

PIRACY ON THE SILVER SCREEN*

RAFAEL ROB†
JOEL WALDFOGEL‡

Using survey data on movie consumption by about 500 University of Pennsylvania undergraduate students, we ask whether unpaid consumption of movies displaces paid consumption. A variety of cross-sectional and longitudinal empirical approaches show large and statistically significant evidence of displacement. In the most appropriate empirical specification, we find that unpaid first consumption reduces paid consumption by about 1 unit. Unpaid second consumption has a smaller effect, about 0.20 units. Our analysis indicates that unpaid consumption, which makes up 5.2 per cent of movie viewing in our sample, reduced paid consumption in our sample by 3.5 per cent.

I. INTRODUCTION

THIS PAPER PROVIDES FIRST PASS EMPIRICAL ESTIMATES of the extent to which piracy is displacing sales in the movie industry.

Piracy (or ‘file sharing’) is the practice of obtaining an unpaid digitized product, in this case a movie, either by illegally downloading it from the Internet, or by illegally making a copy of the movie, using a DVD burner or a VCR. On the face of it, it seems that piracy induces a one-for-one displacement—a person who obtains an unpaid DVD of a movie, no longer has the need to buy one. Several considerations, along with specific features of the movie industry, suggest however that the issue of displacement is more complex and, consequently, that an empirical study is needed to quantify the average rate of displacement.

The first consideration is that the population of consumers (movie watchers) is heterogeneous with respect to their willingness to pay for a

*We thank seminar participants at the 2006 AEA Meetings in Boston, as well as the 2006 Conference on Economics of Information and Communication Technologies at ENST in Paris. Rob thanks the NSF, and Waldfogel thanks the Mack Center and the Wharton Electronic Business Initiative for research support. We are both grateful to Uri Spiegel for letting us administer surveys to his students, and to Liran Einav for useful comments. Sarah Waldfogel provided excellent assistance with data input. The usual disclaimer applies.

†Authors’ affiliations: Department of Economics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, U.S.A.
e-mail: rrob@econ.sas.upenn.edu.

‡University of Pennsylvania, The Wharton School, 1400 Steinberg Hall-Dietrich Hall, Philadelphia, Pennsylvania 19104-6372, U.S.A.; and NBER, 1050 Massachusetts Avenue, Cambridge, Massachusetts 02138, U.S.A.
e-mail: waldfogj@wharton.upenn.edu.

© 2007 The Authors. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics. Published by Blackwell Publishing Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK, and 350 Main Street, Malden, MA 02148, USA.

movie. If the willingness to pay of a consumer is above the market price of a movie and if she obtains this movie without paying for it, displacement is one-for-one: The consumer does not pay for a movie that she otherwise would have purchased (had a file sharing technology not been available), so the movie industry is indeed selling one less movie. On the other hand, if the consumer's willingness to pay is below the market price, the consumer obtains a movie that she would have not purchased in the first place, implying that displacement is zero. Since the willingness to pay is above the market price for some consumers and below the market price for others, the average rate of displacement is somewhere between zero and one. In a previous paper (Rob and Waldfogel [2006]) we analyzed the music industry, in which an analogous situation arises, and found that the average rate of displacement of music albums is 0.20, which is indeed between zero and one.¹

Another consideration is that the existence of file sharing technology may actually stimulate, rather than depress, sales when file sharing is non-anonymous and occurs within small groups, for example between friends or family members.² To see the logic behind this statement, consider two friends, both of whom have willingness to pay below the market price of a DVD, but with a combined willingness to pay above this price. Then, in the absence of file sharing technology, neither of them buys a DVD of the movie. On the other hand, with file sharing technology, they agree to buy one DVD between them and make a copy, which increases sales.³ This sales-stimulating effect of file sharing acts as an offset against displacement, implying again that the rate of displacement is below one. A theoretical model that spells out and explores the implications of this line of reasoning is Varian [2000].⁴

A third consideration is that the movie industry generates revenues (for a given movie) from several sources, which include box office sales (theatrical viewing), rentals, royalties for showing the movie on (cable or public) television, and DVD sales. Furthermore, some consumers pay more than once to view the same movie, e.g., watching it at the movie theatre, and then buying a DVD when one is released. This has two opposing effects on displacement. On the one hand, displacement is sometimes more than one-for-one (there is a 'multiplier' effect), since one unpaid copy may displace several modes of consumption. On the other hand, displacement is sometimes less than one-for-one, since home viewing is not a perfect substitute for theatrical viewing.

¹ See also Blackburn [2004]; Oberholzer-Gee and Strumpf [2007]; Zentner [2006]; Bounie, Bourreau and Waelbroeck [2005]; Hong [2005]; Hui and Png [2003]; and Leibowitz [2006].

² This is sometimes referred to as 'sneakernet': A user dons his sneakers and carries a DVD from his friend's house to his DVD-burning computer. See <http://en.wikipedia.org/wiki/Sneakernet>, accessed December 15, 2005.

³ The same holds for newspapers or magazine and, to a lesser extent, books.

⁴ For a related point see Bakos, Brynjolfsson and Lichtman [1999].

Given all these considerations, our purpose here is to estimate the average rate of displacement from data. The data we use to this end is survey data we collected from about 500 undergraduate students at the University of Pennsylvania.

