Piracy or Promotion?
The Impact of the Broadband Internet Penetration on DVD Sales

Michael D. Smith and Rahul Telang

{mds, rtelang}@andrew.cmu.edu

H. John Heinz III School of Public Policy and Management
Carnegie Mellon University
Pittsburgh, PA 15213

This Version: June 2006

Available from http://ssrn.com/abstract=918240

Acknowledgements: The authors thank Nielsen VideoScan, the United States Department of Commerce, and Wal-Mart for generously providing data for use in this research. Both authors thank the National Science Foundation for generous financial support provided through CAREER award IIS-0118767 (Smith) and CAREER award CNS-0546009 (Telang).

Piracy or Promotion?
The Impact of Broadband Internet Penetration on DVD Sales

ABSTRACT

The movie industry represents one area where digital networks have had a particularly strong impact on economic activity. These digital networks provide copyright holders with new sales and promotional channels for their content, while also providing consumers with new opportunities to obtain high quality free copies of this content. Broadband access is a necessary condition for movie piracy and the movie industry has argued that the dominant impact of increased broadband Internet penetration will be increased piracy, and reduced media sales.

We analyze this hypothesis by applying fixed effects and first difference models to a new dataset quantifying changes in broadband Internet penetration and DVD sales at a local level from 2000 to 2003. Contrary to industry concerns, we find that increased broadband Internet penetration leads to a significant *increase* in DVD sales. Our results are robust across a variety of specifications. Using the most conservative results, we find that 9.3% of the $14.1 billion increase in DVD sales during our study period can be directly attributed to increased broadband Internet penetration. This corresponds to a $1.3 billion increase in DVD revenue and a $630 million increase in profits to movie studios.

Keywords: *Information goods, Internet penetration, movie promotion, DVD Sales.*

1. Introduction

> "Viacom believes that … broadband subscriptions have reached the 'tipping point' at which it can no longer afford to expose its content to piracy."
> *Viacom comments before the Federal Communications Commission in the matter of Digital Broadcast Copy Protection, December 6, 2002. (Lucey 2002, page 8)*

> "The moat that has slowed a wide-spread assault on movies in digital form is the languor with which American computer-homes have valued broadband access… But that moat will gradually be drained as broadband grows, both in its speed-power and in the deployment of broadband to homes. Once that happens…all barriers to high-speed take-downs of movies will collapse… It is the certainty of that scenario which concerns every moviemaker and distributor in the land."
> *Jack Valenti, testimony before the Senate Committee on Foreign Relations, February 12, 2002. (Valenti 2002)*

Digital computer networks create new opportunities and new challenges for sellers of information goods. These opportunities and challenges are particularly apparent in the movie industry, whose business model is driven by its ability to extract revenue from what are essentially information goods.

On one hand, digital networks can create new and lower cost channels for studios to promote, sell, and distribute their content to paying customers. Having high speed access to the Internet might allow consumers to collect and exchange more information about movies they are interested in, might allow access to products that would not have been available in brick-and-mortar channels (Brynjolfsson, Hu, and Smith 2003), and might allow producers a more targeted channel to promote movies of interest to consumers.

On the other hand, digital networks create new opportunities for unscrupulous users to illegally obtain free copies of movies through file sharing. And particularly in the case of movies, whose

file size can range from hundreds of megabytes to gigabytes, such file sharing is predicated on broadband Internet access (Pruitt 2004).

It is this link between broadband Internet access and piracy of movie file that has led Hollywood studios to argue that the dominant impact of broadband Internet access will be a significant increase in Internet piracy and a resulting drop in DVD sales. The quotes at the beginning of this section are representative of this view. Jack Valenti's statement while president of the Motion Picture Association of America is most to the point, arguing that the increase in broadband Internet penetration is the "scenario which concerns every moviemaker and distributor in the land."

A loss in DVD sales due to piracy is not an idle concern for movie studios where media sales, primarily DVDs, made up 46% ($14.9 billion) of their total revenue in 2002 (Epstein 2005, p. 20; see also PBS 2005). In 2002 DVD sales were a little over twice that of theater revenue. Moreover, the studio's margins on DVD sales are higher than their margins in many of their other lines of business. Studios only pay 20% of DVD revenues to the various artist and production unions, keeping the remaining 80% for themselves. This "sweetheart deal" was the result of savvy negotiating on the part of the studios early in the development of the DVD standard. Indeed the DVD standard emerged as an enormous boon to the movie industry. For example, Time Warner's library of movies appreciated by $7 billion between the time the DVD was introduced in 1996 to 2004 (Epstein 2005).

Moreover, studios' concerns about piracy are not without empirical and anecdotal support. A recent study (sponsored by the studios) found that in 2005 Hollywood studios lost $1.3 billion in the United States due to piracy (McBride and Fowler 2005). Of this, $447 million was attributed to Internet piracy (with the remaining amounts coming from physical copies of movies sold by

professional bootleggers ($335 million) and illegal copies made by individual consumers ($529 million).

These large numbers are driven by the widespread availability of pirated movies online. Pirated copies of DVD movie releases are typically available on Internet file-sharing networks several days to a week before the movie is released for retail sale (Byers et al. 2003).[1] Moreover, for some movies, the pirated file can receive hundreds of thousands of downloads prior to its DVD release date. To illustrate this effect, on July 6, 2006, we searched mininova.org, a poplar movie sharing site for the movie "Superman Returns," a movie that had been in theaters for a little over two weeks. On this date, mininova showed that more than 50,000 users were actively downloading (illegal) copies the movie, and that the cumulative number of downloads in the first two weeks of release are at least 10 times this amount.

However, while there is much anecdotal evidence for the potential impact of piracy, there have been no rigorous studies showing a causal relationship between increased access to broadband digital networks and reduced DVD sales. This is because analyses at the individual DVD level are problematic due to causality concerns: popular movies both sell more DVDs and are shared more in file sharing networks. Since controlling for popularity is generally quite difficult, any direct estimate capturing the effect of sharing on individual DVD sales is bound to be spurious, without the presence of appropriate instruments, which are hard to find.

