these estimates, it is important to keep in mind that broadband Internet access stood at low levels for our respondents in 1999. If a

[5] Because purchases are positive for 243 of 347 observations, the associated displacement estimate is $-.18$.

person is not connected, there is no scope for downloading to displace her purchases. Many, if not most, of the persons in the sample purchased 1999 releases prior to having the ability to download. The sales displacement question is addressed cleanly only when people have the ability to download albums at the time they might otherwise have purchased them. These considerations suggest that recent years would be most useful for detecting possible displacement and explain why we see displacement in the current estimates but not for the 1999–2003 hits overall. When we estimate equation (1) only on the hits first certified in 2003, we find displacement of $-.09$ ($p = .10$, two-sided).

The estimates in Table 5 are vulnerable to the concern that downloading is endogenous, or at least that it is subject to unobserved heterogeneity. We might expect music lovers to both download and purchase more music. One way to deal with this is simply to recognize that it would lead OLS regressions to underestimate the amount of sales displacement and to view our coefficients as underestimates of true sales displacement. Alternatively, one can try to find instruments for music downloading that are not themselves related to music demand. The speed of one's access to the Internet can serve as such an instrument, provided that connection speed is exogenous to interest in music. It would undermine this strategy, for example, if people chose high-speed connections because of their interest in music.

We have two approaches to using access speed as an instrument for downloading. First, we can use the school the student attends, and second, for a subsample, we can use whether she has broadband access. Far more than college students at Hunter or CCNY, or M.A. students at Chicago, Penn students live in dormitories with high-speed Internet access. "Only 612 of Hunter's 17,000 students have the opportunity to live in the Residence Hall."[6] And there are no dormitories at CCNY.[7] By contrast, there is space for "roughly 5,600" of the nearly 10,000 undergraduates at Penn.[8] Our own survey shows this: broadband access is ubiquitous among Penn students and far less common among others. Our two instrumental variables (IV) strategies are (1) to use which school they attend to instrument downloading and (2) to use the survey's individual measure of broadband access as an instruments.

The first four columns of Table 6 present first- and second-stage regressions on the hit sample. The first two columns report first-stage (download) regressions, using school attended and the individual broadband access variables, respectively, as instruments for downloading. As one would expect from the raw data in the tables, Penn students download more than other

[6] Hunter: Brookdale Residence Hall (http://studentservices.hunter.cuny.edu/rhhome.shtml/index.html).

[7] About the City College of New York (http://www.ccny.cuny.edu/bulletin02/about_usU.htm).

[8] Dina Matkevich, University of Pennsylvania Housing and Conference Services, e-mail to the author, March 30, 2006.

TABLE 5

ORDINARY LEAST SQUARES (OLS) AND TOBIT SALES DISPLACEMENT ESTIMATES

	HIT PURCHASES (N = 364)			CURRENT PURCHASES (N = 347)			2003 HIT PURCHASES (N = 364):
	OLS (1)	OLS (2)	Tobit (3)	OLS (4)	OLS (5)	Tobit (6)	OLS (7)
Current downloads				−.1945**	−.1722*	−.2579*	
				(.0718)	(.0728)	(.1001)	
Hit downloads	−.0266	−.0085	−.0310				−.0910
	(.0521)	(.0527)	(.0598)				(.0556)
Male	−3.3089**	−3.5692**	−4.4627**	−.6344*	−.6188	−1.0338*	
	(.9708)	(1.0173)	(1.1373)	(.3167)	(.3306)	(.4419)	
Music fan:							
2	4.0584	3.7992	8.0407**	.1022	.0572	1.1415	.3705
	(2.4907)	(2.4789)	(2.9994)	(.8161)	(.8161)	(1.2328)	(.2494)
3	5.6050**	5.3117*	9.2048**	.9109	.8855	2.4968*	.5430*
	(2.1314)	(2.1219)	(2.6438)	(.6982)	(.6979)	(1.0789)	(.2128)
4	6.7212**	6.4736**	10.6962**	1.9665**	1.9228**	3.9187**	.6201**
	(2.1738)	(2.1652)	(2.6787)	(.7144)	(.7141)	(1.0955)	(.2173)
5	9.9391**	9.5997**	14.1591**	2.6707**	2.5686**	4.8472**	.6221**
	(2.3228)	(2.3144)	(2.8243)	(.7518)	(.7527)	(1.1391)	(.2327)
$25K–$50K	2.2034	1.6273	2.0517	.5603	.4071	.4576	.0848
	(1.5629)	(1.5824)	(1.7785)	(.4988)	(.5089)	(.6768)	(.1564)
$50K–$100K	1.0381	1.1821	1.5687	.2371	.2931	.1194	−.0025
	(1.3790)	(1.4304)	(1.6202)	(.4659)	(.4823)	(.6504)	(.1373)