The respondents in our data set are, to be sure, not representative of U.S. consumers generally. Nonetheless, there are reasons to believe that the data set includes individuals in a group highly relevant to the issue of movie piracy. Indeed, according to the Motion Picture Association, individuals aged 16–24, who make up 15 per cent of the U.S. population, accounted for 26 per cent of movie attendance between 2001 and 2005.⁵ And a ‘typical worldwide pirate is 16–24 years old, male, and lives in an urban area.’⁶ Further, the 16–24 ‘age range represents 71 per cent of downloaders’ and ‘44 per cent of MPA company losses in the U.S. are attributable to college students.’ Thus, while our sample is one of convenience, we believe that is of relevance to the issues we study here.

The analysis of this data set poses several challenges, the main one being that consumers differ not only with respect to their willingness to pay for a specific movie (as discussed earlier), but also with respect to the number of (paid and unpaid) movies they watch—their intensity of demand for movies in general. As we explain below, this creates a positive correlation between legally and illegally obtained movies, which offsets the ‘pure’ rate of displacement (associated with a *negative* correlation)—the parameter we are trying to identify. We do several empirical exercises that confirm the presence of this offsetting effect, and propose an approach to purge it from our estimating equations. Another issue we have to contend with is that there may be several episodes of consumption for a given consumer, so when we talk about ‘displacement caused by piracy’ we have to specify at which episode piracy occurs and which episode it affects. In the end, the estimate we obtain, using what we consider to be the most plausible econometric specification, is that the rate of displacement is 1.0 for movies that are pirated on the first episode of consumption. This rate of displacement is composed of 0.76 for contemporaneous first episodes of consumption, and 0.24 for forward second episodes. We also find that the rate of displacement for movies that are pirated on the second episode of consumption and that affect the second episode of consumption is 0.20.

All in all, our finding is that the rate of displacement for movies is far from being negligible, which has important implications for the future effect of piracy as P2P file sharing technology progresses, making it easier to transmit movie files over the Internet. To this point, the effect of piracy on movie

⁵ See MPA’s 2005 *U.S. Movie Attendance Study* (MPA [2005]).

⁶ See ‘The Cost of Movie Piracy’ (MPA/LEK [2005]).

revenues has been limited: Movie revenues continue to increase,⁷ which is in sharp contrast to music revenues.⁸ The apparent reason is that ‘compressed’ movie files are large (around 4 gigabytes), so that even with the newest downloading technology (Bit Torrent) it takes between 30 minutes and 8 hours to download a movie;⁹ by contrast, it takes few seconds to download a song. If this technological obstacle to downloading is lowered, however, then, given that a sizable fraction of movie revenues come from DVD sales and given that the rate of displacement for movies we find here is high (1.0 for the first episode of consumption), our estimation results suggest that movie revenues may greatly suffer.

The rest of the paper is organized as follows. In the next section we provide a brief industry background and discuss our data. In Section III we discuss empirical strategies and report our results; tables containing full details of the results are found at the end of the paper. Section IV compares the music and the movie industries, and Section V concludes.

II. INDUSTRY BACKGROUND AND DATA

II (i). *Industry Background*

Until 1980, box office revenue provided the majority of revenue to the U.S. motion picture industry. The box office share has declined steadily since then. In 2004, the U.S. film industry generated \$45 billion in revenue, of which 17% were box office revenue, and almost 50% were sales of home videos and DVDs.¹⁰ Between 2000 and 2004, overall industry revenue grew 39 per cent in real terms.¹¹ In contrast to the music industry, whose concerns about piracy are backed by sharp recent revenue declines, the movie industry continues to experience robust growth in revenue, driven largely by the growth in DVD sales. Concerns about piracy in the movie industry, while substantiated in many instances (see <http://www.mpaa.org/>), have not been reflected (so far) in overall industry revenue.

⁷ Notwithstanding, the Motion Picture Association of America estimates \$3.5 billion in losses to various forms of piracy in 2002. See the MPAA’s ‘2003 Piracy Fact Sheet Overview’ at <http://www.mpaa.org/PiracyFactSheets/PiracyFactSheetOverview.pdf>, accessed December 15, 2005.

⁸ See the RIAA’s ‘AntiPiracy’ at <http://www.riaa.com/issues/piracy/default.asp> as well as the RIAA’s ‘2004 Yearend Statistics’ at <http://www.riaa.com/news/newsletter/pdf/2004yearEndStats.pdf>, both accessed December 15, 2005.

⁹ ‘Cable and DSL modem users can expect an average of 25 megabytes per hour, sometimes slower if the swarm is small with less than 2 seeders. On a good day with a big swarm, however, you can download a 5MB song within 3 minutes, and a 900MB movie within 60 minutes.’ http://netforbeginners.about.com/od/peersharing/a/torrenthandbook_4.htm, accessed February 24, 2005.

¹⁰ See Edward J. Epstein’s website, ‘Hollywood by the Numbers,’ at <http://www.edwardjayeinstein.com/mpa2004.htm>, accessed December 13, 2005.

¹¹ See MPA [2005].