Therefore in our research, instead of focusing on individual DVD sales, our unit of analysis is the aggregate sales of DVDs at a local level. We have obtained DVD sales data from Nielsen Vid-

[1] DVDs are available before release due to leaks in the production process and leaks in distribution networks. DVDs are pressed several months before releases are shipped to video rental outlets and retail stores several weeks before release.

eoscan for 99 Designated Market Areas (DMAs) for the years 2000, 2001 and 2003. We have matched this data with broadband Internet penetration data for these same regions. Because of the bandwidth required to download ripped DVDs (typically around 3-6 GB uncompressed), broadband penetration serves as a useful proxy for (unavailable) data on file-sharing levels across regions. Our broadband penetration data comes from a U.S. Government current population survey (CPS) of more than 60,000 U.S. Households conducted in 2000, 2001, and 2003. CPS one of the most reliable and detailed data source used by government policy makers and academics.

The use of panel data is critical to our study. A purely cross-sectional analysis of broadband penetration and DVD sales would again be subject to spurious correlation. With a longitudinal dataset across a large number of local areas, we can estimate a difference-in-difference model that overcomes these limitations. The intuition is that by running a regression of changes in DVD sales on changes in broadband penetration over time (after controlling for local area characteristics), we can obtain an unbiased estimate of the impact of broadband penetration on DVD sales.

Using this data we find that, contrary to concerns expressed by the movie industry, increased broadband Internet penetration leads to a large and statistically significant increase in DVD sales. This increase is robust to a variety of different specifications. Using the most conservative results, we find that 9.3% of the $14.1 billion increase in DVD sales during our study period can be directly attributed to increased broadband Internet penetration. This corresponds to a $1.3 billion increase in DVD revenue to the movie industry, and a $630 million increase in profits to Hollywood studios.

The remainder of this paper proceeds as follows. In Section 2, we review the relevant literature on piracy and its impact of media sales. In Section 3, we present a brief theoretical treatment of the possible impacts of increased broadband Internet penetration on DVD sales. Sections 4 and 5 present our empirical estimation models and data. Section 6 presents our results. We close with a discussion of the implications of our results in Section 7.

2. Literature

Our research generally pertains to the literature on the impact of piracy on markets for information goods. Most of the work in this area has focused on software or music "piracy," and particularly on P2P file sharing networks and their impact on firm profitability. A prominent trend in the analytic literature has been to show that that piracy need not be bad for firms. Prasad and Mahajan (2003) argue that piracy may be good for a new product if the firm needs to establish an initial user base and speed up diffusion. Gu and Mahajan (2004) show that because piracy removes the most price sensitive buyers from the market it can reduce price competition, benefiting sellers. Finally, Peitz and Waelbroeck (2003) show that piracy can act as a free "sample" that increases product awareness.

The empirical work on piracy has focused on estimating the effect of piracy on legal demand, with somewhat mixed results. On one hand, Hui and Png (2003) find that piracy negatively affects the legal demand for software, Liebowitz (2005) finds that file sharing has an adverse effect on CD sales, and Rob and Waldfogel (2006) find that piracy decreased media consumption by 3.5% in a sample of 500 University of Pennsylvania students. On the other hand Oberholzer and Strumpf (2004) find that music file sharing on P2P networks has no effect on CD sales and Bhat-

tacharjee et al. (2004) find that the impact of piracy depends on the popularity of the particular music CD, with more popular CDs facing higher piracy risks.

We also note that the impact of piracy on product markets is conceptually similar to the impact of used goods markets on new product sales (Ghose, Smith, and Telang (2006)), competition between traditional print copies of books and PDF copies of books (Kannan and Jain 2002), consumers' decisions to rent or purchase movies (Knox and Eliashberg 2005), and international movie release windows (Elberse and Eliashberg 2003).

However, we are not aware of any studies that have analyzed the net impact of digital networks on sales of DVDs and other digital media. Studying only the impact of piracy effectively ignores positive impacts digital networks can have on DVD sales. Digital networks can have a strong positive impact on media sales through the availability of formal and information promotional information on movies, and through the increased product variety available through online markets (Brynjolfsson, Hu, and Smith 2003). For example Anderson (2004) notes that while a typical Blockbuster video rental outlet carries around 75 documentaries on DVD, Amazon.com carries 17,061. Similarly, Anderson (2006) notes that, DVDs outside the 1,000 most popular titles at any given time sell three times more titles through the Internet than they do through brick-and-mortar stores. We discuss these competing impacts of digital networks on DVD sales in more detail below.

3. Theory of the Impact of Broadband on DVD Sales

Increased penetration of broadband digital networks can have either a positive or negative impact on sales of information goods such as DVDs. In terms of negative impact, broadband digital networks may to spur the growth in movie file sharing. Similar arguments were made by Lie-

bowitz (2005) and Boorstein (2004) when they used Internet penetration as a proxy for music file sharing. While movie file sharing has not received as much press as music file sharing, it is still a significant phenomenon, as highlighted above. The sheer numbers of movies downloaded and shared is in millions. The movie industry has recognized this and a recent industry study finds that losses due to illegal online movie file sharing are in the range of about $550 million (McBride and Fowler 2006). More importantly, movie file sharing cannot occur without broadband connectivity because of the sheer size of the movie files.

However, as noted above, the impact of file sharing itself has been a controversial topic with differing conclusions and opinions. File sharing can act a substitute where a copy freely available online substitutes for a legally available content and thereby hurts and interest of original copyright holders. However, there is an alternative argument put forward, namely sampling. Files available online can make users aware of the availability of certain movies and artists and eventually lead to purchases by users. Thus the argument is that file sharing networks can actually reduce the search cost for users and can make them aware of more obscure and less advertised movies leading to an increase in DVD purchase. There is significant empirical work for music file sharing and how it has impacted the recording industry. However, there is no consensus on the magnitude of file sharing impact on CD sales (See Oberholzer and Strumpf (2004), Liebowitz (2005) and Bhattacharjee et al (2005) for details).