$100K–$250K	.7783	1.3815	2.2364	.3413	.5042	.5354	.0577
	(1.4597)	(1.5420)	(1.7435)	(.4827)	(.5088)	(.6845)	(.1450)
>$250K	2.6972	3.4626*	4.3480*	.4111	.6026	.5559	.0485
	(1.6185)	(1.7117)	(1.9276)	(.5340)	(.5647)	(.7550)	(.1606)
Black	1.7491	.3582	.6459	.3397	.2338	.4188	.7495**
	(1.7698)	(1.8678)	(2.0700)	(.5682)	(.5923)	(.7799)	(.1768)
Hispanic	1.7620	−.0623	.0169	−.2467	−.5010	−.2740	.2155
	(1.6841)	(1.8210)	(2.0131)	(.5553)	(.6063)	(.7901)	(.1689)
Asian	−.8649	−.9258	−1.6524	−.7525*	−.6987	−1.0400*	.2989**
	(1.1388)	(1.1535)	(1.3002)	(.3611)	(.3667)	(.4964)	(.1126)
University of Chicago		−3.5870	−3.9167		−.3204	−.5408	
		(1.9267)	(2.1508)		(.6088)	(.8167)	
Hunter College		−1.4116	−1.3880		.4883	.6874	
		(2.0773)	(2.3001)		(.7008)	(.9171)	
University of Pennsylvania		−4.1112*	−4.5066*		−.6275	−.5965	
		(1.6526)	(1.8442)		(.5351)	(.7167)	
Constant	2.8597	6.4934*	2.1786	1.7047*	2.0768*	.0260	−.1857
	(2.2992)	(2.8164)	(3.3257)	(.7482)	(.9027)	(1.3107)	(.2214)
R^2	.12	.14		.14	.15		

NOTE.—Standard errors are in parentheses. OLS = ordinary least squares.
* Significant at the 5% level.
** Significant at the 1% level.

TABLE 6

INSTRUMENTAL VARIABLES ESTIMATES OF HIT AND CURRENT SALES DISPLACEMENT

	HIT				CURRENT			
	Download		Purchase		Download		Purchase	
SAMPLE	School (1)	Individual (2)	School (3)	Individual (4)	School (5)	Individual (6)	School (7)	Individual (8)
Current downloads							−.8804	−1.4676
							(.4570)	(.8502)
Hit downloads			−.5719	−.2563				
			(.3334)	(.4415)				
Broadband access		3.9311				1.0707*		
		(2.0540)				(.5222)		
Male	.1042	.4737	−2.9788**	−3.0819*	.5737*	.9173**	−.2050	.4658
	(1.0355)	(1.3606)	(1.1299)	(1.1891)	(.2475)	(.3465)	(.4551)	(1.0341)
Music fan:								
2	3.3970	3.9291	5.8858	6.4379*	.6286	.7874	.4870	.7147
	(2.5165)	(3.3562)	(3.0582)	(3.1422)	(.6149)	(.8373)	(.9552)	(1.5463)
3	6.0806**	7.5175**	8.8921**	6.8162	.7307	.9770	1.3940	2.3849
	(2.1350)	(2.7380)	(3.1421)	(3.9991)	(.5251)	(.7038)	(.8494)	(1.4629)
4	5.8373**	6.7521*	9.8673**	8.7015*	.8491	1.0916	2.5449**	3.2605*
	(2.1814)	(2.8182)	(3.1280)	(3.7107)	(.5368)	(.7232)	(.8912)	(1.5214)
5	4.8969*	5.3952	12.4919**	9.9086**	.8452	1.3516	3.1924**	4.8208**
	(2.3410)	(3.0554)	(3.0726)	(3.4224)	(.5661)	(.7707)	(.9150)	(1.7036)
$25K–$50K	.6323	1.1859	2.3412	3.5615	.1678	−.1327	.6081	−.0888
	(1.6102)	(2.2458)	(1.7926)	(1.9947)	(.3839)	(.5638)	(.5640)	(.9854)
$50K–$100K	1.3031	1.8318	2.2499	1.5951	.5565	.4007	.7611	1.0134
	(1.4543)	(1.9315)	(1.7400)	(1.9552)	(.3627)	(.5105)	(.6282)	(1.0181)