II (ii). *Data*

The data for this study are derived from a series of surveys administered to over 500 Penn undergraduates in economics classes in spring and late fall, 2005. In each survey, students see a list of the top 50 movies from each of the previous three years (2002–2004 in spring 2005, and 2003–2005 year to date in fall 2005). For each movie they are asked whether they saw it, as well as the sequence by which they saw it. Respondents can have seen a movie by any (and any combination) of 4 paid methods: *theater*, *television*, *rental*, and *purchase*; as well as by either or both of two unpaid methods: by viewing a *downloaded* copy, or by viewing a *burned* copy of a legally obtained copy (generally DVD). For each movie seen, a respondent enters a ‘1’ under the mode they first saw it, a ‘2’ under the mode by which they saw it next (if they saw it more than once), and so on.

In addition to asking about their movie consumption, the survey also asks about family income, respondent’s race and age, speed of Internet access, and three variables related to their interest in movies: how often they go to movie theaters, how many movies they own, and their (self-reported) level of interest in movies on a five-point scale. The sample includes nearly 21,000 instances of movie consumption. For analysis, we include only respondents who made no mistakes in reporting their sequences.¹² This brings the usable sample down to 470 individuals.

Before we report various statistics and analyze the data, it is worth repeating that the surveyed population is far from being representative of the U.S. population—respondents are richer and are more likely to have broadband Internet connections as well as access to DVD burning hardware. Nonetheless, and as stated above, the effects documented for this population are suggestive of effects that might operate more broadly as technologies continue to diffuse and, consequently, as more people own (or have access to) the same, or better, technology that respondents in our sample currently have.

The mean (median) age is 18.7 (18), and the age range is between 17 and 22. Nearly a quarter of the respondents are Asian, 6 per cent are black, and 6 per cent are Hispanic. Respondents are from families of above-average income: Nearly 30 per cent of the sample reports family income in excess of \$250,000. Another 37 per cent reports family income of \$100,000–\$250,000. Just 10 per cent of respondents report family income below \$50,000.

Nearly half (46%) report a ‘typical’ level of interest in movies. Nearly a third report above-average interest in movies. Only a fifth report below-average interest in movies. Respondents own a mean (median) of 38 (20) movies, and they go to movie theatres 1.6 (1) times per month. Internet

¹² We discard data for individuals who indicate that they saw a movie by multiple means but do not indicate the sequence (inputting, for example, a sequence of ‘1’s or check marks).

access is ubiquitous among the respondents: 98 per cent of respondents report high-speed access in 2005, up steadily from 70 per cent in 2002. Over half (54 per cent) own a DVD burner, a quarter have access to one, although they do not own it, and the remaining fifth have no access.

For the 150 movies in each survey, respondents report an average of 54.2 instances of paid consumption, and 3.0 instances of unpaid consumption. Theatrical consumption is the most common form, accounting for nearly half, followed by rental (20 per cent), television (16 per cent) and purchase (15 per cent).¹³ The two forms of unpaid consumption account for 5.2 per cent of consumption, the majority (3.1 of 5.2) burned and the remainder (2.1 of 5.2 per cent) downloaded.¹⁴

The seemingly low reported unpaid consumption provides some cause for skepticism about the veracity of survey answers to questions about illegal activity. Two points are in order. First, in analogous surveys about music downloading, similar respondents report much higher levels of unpaid consumption, roughly equal to their volumes of paid consumption. This suggests that information about this sort of activity is not obscured by respondents' inhibitions. Second, the relative quantities of paid vs. unpaid consumption in this sample corresponds, at least roughly, to comparable quantities in the aggregate data. The MPA claims \$1.2 billion in losses due to U.S. piracy in 2005, a year when domestic U.S. box office revenue equaled \$9 billion (MPA/LEK [2005]). Since box office revenues made up only a quarter of the sum of box office and DVD sales during the first quarter of 2004, we extrapolate an overall revenue estimate of \$40 billion dollars. And, relative to this ballpark figure, the rate of piracy, according to MPA sources, is 2–3 per cent, which is not far off the 5.2 per cent in our sample (so, if anything, the rate of piracy in our sample *exceeds* the rate of piracy in aggregate data).

Another noteworthy feature of our data is repeat consumption, i.e., the fact that movies are frequently viewed through more than one medium by a respondent. Table I breaks consumption down by medium and by viewing order. While respondents engage in 57.2 episodes of consumption, they see on average only 44.5 separate movies. Virtually all theatrical consumption occurs on first viewing; and theatrical viewing is the way that nearly 60 per cent of consumption sequences begin. The other major medium for first viewing is rental (8.3 of 44.5, or 19 per cent). The other media account for relatively few first viewings; on the other hand, first viewings via those media account for most of the viewings via those media. For example only 4 per

¹³ The frequencies of theatrical attendance are similar to aggregate averages. MPA estimates imply that persons aged 16–24 went to the movies an average of 11.2 times during 2005. Our respondents report an average of 25.5 theater visits to top-50 movies over three years, which is of the same order of magnitude as three times the average MPA figure.

¹⁴ Sample relative frequencies of burned vs. downloaded consumptions echo those in aggregate data. MPA/LEK [2005] reports that two thirds of U.S. piracy losses are due to 'hard goods piracy' (burned), while only a third are 'Internet piracy.'