Independent of file sharing, access to broadband Internet networks may stimulate DVD sales. Recent work by Brynjolfsson et al. (2005) and Brynjolfsson et al. (2003) show that increased product variety and decreased search costs for product information in Internet markets has made is easier for users to find niche products that offer a good fit for their individual tastes. The availability of increased product variety and increased product information could be particularly im-

portant for DVDs where typical brick-and-mortar retailers carry fewer than 1,000 of the more than 40,000 DVDs in print. Given that advertising and promotional costs are very high in these markets, broadband and peer-to-peer networks may help lower promotional costs for niche titles. This conjecture is also consistent with Bhattacharjee et al.'s (2005) finding that peer-to-peer networks can actually help sales of less popular music titles.

In summary, increased penetration of broadband digital networks might hurt DVD sales through cannibalization due to piracy, and increased broadband penetration might help DVD sales through access to increased product variety, lower search costs for product information, the ability to sample product content, and the ability to better promote and market DVDs.

The majority of the extant literature has focused only on the impact of piracy. In this study we focus on the net impact of broadband digital networks on DVD sales. However, we augment this net measure with a variable that may provide a reliable proxy for piracy activity: whether users use broadband connections to watch TV, movies, or listen to radio. The users who indulge in these activities are more likely to be tech-savvy and be more aware of file sharing networks. Thus, while the focus of the paper is identifying the net impact of broadband digital networks, this measure will provide additional insights into how DVD sale are affected when users can use broadband for watching TV, movies and listening to radio.

4. Model

Our econometric model to estimate impact of broadband Internet penetration on DVD sales correlates local changes in Internet broadband penetration with local changes in DVD sales, following a methodology similar to Boorstein (2004) and Liebowitz (2005). However, our data sources

are somewhat different with additional measures, and include better controls for demographic characteristics.

We use an aggregate approach, because measuring the impact of file-sharing on DVD sales at an individual movie level is notoriously tricky due to serious simultaneity problem: DVDs that are sold more are also prone to more piracy. This makes the estimation very difficult unless one can find good instruments, a non-trivial problem. Analyses at the individual level can also ignore trends at the industry level: while an individual album may suffer from piracy, industry as a whole may not suffer as much (or at all).

Our approach with aggregate data avoids these problems by estimating the impact of file sharing on aggregate sales rather than individual albums. Moreover, by using panel data, we can avoid the simultaneity problem as well, as explained below.

Specifically, we first estimate the following model using ordinary least squares for both individual years in our data and our data pooled across years:

$$Sale_{it} = \delta 0 + \delta 1 * bband_penetration_{it} + \delta 2 * X_{it} + \varepsilon_{it} \quad (1)$$

where $bband_penetration_{it}$ and $Sale_{it}$ represent the per capita penetration of broadband Internet access and the per capita level of DVD sales respectively in DMA i and year t; and X is a vector of other exogenous factors that may affect DVD sales. For our individual year regressions, these exogenous factors include demographic characteristic and the number of Wal-Mart stores in the DMA. In our pooled regressions, we also add yearly indicator variables.

Next, we estimate a modified version of (1) using both DMA- and year-level fixed effects. Fixed effects are important in our setting because there may be many unobservable characteristics at a DMA-level that could be correlated with both broadband penetration and DVD sales. For exam-

ple, a more urban DMA or a younger DMA will probably see more broadband penetration as well as higher DVD sales. If this is the case, equation (1) should properly be rewritten as:

$$Sale_{it} = \delta 0 + \delta 1 * bband_penetration_{it} + \delta 2 * X_{it} + u_i + \xi_{it} \quad (2)$$

where u_i represents the unobserved characteristics that may be potentially correlated with broadband penetration, making $\delta 1$ potentially biased.

We overcome this problem by collecting a longitudinal data set allowing us to use fixed effects — essentially a separate identifier variable for each DMA — to avoid the problem of these unobserved effects. This controls for DMA specific unobserved effects and allows u_i to correlate with the regressors.

Similarly, there may be a time dependent trend in the data that is correlated with the explanatory variables. To control for these effects, we also include year-level fixed effects in the regression to capture changes in DVD sales over time that are independent of broadband penetration.

Finally, one of the shortcomings of the panel data is that error terms can be autoregressive. As broadband penetration and DVD sales are generally increasing over time, we also explicitly control for autocorrelation in our estimation. Thus in (2) we allow $\xi_{it} = \rho\xi_{it-1} + z_{it}$, and control for potential autoregressive error structures following Baltagi and Wu (1999).

Besides broadband penetration, in our data, we also have direct measure of individuals' propensity to use broadband for watching TV, movies or listening to radio. One question in the CPS survey (discussed in detail in the data section below) explicitly asks the users whether they used Internet for watching TV, movies or listening to radio. In our sample, we are able to drop responses related to dialup connection, and thus we can interpret this question as whether users used broadband connection for watching TV, movies or listening to radio. Presumably, these us-

ers would be more sophisticated with digital technologies and more likely to download movies and/or make copies from DVDs or share them among their friends. Thus, we expect that a larger share of such users in a DMA may have an adverse impact on DVD sale via piracy. We re-estimate (3) with this data to see if the use of broadband digital networks for watching movies and TV and listing to radio has an adverse impact on DVD sales.

5. Data

Our data consist of DVD sales, broadband penetration, broadband use for TV, movies and radio, demographic characteristics, and market characteristics for the 99 largest media markets in the United States from 2000 to 2003. This is an important time period because 2000 was the beginning of the large-scale rollout of Internet broadband access in the United States. Between 2000 and 2003, broadband penetration increased from about 5% to 25%. Thus, we capture large variation in broadband penetration across years, and disentangle the causality between these two variables through the use of local-level observation.

We obtain our DVD sales data from Nielsen VideoScan. Nielsen captures the aggregate sales of DVDs through both Internet and brick-and-mortar channels within DMAs.[2] A DMA is typically defined as a large media market and its surrounding areas. For example, the Philadelphia DMA includes more than 450 zip codes and includes areas from Philadelphia to some parts of Reading, PA and some zip codes in Delaware. Our data cover the 99 largest DMAs in the United States that cover about 83% of the total U.S. population. The data provided by Nielsen quantifies unit DVD sales in these DMAs for the years 2000, 2001, and 2003. We convert this to per capita DVD sales using census data for each DMA.

[2] Note that Internet sales are allocated at customer's location, not the retailer's location.