$100K–$250K	1.6783	.9198	2.5392	.1394	.6131	.4236	1.0042	.9678
	(1.5669)	(2.0122)	(1.9796)	(1.8480)	(.3825)	(.5225)	(.6970)	(1.0653)
>$250K	.9148	.4349	4.1691*	3.4247	.6004	.2875	1.1005	.6888
	(1.7415)	(2.2544)	(2.0549)	(1.9585)	(.4248)	(.5968)	(.7535)	(1.1256)
Black	−.4683	.1327	1.0382	.8998	.6541	.8319	.6239	.3569
	(1.9010)	(2.5156)	(2.0723)	(2.1152)	(.4455)	(.6096)	(.6679)	(1.2014)
Hispanic	2.2792	4.3455	2.3844	4.9284	.4618	.9973	−.1939	1.1861
	(1.8495)	(2.6193)	(1.9655)	(2.5741)	(.4568)	(.6604)	(.6278)	(1.2501)
Asian	3.4776**	5.8701**	1.3554	.7168	−.0192	.1463	−.7229	−.9885
	(1.1592)	(1.5129)	(1.8671)	(2.8829)	(.2767)	(.3935)	(.4081)	(.6960)
University of Chicago	.0573				.4089			
	(1.9611)				(.4588)			
Hunter College	−1.3440				.0841			
	(2.1131)				(.5288)			
University of Pennsylvania	3.6365*				1.0906**			
	(1.6707)				(.3994)			
Constant	−4.1479	−6.7023*	1.2116	1.7005	−1.0581	−1.4387	1.3955	1.1489
	(2.8580)	(3.3334)	(2.8147)	(2.9569)	(.6787)	(.8401)	(.8686)	(1.3918)
N	364	231	364	231	347	219	347	219

NOTE.—The dependent variables are download and purchase. The instruments are school and individual. Standard errors are in parentheses.
* Significant at the 5% level.
** Significant at the 1% level.

TABLE 7

LONGITUDINAL SALES DISPLACEMENT ESTIMATES: HIT SAMPLE

	Purchases from Year: OLS (1)	Purchases from Year All: FE (2)	Purchases from Year > 5: FE (3)
Downloads from year	−.0447 (.0458)	−.0789* (.0334)	−.1565* (.0656)
Year:			
2000	.0527 (.1138)	.0656 (.1137)	.2061 (.2113)
2001	−.6141** (.0995)	−.6110** (.1131)	−.9055** (.2096)
2002	−1.1111** (.1185)	−1.1141** (.1131)	−2.0608** (.2096)
2003	−1.8747** (.1291)	−1.8987** (.1154)	−3.5251** (.2135)
Music fan:			
2	.9301** (.2558)		
3	1.2497** (.2035)		
4	1.4370** (.2221)		
5	2.0663** (.3374)		
$25K–$50K	.3723 (.2932)		
$50K–$100K	.1326 (.2475)		
$100K–$250K	.0702 (.2776)		
>$250K	.3923 (.2710)		
Black	.4463 (.3307)		
Hispanic	.4105 (.3766)		
Asian	−.0662 (.2273)		
Constant	.8397** (.2218)	2.4469** (.0885)	4.4216** (.1625)
N	1,820	1,820	820
R^2	.14	.22	.39
Number of individuals		364	164

NOTE.—Robust standard errors are in parentheses. OLS = ordinary least squares; FE = fixed effects.
* Significant at the 5% level.
** Significant at the 1% level.

students, after accounting for race, income, and interest in music. Columns 3 and 4 report second-stage hit displacement regressions, and the estimated hit sales displacement coefficients are −.25 and −.57. Columns 5–8 repeat the exercise using the current sample, and, again, individuals with broadband access download more, and the estimated current sales displacement coefficients are −.88 and −1.47. The estimated IV displacement coefficients in Table 6 are much larger than their OLS analogs in Table 5, although they are also less precisely estimated.

The panel structure of our data allows another empirical approach. Table 7 uses the panel data, in which a unit of observation is a person by vintage for the hit sample. We estimate equations of the form

$$P_{it} = X_i\beta + \alpha D_{it} + \phi_t + \varepsilon_{it}.$$

Thus we run regressions of the number of albums released in each year that an individual purchased on the number of that year's albums that the person downloads, along with time dummies (ϕ_t) and our various controls. These pooled cross-section time-series estimates give sales displacement coeffi-

cients of roughly −.04, but they are not significant. The second column of Table 7 incorporates individual fixed effects, identifying the sales displacement coefficient from the relationship between changes in the tendency to download and changes in the tendency to purchase, over and above the common vintage pattern. That is, we make the substitution $\varepsilon_{it} = \mu_i + \nu_{it}$, where μ_i is an individual-specific fixed effect and ν_{it} is an individual- and year-specific error. Here we obtain a coefficient estimate of −.08, and it is statistically significant. When we restrict attention to persons purchasing more than five albums from the hit sample, we get a coefficient twice as large (−.15 and significant).

What do all of these estimates mean? Estimates based on data for periods when downloading is feasible suggest sales displacement of roughly −.2 per album downloaded in the OLS specifications and much higher values in the IV estimates. Average current-year downloads—see Table 4—were 1.24, while current purchases averaged 2.61. Using the −.2 displacement estimate, in the absence of downloading, purchases would have averaged 2.86, which suggests that downloading reduced purchases by individuals in the sample by about 9 percent.