TABLE I
VIEWING MODES ON NTH VIEWING

Viewing mode	All	First	Second	Third
Theater	25.5	25.4	0.1	0.0
Television	8.9	4.9	3.3	0.6
Rental	11.2	8.3	2.7	0.2
Purchase	8.5	4.0	3.8	0.6
Total paid	54.2	42.7	9.8	1.4
Downloaded	1.2	0.8	0.3	0.1
Burned	1.8	1.0	0.6	0.2
Total unpaid	3.0	1.8	0.9	0.3
Total	57.2	44.5	10.7	1.7

Notes: Column labeled 'All' shows number of viewings by each mode, along with total paid and total unpaid. Column labeled 'First' shows mode on first viewing.

cent (1.8 of 44.5) first viewings use unpaid copies. But these viewings represent most of the unpaid consumption. To put this another way, unpaid consumption is rare; but when it occurs, it usually occurs on first viewing.

The vast majority of second viewings are on purchased copies, television, or rental copies. As with first viewings, unpaid consumption accounts for a small fraction of second viewings. At the same time, over a quarter of unpaid consumption occurs on the second viewing.

The first column of Table II shows consumption sequences that begin with theatrical and rental consumption. The length of the sequences varies with the medium of initial consumption. Of the sequences beginning with rental, less than fifteen per cent are viewed through another (paid or unpaid) medium. Of the sequences that start in the theater, far more—over a third—continue to another medium, roughly 10 per cent each to purchase, rental, and television with another few per cent subsequently consumed via unpaid means.

III. EMPIRICAL STRATEGY, IMPLEMENTATION, AND RESULTS

III (i). *Generalities*

Before beginning the detailed analysis, it is helpful to distinguish two questions we ask (or, more precisely, two parameters we estimate). First, we ask what the *rate* of displacement is: how many movies are not purchased for each movie that is pirated? Second—if the rate of displacement is positive—we estimate the *amount* of displacement: how much paid consumption is displaced by unpaid consumption? This is calculated as the rate of displacement times the volume of unpaid consumption.

We have two basic approaches to estimating the rate of displacement. First, we can make straight use of our cross section data, asking whether people who engage in more unpaid consumption engage in less paid

TABLE II
CONSUMPTION SEQUENCES, OVERALL AND BY WHETHER ENGAGE IN UNPAID USE

	Overall	No Unpaid Use	Some Unpaid Use
<i>Following First Viewing in Theater</i>			
	11772	5449	6323
No subsequent	65.3%	65.1%	65.4%
Television	9.8%	10.7%	9.0%
Rental	9.7%	9.8%	9.7%
Purchase	12.8%	14.4%	11.4%
Downloading	0.8%	0.0%	1.5%
Burning	1.6%	0.0%	3.0%
Total paid	32.3%	34.9%	30.1%
Total unpaid	2.4%	0.0%	4.5%
<i>Following First Viewing through Rental</i>			
No subsequent	86.7%	87.5%	85.9%
Theater	0.1%	0.1%	0.1%
Television	6.4%	5.9%	6.9%
Purchase	4.5%	6.5%	2.8%
Downloading	0.6%	0.0%	1.1%
Burning	1.8%	0.0%	3.3%
Total paid	11.0%	12.5%	9.7%
Total unpaid	2.3%	0.0%	4.3%

Notes: Table shows modes of consumption following first viewing in theater (top panel) and rental (bottom panel).

consumption. Implementing this approach requires a way of predicting how many movies the respondents would have paid to consume in the absence of a technology that enables unpaid consumption. The main challenge to making this prediction is that persons who like movies may like both to pay for them and to view them without paying. If it is not possible to control for all determinants of individuals' interest in movies, then unobserved heterogeneity with respect to interest in movies induces a positive relationship between unpaid and paid consumption, even if the causal relationship is negative. This distorts the estimation results of the cross section approach.

The structure of our data allows, however, a second, panel approach. The data include individuals' movie consumption, by medium, yearly 2002(3)–2004(5). We can include individual fixed effects—which control for the individuals' vintage-constant interest in movies—then ask whether persons with higher unpaid consumption in some vintages have higher or lower paid consumption of movies from those vintages.

Our sequencing information allows us to ask further questions, which would have been impossible were we to lump the data together without regard to sequencing. Consider, for example, an individual who watches a movie first at the theatre and then downloads it from the Internet. If we were to ignore this sequencing, it would appear not only that paid and unpaid

consumption are positively related, but that unpaid consumption (on second viewing) 'causes' paid consumption (on first viewing). On the other hand, if we keep track of place in sequence we can specify estimating equations in which past behavior affects future behavior, but not the other way around. This allows us to estimate both contemporaneous and forward displacement. (More generally, this type of data allows one to learn about the temporal structure of consumption, e.g., whether different instances of consumption across time are complements or substitutes.) These are the questions to which we turn next.

III (ii). *Cross Sectional Approaches*

The cross sectional approach to the displacement question asks whether people with more unpaid viewings have fewer paid viewings. To this end we regress individuals' quantities of paid consumptions on their unpaid consumptions and other variables measuring their interest in movies. Unless our controls are exhaustive, unobserved heterogeneity is still a looming concern with models of this sort: people who watch movies without paying may simply like movies a lot, even beyond the capacity of our control variables to distinguish among persons in the sample. If so, then our estimate of the extent of displacement will be biased upward, that is, against a finding of sales displacement.