Our broadband penetration data was collected from the Current Population Survey (CPS), a monthly household survey conducted by the Bureau of the Census for the Bureau of Labor Statistics. The Current Population Survey provides a comprehensive body of information on the employment and unemployment experience of the Nation's population, classified by age, sex, race, and a variety of other characteristics. The CPS covers a representative sample of approximately 130,000 individuals nation-wide and asks a multitude of questions regarding employment, income level, etc. The CPS is one of the most comprehensive and reliable data source and these data are used by many government and research organizations to inform policy decisions (www.bls.gov/cps).

The CPS also periodically issues “Internet use” supplements, and some of the questions on this survey are whether a user has a broadband connection (typically defined as a speed of 200 kilobits per second one way), what type of broadband connection is used (e.g., DSL, Cable, Satellite, Fiber Optic), and whether these connections are used to watch TV or movies, or listen to the radio. With these questions, the CPS data allow researchers to know the proportion of users within a geographical area who have an Internet broadband connection, and to have a limited idea of how these connections are being used.

The CPS aggregates its data by Metropolitan Statistical Area (MSA), which are smaller in size than DMAs. Broadband data is available for the 241 largest MSAs in the country, covering 82% of the U.S. population. The CPS provides broadband information as un-weighted raw number as well as weights for the number by households and the number of persons. We use person-weighted data (PWSSWGT) as it is the standard weight to use for demographic calculations.[3]

[3] Household weights give similar results, but person-weighted data are better suited to our disaggregate analysis.

Further details on the CPS data and weighing scheme can be found (http://www.bls.census.gov/cps).

Our demographic data was collected from Sourcebook America for the year 2000 and 2003 for each DMA. For year 2001 we extrapolated data from 2000 and 2003 and used available data from the MSA level. Our demographic data include control variables for the population of the MSA; and the median age, median income level, proportion of whites, proportion of males, and proportion of the population between the ages of 15 and 24 for each DMA. These variables are included as controls, and also have a theoretical basis, particularly in the case of 15 to 24 year old males. According to a recent study conducted by LEK Consulting for the Motion Picture Association, 16 to 24 year olds and males made up 71% and 58% respectively of active Internet downloaders in 2004 (LEK 2005).

One remaining challenge is to match our DMA-based DVD sales data to the MSA-based broadband usage and demographic data. In previous work (e.g. Boorstein 2004), this mapping was based on the assumption that if the largest zip code in an MSA was contained in a DMA, then the whole MSA was mapped into that DMA. This approach, while reasonable, can create problems as DMAs typically encompass a larger area than MSAs do.

In an attempt to map MSAs into DMAs more precisely, we create a mapping based on the individual zip codes for each MSA and DMA, along with the population of each zip code, obtained from the U.S. census[4]. We then match MSAs into DMAs using zip codes. Not every zip code of a DMA is matched into MSAs. In short, a DMA is not completely mapped into MSAs. However, our mapping is still quite good. On average, we can map about 88% of a DMA into one or more

[4] http://mcdc2.missouri.edu/websas/geocorr2k.html

MSAs. For larger and more urban DMAs almost 100% mapping is achieved (for example New York or Philadelphia). However, there are a few DMAs where coverage is less than 50% (for example, Johnstown- Altoona DMA). Thus, while imperfect, we believe this mapping is accurate for our analysis and represents an advance of techniques in the extant literature. In the Appendix, we perform sensitivity analysis to verify how robust our results are to DMA coverage.

Using this mapping to measure broadband penetration level at a DMA level, we weight the MSA level broadband penetration by the proportion of MSAs that belongs to a DMA. For example, if 30% of MSA-A's zip codes (mapping fraction of MSA) and 70% of MSA-B's zip codes map into DMA-1, we first calculate per capita broadband penetration at the MSA level, weight them by the mapping fraction, and sum them up to calculate per capita broadband penetration at DMA level.

With regard to market data, one shortcoming on the Nielsen DVD sales data is that it does not include sales figure from Wal-Mart. In our sample time period, Wal-Mart accounted for approximately 29% of all DVD sales in the United States (Alexander and Associates, cited in Mortimer 2006), a share that seems to be increasing over time. For example, Wal-Mart notes that in 2006 it was "Hollywood's biggest customer, accounting for 40% of [Hollywood's] DVD sales" (Wal-Mart, 2006).[5] However, because of the absence of Wal-Mart sales in our data it is important to control for changes over time in the number of Wal-Mart outlets in the DMAs we analyze. To do this, we obtained a list of all Wal-Mart locations in the United States, their address (including zip code), the type of store (Superstore, Wal-Mart, local store), and their open-

[5] Similarly, a 2005 story in Variety placed Wal-Mart's sales at 37% of total Hollywood output (Learmonth 2005).

ing date.[6] We use this data to construct a variable for the number of new Wal-Mart retail stores in each DMA that opened since the baseline 2000 year.

Summary statistics for our data are presented in Table 1 below.

Table 1: summary Statistics

	Year 2000		*Year 2001*		*Year 2003*	
	Mean	*Median*	*Mean*	*Median*	*Mean*	*Median*
Per capita DVD sales (pc_sale)	0.19 (0.09)	0.185	0.38 (0.17)	0.37	0.698 (0.243)	0.682
Per capita broadband usage (pc_bband)	0.046 (0.03)	0.044	0.114 (0.05)	0.107	0.25 (0.07)	0.255
Per capita use of broadband for TV, movies and radio (pc_movies)	NA		0.027 (0.017)	0.024	0.063 (0.023)	0.062
Population (in 000)	2,424 (2,754)	1,580	2,454 (2,785)	1,595	2,516 (2,846)	1,643
Median income (in 000)	40.92 (6.0)	41.05	42.26 (6.3)	42.15	44.93 (6.95)	44.6
Median age (in years)	35.8 (2.4)	35.7	36.06 (6.05)	36.0	36.56 (2.6)	36.45
Proportion of Males	49% (0.68)	48.9%	49.03% (0.67)	48.93%	49.08% (0.64)	49%
Proportion of Whites	79.85% (11.4)	82.7%	79.55% (11.7)	82.45%	78.96% (11.7)	82%
Age between 15 and 24	14.07% (0.01)	14.1%	14.14% (0.01)	14.1%	14.28% (0.015)	14.2%
No. of new Wal-Marts per DMA since 2000			1.01 (0.76)	0.0	1.93 (2.93)	1.0
N	98	98	98	98	98	98
DMA_Coverage_factor	0.76 (0.21)	0.83				

[6] We eliminate "local stores" from this is data as discussions with Wal-Mart public relations staff suggested that these stores typically sell very few DVD titles. Our results are not sensitive to this assumption.