B. Do Consumers Download High- or Low-Value Music?

Before analyzing the valuation data, it is worthwhile to discuss the virtues and shortcomings of survey-based valuation information. Economists have traditionally been skeptical of surveys of valuation because a respondent's valuation of an object he has never thought about—or never will see—may not have any meaning (Diamond and Hausman 1994). In our context, however, we are asking about familiar objects. Indeed, we are in particular asking about objects that the respondents actually possess. Moreover, these are objects of a sort the respondents have obtained repeatedly, so their valuation responses may be more meaningful than responses about unfamiliar and largely hypothetical goods.

Hence, we are willing to put aside qualms about survey valuation data to ask whether downloading is a means of obtaining high- or low-valuation albums. To this end, Tables 8 and 9 report album-level regressions of log valuation on a downloading dummy. Table 8 reports regressions of log (ex post) valuation on a downloading dummy, using data from survey 1. Only buy valuation data are included in these regressions (to ensure comparability with survey 2 results). The first column shows that downloaded hits are valued 33 percent below purchased hits. The second column shows that downloaded current albums are valued 39 percent less than purchased current albums. Standard errors in all regressions are adjusted for clustering on individuals. Columns 3 and 4 repeat the exercise with individual fixed effects, showing not only that downloaded albums are less valuable than purchased albums but also that this is true within individuals.

TABLE 8

RELATIVE VALUATION OF DOWNLOADED CDs: SURVEY 1, EX POST VALUATION (Log Values) (N = 1,133)

	ORDINARY LEAST SQUARES		FIXED EFFECTS	
	Hit (1)	Current (2)	Hit (3)	Current (4)
Downloaded	−.330** (.072)	−.390** (.089)	−.177** (.023)	−.238** (.041)
Constant	2.382** (.037)	2.605** (.038)	2.331** (.011)	2.557** (.018)
N	4,017	1,011	4,017	1,011
R^2	.05	.08	.02	.04
Number of individuals			284	226

NOTE.—Robust standard errors are in parentheses..
** Significant at the 1% level.

TABLE 9

RELATIVE VALUATION OF DOWNLOADED CDs: SURVEY 2, EX ANTE AND EX POST VALUATIONS, HITS ONLY (Log Values) (N = 1,133)

	ORDINARY LEAST SQUARES		FIXED EFFECTS	
	Ex Ante (1)	Ex Post (2)	Ex Ante (1)	Ex Post (2)
Downloaded	−.4623** (.0822)	−.2654* (.1043)	−.2857** (.0283)	−.0857 (.0511)
Constant	2.7104** (.0348)	2.3879** (.0765)	2.6253** (.0175)	2.3013** (.0317)
R^2	.15	.02	.09	.00
Number of individuals			93	92

NOTE.—Robust standard errors are in parentheses.
* Significant at the 5% level.
** Significant at the 1% level.

Table 9 performs a similar exercise on the ex ante and ex post hit valuation information from survey 2. As in survey 1, the ex post valuations of downloaded albums are nearly a third (27 percent) less than the valuations for purchased albums. The ex ante valuations of downloaded albums, relative to purchased albums, however, are even lower than the ex post valuations (46 versus 27 percent). At the time of obtaining the music, then, persons expect the albums they download to be much less valuable to them than purchased music. This suggests that they would not have been willing to pay much for this music and that they would not likely have purchased it. But the coefficient estimate describes only the average valuation difference between downloaded and purchased CDs. A more thorough analysis of downloading requires information about the full structure of ex ante valuation (for demand) and ex post valuation (for welfare).

TABLE 10

EX ANTE AND EX POST VALUATIONS

	Mean	Ex Post Valuation (1)	Ex Post − Ex Ante Valuation (2)
Ex ante valuation	$13.34	.8178** (.0183)	
Ex post valuation	$11.94		
Ex ante and ex post correlation	.6389		
Pleasantly surprised	16.4%	4.2537** (.4657)	
Grew on me	16.5%	4.7717** (.4656)	
Familiar before I bought	21.8%	.2707 (.4921)	
Guessed right	11.5%	.1277 (.5110)	
Disappointed from start	9.5%	−5.5276** (.5782)	
Grew tired of it	31.7%	−5.2951** (.5028)	
Released in 1999			−3.1906** (.7635)
Released in 2000			−1.7907* (.7534)
Released in 2001			−1.7723* (.7931)
Released in 2002			−.0587 (.8038)
Constant		1.6733** (.5214)	.3375 (.6831)
R^2		.74	.03

NOTE.—Standard errors are in parentheses. N = 1,209.
*Significant at the 5% level.
**Significant at the 1% level.