Table III reports results, overall and disaggregated by each paid medium. The first five columns are regressions of the number of first paid viewings in this medium on the number of first unpaid viewings. People with larger libraries, more frequent theater patronage, and higher levels of movie fandom have more paid first viewings. The coefficient on unpaid first viewings is negative but insignificantly different from zero. The second column examines theatrical first viewing. Here the coefficient of interest is -0.12 , below its standard error of 0.16 and therefore not statistically significant either.

Columns (6)–(10) are regressions of second viewings on controls and the numbers of unpaid first and second viewings, equation (2.2). Persons with one more unpaid first viewing have 0.34 fewer second paid viewings. The coefficients for television, rental, and purchase are all negative, and the rental coefficient is large and significant (-0.26).

Column (6) also indicates that persons with more unpaid second viewings have more paid second viewings as well (0.42 and significant). Most of this reflects additional rental: persons who watch more unpaid second viewings engage in more rented second viewings.¹⁵

¹⁵ We also estimated models that disaggregate unpaid consumption into distinct downloaded and burned components. The distinct categories have statistically indistinguishable effects, so here and below we report only models with overall unpaid consumption as an explanatory variable.

TABLE III
CROSS SECTIONAL DISPLACEMENT ESTIMATES FOR FIRST AND SECOND VIEWINGS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Viewing:	First buy	First theater	First tv	First rental	First purchase	Second buy	Second theater	Second tv	Second rental	Second purchase
Unpaid First Viewings	0.062 (0.1965)	-0.1200 (0.1622)	0.0926 (0.0766)	-0.0030 (0.11150)	0.0367 (0.0677)	-0.3351 (0.1100)**	0.0120 (0.0064)	0.0222 (0.0695)	-0.2636 (0.0499)**	-0.1057 (0.0624)
Unpaid Second Viewings						0.4186 (0.1587)**	-0.0070 (0.0092)	0.0136 (0.1002)	0.5347 (0.0720)**	-0.1227 (0.0900)
Fan = 2	9.6677 (3.4942)**	4.1441 (2.8841)	1.8201 (1.3620)	4.0491 (2.0455)*	-0.3456 (1.2045)	1.4561 (1.8299)	-0.0532 (0.1062)	0.8772 (1.1556)	0.4692 (0.8300)	0.1630 (1.0380)
Fan = 3	16.9425 (3.1999)**	9.8594 (2.6412)**	1.6728 (1.2473)	5.1814 (1.8732)**	0.2289 (1.1030)	3.5706 (1.6757)*	-0.0086 (0.0973)	1.0674 (1.0582)	1.4291 (0.7601)	1.0827 (0.9506)
Fan = 4	24.2538 (3.4618)**	15.5082 (2.8574)**	3.8026 (1.3494)**	5.5389 (2.0266)**	-0.5959 (1.1933)	6.9474 (1.8137)**	-0.0038 (0.1053)	3.4337 (1.1454)**	1.2330 (0.8227)	2.2845 (1.0288)*
Fan = 5	28.1352 (4.0252)**	17.7141 (3.3224)**	3.9585 (1.5690)*	7.8630 (2.3564)**	-1.4004 (1.3875)	6.5046 (2.1080)**	-0.0696 (0.1224)	3.6289 (1.3312)**	0.5197 (0.9562)	2.4256 (1.1958)*
Movies owned	0.0288 (0.0120)*	0.0045 (0.0099)	-0.0029 (0.0047)	-0.0061 (0.0071)	0.0333 (0.0042)**	0.0320 (0.0064)**	0.0002 (0.0004)	0.0008 (0.0040)	-0.0001 (0.0029)	0.0311 (0.0036)**
Moviegoing	3.2080 (0.6959)**	3.2493 (0.5744)**	-0.5044 (0.2713)	0.6606 (0.4074)	-0.1975 (0.2399)	1.1235 (0.3655)**	0.0114 (0.0212)	0.1493 (0.2308)	0.6326 (0.1658)**	0.3302 (0.2074)
Constant	19.2457 (3.0131)**	10.1857 (2.4870)**	3.3753 (1.1745)**	2.4537 (1.7639)	3.2310 (1.0386)**	3.0394 (1.5780)	0.0525 (0.0916)	1.2220 (0.9965)	0.6058 (0.7158)	1.1591 (0.8951)
Observations	454	454	454	454	454	454	454	454	454	454
R-squared	0.28	0.25	0.04	0.04	0.13	0.22	0.01	0.06	0.18	0.21

Notes: Standard errors in parentheses. *significant at 5%; **significant at 1%. Dependent variables are measures of paid consumption. In columns (1)–(5) the dependent variable is paid first consumptions, while in the remaining columns it is paid second consumptions. 'Buy' in columns (1) and (6) is the sum of all paid modes. Fan = x is a dummy indicating the respondent's level of interest on a 5-point scale where 5 is the highest. 'Moviegoing' indicates how often the respondent goes to the movies per month.

If one takes these estimates as reflecting causality links, then the effect of each additional first unpaid viewing is the sum of 0.01 and -0.34 , or a reduction of 0.33 in paid viewings. Each second-time unpaid viewing, on the other hand, raises paid viewing by 0.42. For reasons we discussed earlier (namely, unobserved heterogeneity) one doubts that these estimates reflect causality.