6. Results

6.1 Impact of Broadband penetration on DVD Sale

As mentioned above, our analysis is conducted at the DMA level and our dependent variable is per capita DVD sales. Our key explanatory variable is per capita broadband penetration. We also use various control variables for demographic characteristics and the number of new Wal-Mart stores opened in 2001 and 2003.

We first run an OLS regression of per capita sales for pooled data (for all three years, as well for each year separately). We report our results in Table 2.

Table 2: Pooled Regression

Indep. Vars.	Pooled data	Year 2000	Year 2001	Year 2003
Per Capita Broadband Penetration	0.422** (0.12)	0.252* (0.15)	0.27 (0.187)	0.24 (0.22)
Log(Population)	0.046** (0.14)	0.018 (0.011)	0.06** (0.02)	0.11** (0.03)
Median Income (in 000)	0.016** (0.001)	0.009** (0.001)	0.017** (0.002)	0.022** (0.003)
% Whites	-0.002** (0.0006)	-0.001** (0.0005)	-0.002** (0.001)	-0.002* (0.001)
% Males	0.04** (0.01)	0.028** (0.008)	0.048** (0.015)	0.054** (0.025)
Median Age	0.009* (0.005)	0.002 (0.005)	0.006 (0.009)	0.015 (0.012)
% 15-24 Year Olds	0.024** (0.008)	0.44** (0.77)	0.02 (0.013)	0.042** (0.017)
Wal-Mart	-0.001 (0.003)		-0.013** (0.006)	-0.011* (0.007)
Year dummy1	-0.33**(0.03)			5.51** (1.631)
Year dummy2	-0.20**(0.02)			
Constant	-3.33**(0.75)	-1.91** (0.625)	-3.9** (1.134)	5.51** (1.631)
No. of Obs.	294	98	98	98
R^2	0.86	0.73	0.76	0.75

The dependent variable is per capita DVD sales. Standard errors are listed in parenthesis; ** and * denote significance at 0.05 and 0.10, respectively.

In the pooled regression, broadband penetration is large, positive, and statistically significant (at 1% level) for the pooled data. An estimate of 0.42 on the per capita broadband for the pooled

data suggests that a unit increase in per capita broadband Internet penetration increases the per capita DVD sales by about 0.42 units. The other control parameters generally have the expected signs. Larger DMAs and DMAs with higher income, lower proportions of 15-24 year olds, and lower proportion of males have higher DVD consumption. Also as expected, an increase in Wal-Mart retail outlets leads to a decrease in DVD sales in our data (which do not include Wal-Mart sales figures) but is not statistically significant. The high R^2 in our results suggests that our exogenous variables explain most of the variation in per capita DVD sale.

In the non-pooled, individual year regressions, the coefficient on per capita broadband penetration is also large and positive, however, here it is not statistically significant (it is only significant at 10% level in year 2000 regression). This variable and the other control variables are, however, consistent across specifications and consistent with the pooled results. The insignificance of the per capita broadband penetration variable is likely because, for our yearly data, our already small sample is divided by three, leaving less precise estimates.

Thus, our OLS results seem to suggest that increases in broadband Internet penetration leads to increased DVD sales. However, as noted above, OLS results are suspect because there may be unobserved DMA specific characteristics which may be correlated with broadband penetration, leading to higher DVD sales, and biasing our estimate of broadband penetration.

To overcome this, we estimate a fixed effect specification as in (2). The fixed effects estimates control for any unobserved DMA-level (and in models 2 and 3 time-level) effects even if they are correlated with broadband penetration. When controlling for autocorrelation we relax the assumption that error terms are not auto-correlated.

The results are presented in Table 3. In Column 1 of Table 3 we present the results of our model using only DMA-level fixed effects. In Column 2 we add controls for year-level effects. And in Column 3, we add controls for autoregressive errors as outlined above.

Table 3: Fixed Effect Regressions

Indep. Vars.	Fixed Effect (1)	Fixed Effect with Time Dummies (2)	Fixed Effect with Time Dummies and Autoregressive Errors (3)
Per Capita Broadband Penetration	0.63** (0.10)	0.39** (0.11)	0.23* (0.126)
Log(Population)	0.11 (0.43)	0.07 (0.39)	-0.65 (0.62)
Median Income	0.048** (0.006)	0.035** (0.006)	0.020** (0.009)
% Whites	-0.121** (0.023)	-0.102** (0.02)	-0.075* (0.035)
% Males	-0.145** (0.052)	-0.168** (0.05)	-0.214** (0.076)
Median Age	0.059** (0.017)	0.013 (0.02)	0.03 (0.028)
% 15-24 Year Olds	0.013** (2.675)	0.013** (2.56)	0.11** (3.78)
Wal-Mart Supercenters	-0.005 (0.005)	-0.005 (0.005)	-0.008 (0.006)
Time Dummies	No	Yes	Yes
Autoregressive Errors (AR1)	No	No	0.338
Constant	9.4 (8.4)	11.88** (7.77)	23.08** (8.0)
No. of Obs.	294	294	196
R^2 (Within)	0.94	0.95	0.94

The dependent variable is per capita DVD sales. Standard errors are listed in parenthesis; ** and * denote significance at 0.5 and 0.10, respectively. All models use DVD-level fixed effects.

The effect of broadband penetration is still large, positive and statistically significant in the fixed effects regression (column 1). Adding additional controls for time effects and autocorrelation reduces the magnitude of this coefficient, but it remains large, positive, and statistically significant in both the time effects (column 2, p=0.01) and autoregressive (column 3, p=0.10) results. Note that controlling for autoregressive errors reduces the number of observations, which impacts the standard errors and resulting significance in column 3.