C. *Ex Ante and Ex Post Valuation of Music*

This section provides some descriptive characterization of the ex ante and ex post valuation data. For each album, we asked for an ex ante valuation, a price (if they purchased the album), an ex post valuation, as well as indicators for why the current valuation deviated (from the initial valuation). If the current valuation exceeded the initial valuation, we allowed respondents to indicate (1) that they were pleasantly surprised, (2) that the music grew on them, or (3) both. If the current valuation equaled the initial, we allowed respondents to indicate that (1) they were familiar with the music before purchase, (2) they guessed right, or (3) both. Finally, if the current valuation fell short of the initial valuation, we allowed respondents to indicate that (1) they were disappointed from the start, (2) they grew tired of the music, or (3) both.

The data are described in Table 10. The average ex ante valuation is $13.34, while the average ex post valuation is $11.94 for the 1,209 albums in survey 2, which indicates depreciation. Respondents report higher ex post than ex ante valuations because of pleasant surprise at 16.4 percent of albums and that 16.5 percent of albums grew on them. Respondents report equal ex post and ex ante valuations: 21.8 percent of albums were familiar before purchase, and respondents guessed accurately on the value of 11.5 percent of albums. Finally, respondents report lower ex post than ex ante valuations for many albums: 9.5 percent of albums are disappointing from the start, while respondents grew tired of 31.7 percent of albums.

These data also allow us to explore the distributions of ex ante and ex post valuations and therefore what the determinants of purchase decisions and welfare look like. Figure 5 depicts the distribution of ex ante valuations, ordered from highest to lowest. The curve is, in effect, a demand curve, since points above any price indicate a willingness for an individual to purchase an album. It is worth noting that the demand curve does not appear to belong to any familiar parametric family. If ex ante valuations, which are guesses about how much the person will like an album, are not accurate, then the area under the demand curve does not represent welfare.

Figure 6 shows the distribution of ex post valuations. This curve is not relevant to purchase decisions, since the valuations are not known at the time of purchase. These valuations are relevant to welfare, however. The curves in Figures 5 and 6 look very similar, but one needs to bear in mind that they are ranked by different variables.

Figure 7 illustrates the distinction between ex ante and ex post valuation more directly; that is, it presents the information in a different way, with a plot of ex post valuation on ex ante valuation and a 45-degree line. This diagram too shows that, while ex ante and ex post valuations are correlated, the correlation, at .63, is far from perfect. Moreover, the preponderance of dots below the 45-degree line indicates depreciation overall.

Column 1 of Table 10 reports a regression of ex post valuation on ex ante valuation and dummies that reflect, for example, whether an individual has grown tired of an album. The coefficient on "grew tired of it" is negative and significant.

The true ex post valuation of an album is the sum over the flow utilities experienced, appropriately discounted. If the flow utility declines over time, then the current valuation of a 3-year-old album will understate its original ex post value. To address this, we examine average depreciation directly, via a regression of (ex post valuation − ex ante valuation) on dummies for year of release.[9] Albums released in 1999 are on average $3 less valuable than albums released in 2003, and the earlier the release, the greater the gap, which we interpret as depreciation. Consequently, we adjust ex post valuation by adding back the average depreciation depicted in column 2 of Table 10. By construction this equalizes average ex ante and ex post valuation.

D. Simulating the Effects of Downloading on Sales and Welfare

We now turn to measuring the welfare effects of downloading directly for a subsample. Given that we have data on ex ante and ex post valuations for respondents in the second Penn sample, we can calculate their expenditure, consumer surplus, and the number of albums purchased and downloaded

[9] We do not know the year the individual obtained the album, but we assume earlier acquisition of albums released sooner.