More convincing estimates require ways of circumventing the unobserved heterogeneity. A natural strategy is instrumental variables. Our survey includes information on broadband and DVD-burner access, which seem like promising instruments for unpaid consumption (see Rob and Waldfogel [2006]; Zentner [2006]). In our data these technological access variables bear no significant relationships with unpaid consumption. These variables fail as instruments for three reasons. First, broadband access is virtually ubiquitous, so there is little variation. Second, even if there were more variation in broadband access, most unpaid consumption copies are obtained via copying rather than downloading, which reduces the usefulness of access speed in predicting unpaid consumption. And while access to DVD-burning hardware would seem a promising predictor, whether one has access to this technology may just as important as whether one's friends have such access. We are left without a promising IV approach.

As mentioned earlier, the sequential nature of consumption allows a different strategy. We can restrict attention to second viewings for the subset of movies with a first paid consumption, and then ask how much paid second consumption is displaced by an unpaid second consumption. On average, a third of movies first viewed at the theater are subsequently viewed by another medium (see Table II). By restricting attention to the movies first viewed, say, at the theater, we grow closer to holding constant individuals' interest in the movies in question. We can then ask how the number of paid second viewings, per movie first seen in the theater, relates to the number of unpaid second viewings, again relative to the number first seen in the theater.

Table II offers a simple implementation of this approach. The table divides survey respondents into two groups, those who engage in no unpaid consumption during the sample, and those who engage in some. Table II also reports these sequences separately for persons in these two groups who report no unpaid consumption as well as those engaging in unpaid consumption. A comparison of these groups' consumption sequences provides some suggestive evidence of displacement following a first (paid) viewing.

For persons with no unpaid consumption, 34.9 per cent of movies first viewed in a theater are viewed a second time by another paid means. For the persons who engage in some unpaid consumption, only 30.1 per cent of movies first viewed in a theater have a second, paid viewing, while 4.5 per cent have an unpaid second viewing. That is, both groups follow up a first theatrical viewing with the same tendency to view the movie through a

second medium. But the paid viewing by those who sometimes steal is lower by almost exactly the amount of their unpaid consumption. This suggests one-for-one displacement.¹⁶ A similar exercise to second consumption following first viewing of a rental copy yields similar implied displacement.¹⁷ Regression analogues of this approach suggest roughly 0.25 displacement of second paid consumptions by second unpaid consumptions, although the estimates (see the working paper version) are only significant in one-sided tests.¹⁸

III (iii). *The Longitudinal Approach*

The problem with the cross sectional estimates is that the error term, containing the individual's unobserved tendency to pay for movies, is correlated with his tendency to engage in unpaid consumption. The solution is to purge the error of this component. Longitudinal data allow this, provided that the unobservable determinant of individuals' paid consumption is constant across vintages. Because these estimates avoid this important form of unobserved heterogeneity, this is our preferred estimation approach, particularly for estimating the effect of first unpaid consumption on first paid consumption.

The movies in the survey represent the 50 most popular (highest box office) films over each of the three years prior to the survey. We can calculate the number of paid and unpaid consumption episodes among the films of each vintage. We can then use the data as a panel, including a fixed effect to purge the error term of individuals' vintage-invariant interest in buying movies. In effect, we ask whether persons who watch more unpaid 2005 movies, relative to 2003 movies, watch fewer paid 2005 movies, relative to 2003 movies.

For the regression framework, define:

B_{it}^j = number of paid j^{th} (where $j = 1$ or 2) consumption episodes by person i for vintage t movies;

S_{it}^j = number of unpaid j^{th} (where $j = 1$ or 2) consumption episodes by person i for vintage t movies;

μ_i = unobserved determinants of an individual's tendency to buy movies;

ϕ_t = vintage effect; and

ε_{it} = idiosyncratic error.

$$^{16} \frac{\partial \text{paid}}{\partial \text{unpaid}} = \left(\frac{.301-.349}{.045-0} \right) \approx -1.$$

$$^{17} \frac{\partial \text{paid}}{\partial \text{unpaid}} = \left(\frac{-0.043}{0.125-0.097} \right) \approx -1.5.$$

¹⁸ See Rob and Waldfogel working paper [2006] for more discussion of these estimates.

We then estimate:

$$(3.1) \quad B_{it}^1 = \alpha^1 + \beta^1 S_{it}^1 + \mu_i + \phi_t + \varepsilon_{it}^1$$

and

$$(3.2) \quad B_{it}^2 = \alpha^2 + \beta^{22} S_{it}^2 + \beta^{21} S_{it}^1 + \mu_i + \phi_t + \varepsilon_{it}^2.$$

Table IV reports results.¹⁹ As before, the first five columns refer to first consumption episodes. The coefficient on unpaid consumption in the first column, -0.76 (s.e. = 0.10) indicates that persons with one additional unpaid first consumption engage in 0.76 fewer paid first consumption episodes. Each of the by-medium coefficients is also negative and significant, the largest being -0.27 (theater) and -0.29 (rental).

The last five columns of the table examine the relationship between unpaid first and second consumption and paid second consumption. Here, the coefficient on unpaid second consumption is -0.19 (s.e. = 0.09), while the coefficient on unpaid first consumption is -0.24 (s.e. = 0.06).

Interpreting the coefficients causally, a first unpaid consumption reduces paid consumption by 1 unit ($-0.76 + -0.24$). Recall that most first consumption episodes are theatrical and that a third of these are followed by a second paid consumption episode. Hence the maximum displacement exceeds one, and so the estimate we report should be considered as less than one-for-one displacement.