In addition, the signs of the control variables are consistent across specifications.[7] They are also consistent with the pooled regression results. In addition, the time dummies in column 2 (not shown) are highly significant indicating that controlling for time effects is important. The autoregressive coefficient ($\rho = 0.31$) in column 3 is not particularly large. However, including it does have a significant impact on the magnitude of impact of penetration. Characteristic of fixed effects regressions, the fit of the model is excellent across specifications.

Table 4: First Difference Estimation

Indep. Vars. (with differencing)	Estimates
Per Capita Broadband Penetration	0.254** (0.10)
Log(Population)	0.027 (0.44)
Median Income	0.034** (0.007)
% Whites	-0.095** (0.025)
% Males	-0.17** (0.05)
Median Age	0.017 (0.02)
% 15-24 Year Olds	0.12** (0.03)
Wal-Mart Supercenters	-0.006 (0.004)
Time dummy	-0.024 (0.016)
Constant	0.12** (0.025)
No. of Obs.	196
R^2	0.62

An alternate way to estimate this model is by the first difference method. In first differences, instead of running a regression of Y_{it} on X_{it}, we run the regression of Y_{it}-$Y_{i(t-1)}$ on X_{it}-$X_{i(t-1)}$. When T =2, fixed effects and first difference produce the same results. When T > 2, the models are different. In particular, first differencing allows error term to be auto-correlated. In fact, if the auto-

[7] Note that the coefficient on Log(Population) is statistically insignificant in each specification.

correlation is high, then first differencing will produce more efficient estimates. We re-estimate our model using first differencing method as well and present the results in Table 4.

As before, the estimate on per capita broadband is large and statistically significant. It is also quite similar in magnitude to the coefficient in the autoregressive results of Table 3, Column 3. The other control variables have the same signs as the fixed effects results. Thus, it seems clear that, even under various specifications, per capita broadband is positive and statistically significant. And, in our sample, this result is quite robust across a variety of different specifications.

What is the economic significance of a positive and significant estimate on per capita broadband? If we take the most conservation estimate of 0.23 (Table 3, Column 3), this suggests that a unit increase in per capita broadband penetration has caused an increase of about 0.23 units in per capita DVD consumption. From year 2000 to 2003, the average per capita penetration of Internet broadband access increased from 0.046 to 0.25 or about 0.204 units (Table 1). Since β = 0.23, this suggests that this increase in broadband penetration accounted for a per capita increase in DVD sales of 0.23*0.204 = 0.047 units. From Table 1, note that during this time period per capita DVD sales increased from 0.19 units to about 0.697 units. Thus, our results indicate that 9.3%[8] of the increase in DVD sales from 2000 to 2003 is directly attributable to increased Internet broadband penetration. In dollar terms, given that DVD sales increased by $14.1 billion during our study period (MPA 2004), this corresponds to a $1.3 billion increase in DVD sales that is directly attributable to increased access to broadband digital networks. In terms of profit, previous studies have found that the wholesale price of DVDs is approximately 60% of the retail price (Mortimer 2006), and conversations with industry executives suggest that 80% of studio revenues translate into direct profit. Thus, the $1.3 billion in increased revenue translates into $630

[8] I.e., 0.047 / (0.697 – 0.19)

million of increased profit to studios directly attributable to increased broadband penetration from 2000 to 2003.

Thus, we find that the net effect of broadband is significant both statistically and economically. However, it is still possible that broadband promotes piracy, which in turn adversely impacts DVD sale. We now focus on how consumers use their digital networks to gain additional insights in this regard.

6.2 Impact of Using Broadband for watching TV and movies and listening to Radio

One of questions in the CPS survey is whether users used a broadband connection to watch TV, movies, or listen to radio. It seems likely that users who answer yes to this question are more technology savvy, more likely to be aware of movie availability on digital networks, and possibly more likely to share movies amongst friends and/or download movies.[9] Thus, this may be a useful proxy of piracy through file sharing. We now estimate whether the propensity of users to use their broadband connection to watch TV, movies, and listen to radio has an impact on DVD sales. The data was available only for year 2001 and 2003. So we present the results with first differences only (recall that first differencing and fixed effects are identical because T = 2). Controlling for autocorrelation is now not possible because we only have 2 samples per DMA.

The best way to estimate the effect of per capita broadband use for TV, movies and radio (pc_movies) is to estimate (2) by including pc_movies in it. However, pc_movies and pc_bband are highly correlated, as expected. Dropping pc_bband and running a regression with pc_movies alone is incorrect because pc_bband is a missing variable that affects both per capita DVD sales and pc_movies. In short, to estimate the marginal effect of per capita broadband use for TV,

[9] Note that there were no "legitimate" download services for Hollywood movies during the 2001-2003 time period.

movies and radio (pc_movies) we need to account for the fact that per capita broadband penetration is also increasing. Therefore, we normalize pc_movies by dividing it with pc_bband, so that it can be interpreted as per capita broadband use for TV, movies and radio conditional on per capita broadband penetration. The results with this normalized value are presented below.

Table 5:Fixed Effect Estimation With "Use of Broadband for Watching TV, Movies and listening to Radio"

Indep. Vars.	No bband controls (1)	Full Controls (2)
Normalized Per Capita broadband use for TV, Movies, and Music	0.09 (0.07)	0.043 (0.07)
Per capita broadband		0.37** (0.15)
Log(Population)	-0.25 (0.53)	-0.55 (0.56)
Median Income	0.027** (0.009)	0.025** (0.008
% Whites	-0.097** (0.033)	-0.088** (0.032)
% Males	-0.16** (0.068)	-0.21* (0.07)
Median Age	0.027 (0.025)	0.025 (0.026)
% 15-24 Year Olds	0.11** (0.035)	0.12** (0.035)
Wal-Mart Supercenters	-0.007 (0.006)	-0.006 (0.007)
Time Dummy Vars.	Yes	Yes
Constant	16.2 (11.05)	21.8* (11).
No. of Obs.	196	196
R^2	0.95	0.95

The dependent variable is change in DVD sales per capita. Standard errors are listed in parenthesis; ** and * denote significance at 0.05 and 0.10, respectively. All models use DVD-level fixed effects.