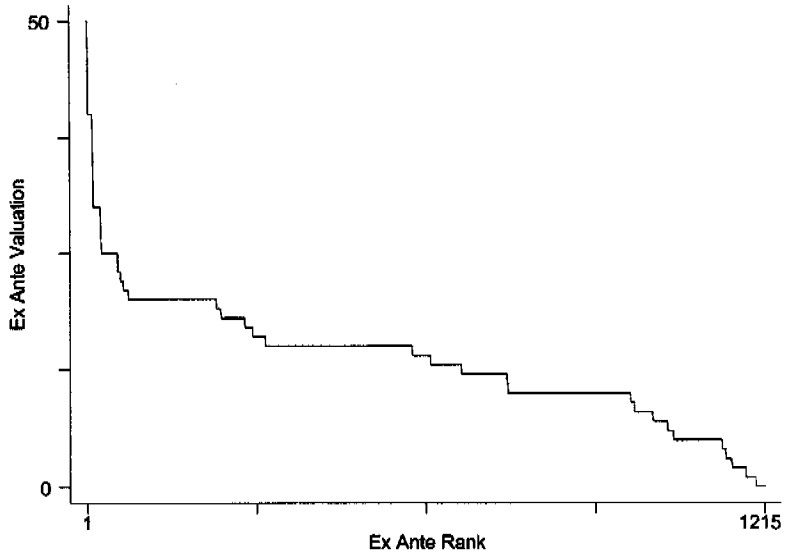

FIGURE 5.—Ex ante valuation: demand

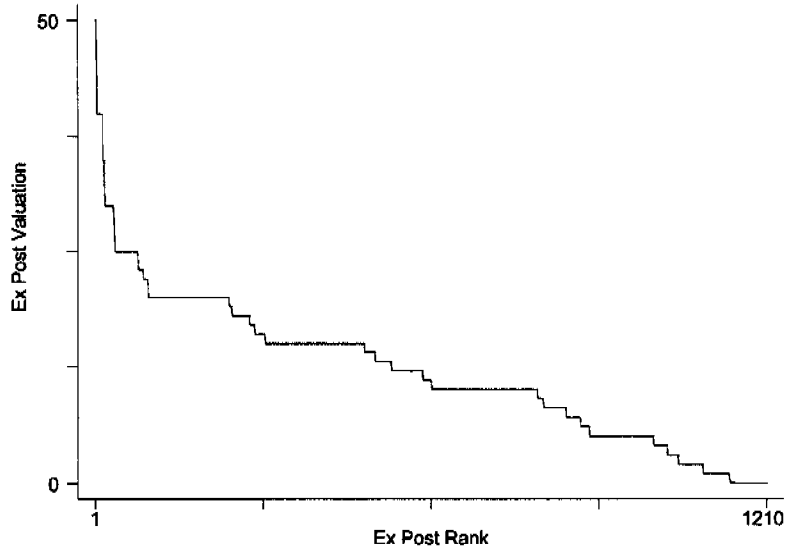

FIGURE 6.—Ex post valuation: welfare

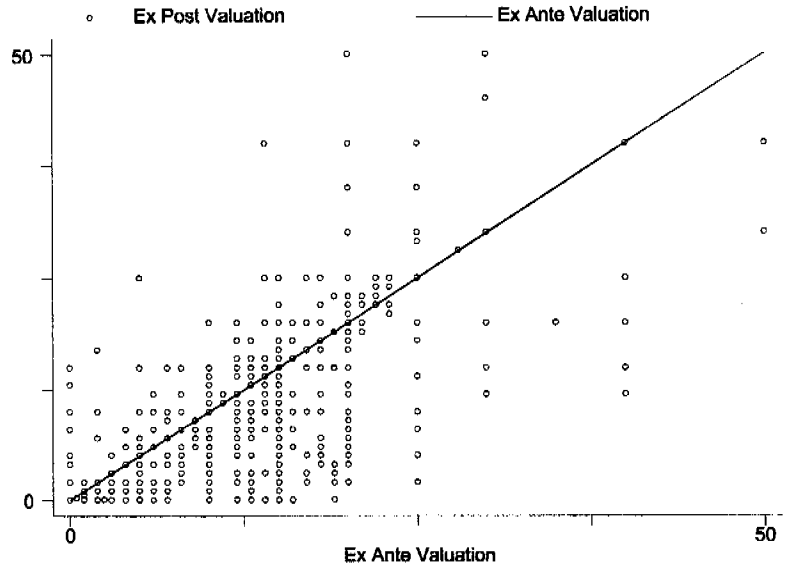

FIGURE 7.—Demand curve and ex post valuation

(in the downloading regime) straightforwardly according to the scheme in Table 1.

Assuming that p_i is \$15, all of the information required to simulate expenditure, consumer surplus, and deadweight loss is available in Tables 11 and 12.[10] Of the 1,209 albums in sample 2 obtained under the downloading regime, 617 albums were purchased, while 592 albums were downloaded by the 92 persons in the (ex post − ex ante) valuation subsample. The buyers paid \$15 per album, generating revenue of \$9,255. Buyers experienced ex post valuations of \$14.94 per album, giving rise to buyer consumer surplus of −\$37.[11] The 154 downloaded albums that would have been purchased are valued at \$17.91 each, generating \$2,738 in surplus. In conjunction with the 438 albums that would not have been purchased and that are ultimately valued at \$9.48 each, the total downloader surplus is \$6,910 and the overall consumer surplus under downloading is \$6,873.

If downloading were not feasible, the 617 purchased albums would still be purchased, as would 154 downloaded albums with ex ante valuations of

[10] We asked prices paid for albums in survey 2. The mean price paid was \$15.30. The simulation results do not change in important ways if we use different prices (for example, \$13, \$14, or \$16).

[11] Note that negative consumer surplus is ruled out by construction with ex ante valuations but is possible with ex post valuation realizations below prices paid.

TABLE 11

REVENUE AND WELFARE WITH AND WITHOUT DOWNLOADING

	N	Price	Ex Ante Valuation	Ex Post Valuation	Ex Post Valuation (Adjusted)
Did buy	617	15	15.91	13.34	14.94
Downloaded, but:					
Would buy if downloading were not feasible	154	15	17.82	16.72	17.91
Would not buy	438	N.A.	8.15	8.28	9.48

NOTE.—Based on 92 individuals in the survey 2 sample. We assume that the price of albums is $15. Welfare calculations are based on ex post valuation adjusted for depreciation. We assume that a downloaded album would have been purchased if the consumer's ex ante valuation were ≥$15. N.A. = not applicable.