The longitudinal displacement estimate of unpaid second consumption of -0.19 is similar to the unreported estimate of roughly -0.27 following theatrical first consumption.

While these longitudinal estimates get around the problem of fixed unobservable tastes for movies inducing a positive bias in the relationship between paid and unpaid consumption, they are of course not immune to all possible biases. It is possible that persons whose unpaid consumption is rising faster over time are simply people growing busier over time and become less able to patronize movie theaters at scheduled show times. Their increased unpaid patronage would appear to cause their reduced paid consumption. Of course, this particular confounding argument would not explain away our displacement findings for non-theatrical consumption modes.

What do these estimates mean? On average, in our data, consumption sequences include 1.3 viewings, so the maximum amount of conceivable displacement exceeds one.²⁰ An unpaid first consumption reduces paid

¹⁹ The models are estimated combining data from the spring ('early') and fall ('late') surveys. Because early respondents report on movies 2002–2004 while late respondents report on movies 2003–2005, each group gets its own vintage dummies in these models.

²⁰ To see this, imagine a first consumption in the theater followed, sometimes, by rental and/or purchase. Alternatively, one might begin by watching an unpaid copy and then engage on no

TABLE IV
LONGITUDINAL DISPLACEMENT ESTIMATES FOR FIRST AND SECOND VIEWINGS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Viewing:	First buy	First theater	First tv	First rental	First purchase	Second buy	Second theater	Second tv	Second rental	Second purchase
Unpaid First Viewings	-0.7579 (0.0958)**	-0.2652 (0.0694)**	-0.1186 (0.0507)*	-0.2947 (0.0594)**	-0.0793 (0.0373)*	-0.2443 (0.0585)**	0.0013 (0.0037)	-0.1188 (0.0415)**	-0.0351 (0.0283)	-0.0917 (0.0360)*
Unpaid Second Viewings						-0.1944 (0.0862)*	-0.0034 (0.0055)	-0.0423 (0.0611)	0.0176 (0.0417)	-0.1663 (0.0530)**
Early, 2003	0.9531 (0.2818)**	0.3567 (0.2041)	0.2336 (0.1490)	0.1297 (0.1748)	0.2331 (0.1097)*	0.5010 (0.1722)**	-0.0136 (0.0109)	0.1813 (0.1220)	0.1098 (0.0833)	0.2236 (0.1058)*
Early, 2004	-1.9830 (0.2822)**	-0.3221 (0.2044)	-0.4335 (0.1493)**	-0.9954 (0.1750)**	-0.2320 (0.1099)*	-1.2785 (0.1718)**	-0.0138 (0.0109)	-0.5000 (0.1217)**	-0.2274 (0.0831)**	-0.5373 (0.1056)**
Late, 2004	-0.2637 (0.2577)	0.0741 (0.1867)	-0.5439 (0.1363)**	0.1750 (0.1598)	0.0311 (0.1003)	-0.2853 (0.1567)	0.0147 (0.0099)	-0.3420 (0.1110)**	0.0669 (0.0758)	-0.0248 (0.0963)
Late, 2005	-6.3243 (0.2246)**	-0.9934 (0.1627)**	-1.7435 (0.1188)**	-2.3790 (0.1393)**	-1.2084 (0.0875)**	-3.2709 (0.1365)**	0.0149 (0.0086)	-1.1645 (0.0967)**	-0.8565 (0.0660)**	-1.2648 (0.0839)**
Constant	13.9864 (0.1334)**	7.5657 (0.0967)**	1.9289 (0.0706)**	3.0669 (0.0828)**	1.4249 (0.0520)**	3.7833 (0.0824)**	0.0189 (0.0052)**	1.3420 (0.0583)**	0.9495 (0.0398)**	1.4729 (0.0506)**
Observations	2164	2164	2164	2164	2164	2164	2164	2164	2164	2164
Number of individuals	541	541	541	541	541	541	541	541	541	541
R-squared	0.43	0.04	0.14	0.24	0.15	0.35	0.00	0.11	0.15	0.19

Notes: Standard errors in parentheses. *significant at 5%, **significant at 1%. Dependent variables are measures of paid consumption. In columns (1)–(5) the dependent variable is paid first consumptions, while in the remaining columns it is paid second consumptions. 'Buy', in columns (1) and (6) is the sum of all paid modes. 'Early, 2003' is a 2003 film vintage dummy from the earlier sampling which includes films from 2002–2004. 'Late, 2004' is a 2004 film vintage dummy for the later sampling which includes films from 2003–2005.

additional paid consumption. A decision to begin with an unpaid copy can therefore displace more than one paid consumption.

consumption by 1 unit. This is not 100 per cent displacement, but it is nearly 80 per cent of the maximum possible displacement. An unpaid second consumption reduces second paid consumption by only about 0.2, or 20 per cent of the maximum possible displacement, by contrast.

In our sample, individuals have an average of 1.8 unpaid first consumption episodes. According to our longitudinal estimates, these unpaid first consumption episodes displace 1.4 paid first consumption episodes ($= 1.8 \times 0.76$) per person and 0.4 second consumption episodes ($= 1.8 \times 0.24$). Our respondents have an average of 0.9 unpaid second consumption episodes, which displace 0.2 ($= 0.9 \times 0.19$) paid second consumption episodes. Overall, then, unpaid consumption displaces an average of 2.0 paid consumption episodes per person in the sample. To put this another way, unpaid consumption reduces paid consumption episodes in the sample from a counterfactual 56.2 to the observed 54.2, or by 3.5 per cent.