We run two specifications. In one, we include only the normalized value of pc_movies, and in the second we also include both pc_movies and per capita broadband penetration. Note that in specification (2), per capita broadband penetration is positive and significant, consistent with our prior results. However, in both specifications, per capita broadband use to watch TV and movies and listen to radio is insignificant. Recall that we argued that the answer to this question may

serve as a useful proxy for Internet piracy. If this is so, our results show that engaging in downloads through broadband digital networks had no statistical impact on DVD sales (indeed, the coefficient, while insignificant, is positive in sign). This finding is consistent with Oberholzer and Strumpf (2004) who find no impact of file sharing on album sales. However, this result should be interpreted with caution as it is not a perfect proxy for Internet piracy. It is possible that some users were answering this question from the perspective of legitimate online services, such as listening to streaming radio broadcasts, downloading movie trailers, or watching other video services available for download. Thus, while we believe this finding is suggestive, it is not conclusive with respect to the impact of piracy on DVD sales.

In summary, we find that increased broadband penetration is causally associated with a statistically significant increase in DVD sales. This increase conservatively accounts for 9.2% of the $14.1 billion increase in DVD sales over this time period. We also use the propensity to use broadband for watching TV, movies or listening to radio to partially untangle the aggregate effects of broadband penetration vis-à-vis piracy. We find that the impact of using broadband connections to watch TV and movies and listen to radio has a directionally positive, although statistically insignificant, impact on DVD sales. Thus, while our findings do not rule out the possibility that piracy hurts DVD sales, together they suggest that increased broadband penetration provides significant additional sales benefits to movie studios that overwhelm any negative impact of piracy through file-sharing.

7. Discussion

The movie industry has long argued that increased Internet broadband penetration would lead directly to increased piracy and decreased media sales. However, due to data limitations and

econometric challenges, there have been no rigorous studies analyzing this view. We overcome these challenges by compiling a dataset that quantifies broadband Internet penetration and DVD sales at a regional level. We analyze this data with fixed effects and first difference statistical model, which allow us to assign causality to our results.

In contrast to studio concerns, our results show a large, statistically significant, positive impact of increased broadband penetration on DVD sales. Using our most conservative estimates, we find that increased broadband penetration accounts for 9.3% of the $14.1 billion increase in DVD sales over our study period — representing $1.3 billion in increase revenue and $630 million in increased profit to the studios. We also find that an increase in the use of broadband digital networks to watch TV and movies and listen to the radio, activities which might be related to piracy given the relative lack of legitimate media download services in our study timeframe, is not associated with a statistically significant change in DVD sales. Together these results shed significant light on the interplay between broadband digital networks, and sales and piracy of digital media by presenting strong evidence that the dominant impact of increased access to broadband digital networks is increased media sales.

As discussed in the theory section, we believe there are several possible explanations for this finding. First, it is possible that having access to broadband digital networks lowers the search costs for users to find information about films, leading to an increase in DVD purchases. Many of the most popular sites on the Internet are primarily associated with providing consumers information about movies (e.g., the Internet Movie Database, yahoo movies, MSN Entertainment, Apple movie trailers, Rotten Tomatoes, Moviefone, Hollywood.com, Fandango.com). It is reasonable to suggest that allowing consumers to easily access such movie information through broadband connections could lead to increased DVD sales.

A related candidate explanation for the broadband digital networks leading to increased DVD sales is that digital networks allow movie studios great flexibility to target advertisements to their consumers. These dual explanations of increased access to advertising and decreased search costs for product information resonated with a representative of a major Hollywood studio who, being shown our results, commented:

> "It seems like pretty straightforward marketing/advertising to me. Fundamentally, people buy DVDs when they really like a movie, and want to watch the movie (or the special features) again and again, either now or at some point in the future…So marketing/advertising, either via [broadband] or TV, doesn't over-expose these people to the movie, it just reminds them how much they like it and want to own it..."

In conjunction with increased access to product and advertising information, broadband digital networks also provide consumers with access to an increased variety of DVDs than what could be found in brick-and-mortar stores. Typical brick and mortar video stores sell fewer than 1,000 titles, and Wal-Mart superstores typically sell fewer than 500 unique titles. In contrast, Amazon.com sells nearly all of the more than 30,000 DVD titles in print — and consumers seem to have strong demand for these niche titles. Anderson (2006) reports that DVD titles ranked below 1,000 sell three times as many copies online as they do in brick-and-mortar markets.

It is also possible that having a broadband connection may spur the purchase of related equipment, such as laptop or desktop computers, which allow users to watch DVDs in new settings, thus increasing DVD sales. However, whatever the specific explanation, it is clear from our results that the dominant impact of increased broadband penetration is increased sales of DVDs, not decreased sales as previously feared by the movie industry.

Finally, we note that our results are conservative in two important respects. First, increased Internet usage should have the effect of "crowding out" other leisure activities, such as watching DVDs, which might lead to a decrease in DVD purchases *ceteris paribus*. Additionally, the fact that the proportion of DVDs sold at Wal-Mart (which is not represented in our sales data) is increasing over time, should have the effect of reducing any observed increases in DVD sales in our data *ceteris paribus*. In each case, these effects will bias our results away from finding that increased broadband penetration leads to an increase in DVD sales, making our reported results conservative.

Our results are not without limitations. Even though we control for DMA level fixed effects, finding an instrument for broadband penetration would be ideal. That would take away any potential bias of time varying unobserved which may be correlated with changes in broadband penetration. Our proxy for piracy effects is also not without limitations. In future work, a more direct measure for piracy should be investigated.

Bibliography

Anderson, C. 2004. The Long Tail. Wired Magazine. October.

Anderson, C. 2006. The Long Tail: Why the Future of Business Is Selling Less of More. Hyperion Press, New York, NY.

Baltagi B and P Wu. 1999. A Unequally Spaced Panel Data Regressions with AR(1) Disturbances. Econometric Theory, Vol. 15, 814-823.

Bhattacharjee S., R. Gopal, J. Marsden, R. Telang. 2005. The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts. Proceedings of the 2005 CIST INFORMS Conference.

Boorstein 2004. Music Sales in the Age of File Sharing. Senior Thesis, Princeton University, April.