$15 or more, generating revenue of $11,565 from this sample. Sales and revenue—since prices are assumed constant—are 20 percent lower with downloading than without it. Consumer surplus would include both the −$37 from those albums that were purchased along with $448 from the 154 albums that would have been purchased, generating $17.91 in ex post valuation on average, for a total of −$411 in consumer surplus. Finally, the sum of the ex post valuations of the albums that would not have been purchased (with ex ante valuations below $15) generates deadweight losses of $4,152.

In per capita terms, consumers spend $126 without downloading and $101 with downloading. Downloading increases consumer welfare by $70 per capita for sample individuals. Nearly two-thirds (64 percent) of this derives from the $45 per capita reduction in deadweight loss. The remaining 36 percent comes from the $25 per capita reduction in spending.

If we compute consumer surplus and deadweight loss using ex ante rather than ex post music valuations (see Table 12), the results are similar. Downloading raises consumer welfare for sample individuals by $64 per capita, and 60 percent of this results from the $39 per capita reduction in deadweight loss.[12]

[12] We administered a small-scale follow-up survey to address two questions left unanswered by the previous analysis, the relationship between song and album downloads and the tendency for people to purchase albums following downloading. Our surveys asked respondents to indicate whether they had obtained "albums" via sharing. As Table 3 indicates, Penn students report downloading roughly as many albums as they purchase (about 7.2). Yet music downloading is done song by song, which raises the question of what it means to download an album. To address this we administered a follow-up survey asking 25 Penn students how many songs they had obtained via sharing from each of the albums in the hit sample. For these students, the mean number of purchased albums is 6.96, just slightly below the full Penn sample average. The average number of albums from which they had obtained at least one song via sharing was 19.7, or more than double the number of "albums" they reported obtaining via sharing. On the other hand, they report downloading an average of only 4.0 entire albums. It appears, then, that including all albums from which they download six or more songs, they obtained 4.1 albums via sharing. Constructing successively more inclusive sharing measures, we obtain the following: including five or more songs yields 4.6, four or more yields 6.3, three

TABLE 12

REVENUE AND WELFARE WITH AND WITHOUT DOWNLOADING

	No Downloading		Downloading		Change	
	Total	Per Capita	Total	Per Capita	Total	Per Capita
Quantity sold	771	8.38	617	6.71	−154	−1.67
Quantity downloaded	0	0	592	6.43	592	
Quantity consumed	771	8.38	1,209	13.41	438	4.76
Revenue ($)	11,565	126	9,255	101	−2,310	−25.1
Ex post valuations:						
Buyer CS ($)	411	4.47	−37	−.40	−448	−4.87
Downloader CS ($)	0	0	6,910	75.1	6,910	75.1
CS total ($)	411	4.47	6,873	74.7	6,462	70.2
DWL ($)	4,152	45.1	0	0	−4,152	−45.1
Ex ante valuations:						
CS total ($)	996	10.82	5,880	74.73		
DWL ($)	3,570	38.80	0	0	−3,570	−38.80

NOTE.—Based on 92 individuals in survey 2 sample. We assume that the price of albums is $15. Welfare calculations are based on ex post valuation adjusted for depreciation. We assume that a downloaded album would have been purchased if the consumer's ex ante valuation were ≥$15. Deadweight loss in this table excludes forgone welfare from albums that consumers are unwilling to either purchase or download. See Table 1 and the surrounding text. CS = consumers' surplus; DWL = deadweight loss.

VI. CONCLUSION

We argue that successfully measuring the possible sales-displacing effect of unpaid music downloading requires data on the quantities of purchases and downloads made by individuals, which led us to conduct original surveys. Using a variety of empirical approaches, we document that downloading displaces sales among a convenience sample of college students. The estimate we consider most conservative indicates that an additional download reduces sales by between .1 and .2 units. As a result, for the individuals in our sample, downloading reduced expenditure by about 10 percent but possibly by much more. Supporting incomplete sales displacement is our finding that downloaded music is valued much less than purchased music.

While downloading reduces expenditure (on hit albums, 1999–2003) by $25 per capita in the subsample for which we perform a direct welfare analysis of downloading, it raises sample consumers' welfare associated with these albums by $70 per capita. Some of the benefit to consumers is transfers from sellers, but most of the benefit ($45 per capita) comes from reductions in deadweight loss.