IV. MOVIES VERSUS MUSIC

Our results in this study, that movie downloading has a low amount, but a high rate, of displacement stands in sharp contrast with music downloading results derived from a similar sample of individuals. In that context, we found, a high volume of unpaid consumption but a low rate of displacement. Why the difference?

We suspect that the difference arises for two reasons, different costs of obtaining unpaid copies and different total costs of consumption. Music files are small enough that they can be downloaded, or copied *en masse*, very quickly. A song can be downloaded in seconds, and a CD containing hundreds of MP3 files can be copied in minutes. Movie files are, by contrast, much larger. A full-fidelity DVD contains nearly five gigabytes of information, so DVD copying takes far longer than MP3 copying. And movie downloading, even of compressed files, tends to take hours rather than seconds. All of this explains the much larger volume of unpaid music, as opposed to movie, files obtained in our sample.

As downloading speeds and compression technologies improve, it is likely that downloading will become much quicker, giving viewers access to movies they value less highly. By itself, this would tend to raise the amount of unpaid consumption, but reduce the rate of displacement. At the same time, the movie studios are moving toward releasing movies in higher-resolution formats, on discs that hold 5 to 10 times the capacity of a current DVD, which, one expects, would lower the amount of unpaid consumption.

Whatever one pays, in time or money, for a movie or music file, there is another, attention cost to actually watching it. And this cost is much higher for movies than for music. Watching movies requires a viewer's undivided attention, so even if movies could be obtained without payment, an

individual's time available to watch movies is finite. Unless individuals find a great deal of additional time for viewing movies as the effective cost of obtaining them declines, the number of movies viewed will not change. Consequently, the viewing of unpaid movies would largely displace paid movie viewing. Music is different. Because music listening consumes only divided attention, an individual can listen to a vast amount of unpaid music without necessarily using much music he would otherwise have purchased. Hence the low rate of displacement accompanying the high volume of unpaid music consumption.

V. CONCLUSION

In closing, we recap our findings and discuss to what extent they can be used to predict the future impact of piracy on the movie industry. Our overall finding is that, although the rate of displacement is large (1.0), the amount of displacement is small. Indeed, only a small fraction of consumption in our sample, 5.2%, is unpaid consumption. Paralleling this movie revenues have grown significantly (39%) over the last four years, and the bulk of this increase is attributed to DVD sales (despite the fact that DVD sales suffer the most from piracy). The rate of displacement we have estimated is sufficiently high that piracy continues to be a serious concern as the technologies allowing unpaid consumption improve and diffuse widely. Sharp sales displacement effects raise difficult questions about the marketing of movies (and other information goods) that future research will need to address.

REFERENCES

- Bakos, Yannis; Brynjolfsson, Erik and Lichtman, Douglas, 1999, 'Shared Information Goods,' *Journal of Law and Economics*, 42, pp. 117–155.
- Blackburn, David, 2004, 'On-line Piracy and Recorded Music Sales', available at http://www.economics.harvard.edu/~dblackbu/papers/blackburn_fs.pdf
- Bounie, David; Bourreau, Marc and Waelbroeck, Patrick, 2005, 'Pirates or Explorers? Analysis of Music Consumption in French Graduate Schools,' at <http://www.ecare.ulb.ac.be/ecare/ws/honorvictor/papers/waelbroeck.pdf>
- Hong, Seung-Hyun, 2005, 'The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster,' at <https://netfiles.uiuc.edu/hyunhong/www/napster.pdf>
- Hui, Kai-Leung and Png, Ivan, 2003, 'Piracy and the Legitimate Demand for Recorded Music,' *Contributions to Economic Analysis*, 2, article 11.
- Liebowitz, Stanley J., 2006, 'File Sharing: Creative Destruction or Just Plain Destruction,' *Journal of Law and Economics*, 49, pp. 1–28.
- Motion Picture Association 2005, 'U.S Movie Attendance Study, 2005' at <http://www.mpa.org/index.asp>
- Motion Picture Association/L.E.K. 2005, 'The Cost of Movie Piracy' at <http://www.mpa.org/index.asp>
- Oberholzer-Gee, Felix and Strumpf, Koleman, 'The Effect of File Sharing on Record Sales: An Empirical Analysis,' *Journal of Political Economy*, 115, pp. 1–42.

- Rob, Rafael and Waldfogel, Joel, 2006, 'Piracy on the High C's: Music Downloading, Sales Displacement and Social Welfare in a Sample of College Students,' *Journal of Law and Economics*, 49, pp. 29–62.
- Rob, Rafael and Waldfogel, Joel, 2006, 'Piracy on the Silver Screen,' NBER Working Paper 12010.
- Varian, Hal, 2000, 'Buying, Sharing, and Renting Information Goods,' *Journal of Industrial Economics*, 48, pp. 473–488.
- Zentner, Alejandro, 2006, 'Measuring the Effect of Online Piracy on Music Sales,' *Journal of Law and Economics*, 49, pp. 63–90.