Brynjolfsson, E., Y. Hu, M. Smith. 2003. Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety. *Management Science* **49**(11) 1580-1596.

Brynjolfsson, E., Y. Hu, D. Simester. 2005. Goodbye Pareto Principle, Hello Long Tail. Working Paper, MIT Center for eBusiness, Cambridge, Massachusetts.

Byers, S., L. Cranor, D. Korman, P. McDaniel, E. Cronin. 2003. Analysis of Security Vulnerabilities in the Movie Production and Distribution Process. Proceedings of the 3rd ACM Workshop on Digital Rights Management, Washington D.C., pp. 1-12.

Epstein, E.J. 2005. *The Big Picture: The New Logic of Money and Power in Hollywood*. Random House, New York, NY.

Ghose, A., M.D. Smith, R. Telang. 2006. Internet Exchanges for Used Books: An Empirical Analysis of Product Cannibalization and Welfare Impact. *Information Systems Research*, **17**(1) 3-19.

Gopal, R., L. Sanders. 1998. International Software Piracy: Analysis of Key Issues and Impacts. *Information Systems Research*, **9**(4) 380-397.

Gu, B., V. Mahajan. 2004. The Benefits of Piracy: A Competitive Perspective. Sixteenth Workshop on Information Systems and Economics.

Hui, K., Png., I., 2003. Piracy and the Legitimate Demand for Recorded Music. Contributions to Economic Analysis and Policy **2**(1).

Learmonth, M. 2005. Store wars! Variety, November 13. Available from http://www.variety.com/article/VR1117932851. Last accessed July 11, 2006.

LEK. 2005. The Cost of Movie Piracy. Consulting Study Prepared for the Motion Picture Association of America. Executive Briefing available from http://www.mpaa.org/researchStatistics.asp.

Liebowitz, S. L. 2004. The Elusive Symbiosis: The Impact Of Radio on the Record Industry, *Review of Economic Research on Copyright Issues*, **1**(1) 93-118

Liebowitz 2005. Testing File-Sharing's Impact by Examining Record Sales in Cities, Working paper, University of Texas, Dallas.

Lucey, A. 2002. Comments of Viacom. Before the Federal Communications Commission, In the Matter of Digital Broadcast Copy Protection. MB Docket No. 02-230. December 6.

McBride, S., G.A. Fowler. 2006. Studios See Big Rise in Estimates of Losses To Movie Piracy. *Wall Street Journal*, May 3, page B1.

Mortimer, J. 2006. Price discrimination and technological innovation: Evidence from the introduction of DVDs. Working Paper, Harvard University, Cambridge, MA.

MPA. 2004. Piracy Fact Sheets: U.S. Overview. Motion Picture Association of America. Available from http://www.mpaa.org/USPiracyFactSheet.pdf. Last Accessed July 11, 2006.

Oberholzer F., K. Strumpf. 2004. The Effect of File Sharing on Record Sales: An Empirical Analysis. Working paper, Harvard University, Cambridge, MA.

PBS. 2005. Now playing...and playing...and playing. PBS Frontline. Available from http://www.pbs.org/wgbh/pages/frontline/shows/hollywood/business/windows.html. Last accessed January 5, 2006.

Peitz, M., and Waelbroeck, P., 2003. Piracy of Digital Products: A Critical Review of the Economics Literature. CESifo Working Paper Series No. 1071.

Prasad, Ashutosh and Vijay Mahajan 2003. How many pirates should a software firm tolerate? An analysis of piracy protection on the diffusion of software. *International Journal of Research in Marketing*, 20(4), 337-353

Pruitt, S. 2004. MPAA: Movie Piracy Is the New Plague: More and more Web users are illegally downloading movies, study reveals. IDG News Service. Available from http://www.pcworld.com/news/article/0,aid,116842,00.asp. Last Accessed July 11, 2006.

Valenti, J. 2002. A Clear Present and Future Danger: The potential undoing of America's greatest export trade prize. Testimony before the Senate Committee on Foreign Relations. Washington, D.C., February 12.

Wal-Mart. 2006. Wal-Mart Facts: News Desk. Available from http://www.walmartfacts.com/newsdesk/article.aspx?id=1530. Last Accessed March 30, 2006.

Appendix A: Sensitivity of Results to DMA-MSA Matching

As noted in the body of the text, one of the issues with our data is DMA coverage factor. As we noted, we map MSAs into DMAs to calculate broadband level penetration at a DMA level. While overall mapping is pretty good (mean of 0.77 and median of 0.83) some of the DMAs are not very well matched. We now re-estimate the model for various levels of DMA coverage (coverage > 0.5 and coverage > 0.75). As we increase the coverage restriction, we lose data because the DMAs below the coverage specification drop out. We present the results in Table A-1.

Table A-1: Fixed Effect Regressions with DMA Coverage

Indep. Vars.	Coverage >0.5	Coverage > 0.75
Per Capita Broadband Penetration	0.125 (0.160)	0.21 (0.211)
Log(Population)	-0.77 (0.667)	-1.457* (0.843)
Median Income	0.027** (0.012)	0.012 (0.015)
% Whites	-0.074* (0.038)	-0.089* (0.047)
% Males	-0.303** (0.132)	-0.529* (0.282)
Median Age	0.025 (0.029)	-0.043 (0.074)
% 15-24 Year Olds	0.098** (4.287)	0.059 (8.011)
Wal-Mart Supercenters	-0.011 (0.006)	-0.011 (0.007)
Autoregressive Errors (AR1)	0.33	0.313
Time Dummies	Yes	Yes
Constant	29.11** (9.654)	54.9** (16.65)
No. of Obs.	168	114
R^2 (Within)	0.94	0.95

The dependent variable is per capita DVD sales. Standard errors are listed in parenthesis; ** and * denote significance at 0.05 and 0.10, respectively. All models use DVD-level fixed effects.

While the estimate on per capita broadband is smaller when DMA coverage is > 0.5 but reassuringly it is back to the original estimate when DMA coverage is > 0.75. However, they are not statistically significant. But it is probably due to smaller sample in these regressions. In fact, none of the variables are significant at 5% any more when DMA coverage is > 0.75. However, results are still consistent in sign with the finding that per capita broadband penetration aids the sale of DVDs.