Our results contrast somewhat with results of other recent studies. Like Zentner (2003) we use individual-level data, but unlike Zentner we find

or more yields 8.8, and two or more yields 14.3. It therefore appears that respondents reported obtaining an album if they have downloaded three or four (or more) of its songs. The follow-up survey also allows us to examine whether people download and then buy. Of the 476 albums—or album fragments—obtained via sharing, respondents subsequently purchased 21, or 4.4 percent.

evidence of displacement in OLS regressions. The difference may be due to the fact that we have continuous measures of album consumption, while his measures are binary. Our results contrast more sharply with Oberholzer and Strumpf (2004), who use a cross-album rather than cross-individual research design. How should one reconcile our results with theirs? In our view, unobservable determinants of music demand are likely to bias both approaches away from finding displacement. Just as albums that are popular to purchase are also popular to download, persons who like music may like to both download and purchase music. These possibilities would lead to positive relationships between downloading and purchases, both across albums and across individuals, even if downloading did not actually stimulate sales. Hence, while a finding of a positive relationship from either of these approaches might be dismissed as a residue of unobserved heterogeneity, we interpret our negative findings as indications that downloads displace sales.

Two facts bear emphasis again. First, our sample is not representative, so our results should not be generalized. Second, our evaluation of welfare takes supply as given. It is entirely possible that downloading has important effects on the quantity and types of music recorded and marketed in the first place. This is an important area for further research.

REFERENCES

Apple Computer. 2004. "iTunes Music Store Downloads Top 50 Million Songs." April 1. http://www.apple.com/pr/library/2004/mar/15itunes.html.

Bakos, Yannis, Erik Brynjolfsson, and Douglas Lichtman. 1999. "Shared Information Goods." *Journal of Law and Economics* 42:117–55.

Besen, Stanley M. 1986. "Private Copying, Reproduction Costs, and the Supply of Intellectual Property." *Information Economics and Policy* 2: 5–22.

Conner, Kathleen Reavis, and Richard Rumelt. 1991. "Software Piracy: An Analysis of Protection Strategies." *Management Science* 37:125–39.

Consumer Electronics Association. 2004. "2004 Sales of Consumer Electronics to Set New Record." January 5. http://www.ce.org/press_Room/press_release_detail.asp?id=10384.

Diamond, Peter A., and Jerry A. Hausman. 1994. "Contingent Valuation: Is Some Number Better than No Number?" *Journal of Economic Perspectives* 8:45–64.

Dixit, Avinash K., and Joseph E. Stiglitz. 1977. "Monopolistic Competition and Optimum Product Diversity." *American Economic Review* 67: 297–308.

Fader, Peter S. 2000. "Expert Report." *A&M Records, Inc. v. Napster, Inc.* 114 F. Supp. 2d 896.

Graves, Tom. 2004. "Movies and Home Entertainment." *Standard & Poor's Industry Surveys,* February 12.

Helmore, Edward. 2003. "Dance of Despair as Big Names in Music Merge." *Observer,* November 9.

Hui, Kai-Leung, and Ivan Png. 2003. "Piracy and the Legitimate Demand for Recorded Music." *Contributions to Economic Analysis* 2, art. 11, pp. 1–24. http://www.bepress.com/bejeap/contributions/vol2iss1/art11.

Liebowitz, Stan. 2003. "Will MP3 Downloads Annihilate the Record Industry? The Evidence So Far." Unpublished manuscript. University of Texas at Dallas, June.

Liebowitz, Stanley J. 1985. "Copying and Indirect Appropriability." *Journal of Political Economy* 93:945–57.

McCarthy, Michael, and Lorrie Grant. 2003. "Rivals Don't Quickly Follow Universal's CD Price Cuts." *USA Today,* September 5.

Oberholzer, Felix, and Koleman Strumpf. 2004."The Effect of File Sharing on Record Sales: An Empirical Analysis." Unpublished manuscript. Cambridge, Mass., Harvard Business School, January.

RIAA (Recording Industry Association of America). 2003. "Anti-piracy." http://www.riaa.com/issues/piracy/default.asp.

Schwartz, John J. 2004. "In Survey, Fewer Are Sharing Files (or Admitting It)." *New York Times,* January 5.

Shapiro, Carl, and Hal Varian. 1999. *Information Rules.* Cambridge, Mass.: Harvard Business School Press.

Spence, Michael. 1976a. "Product Selection, Fixed Costs, and Monopolistic Competition." *Review of Economic Studies* 43:217–35.

———. 1976b. "Product Differentiation and Welfare." *American Economic Review* 66:407–14.

Takeyama, Lisa. 1997. "The Intertemporal Consequences of Unauthorized Reproduction of Intellectual Property." *Journal of Law and Economics* 40: 511–22.

Varian, Hal. 2000. "Buying, Sharing, and Renting Information Goods." *Journal of Industrial Economics* 48:473–88.

Zentner, Alejandro. 2003. "Measuring the Effect of Online Piracy on Music Sales." Unpublished manuscript. University of Chicago, December.

Ziemann, George. 2002. "RIAA's Statistics Don't Add up to Piracy." December 11. http://www.azoz.com/music/features/0008.html.