Public Knowledge

Statement of Gigi B. Sohn, President, Public Knowledge
Before the Federal Communications Commission
National Broadband Plan Workshop: The Role of Content in the Broadband Ecosystem
September 17, 2009

I'd like to thank the Commission for inviting me to speak at today's workshop on the role of content in the broadband ecosystem. There's little doubt that online content produced by the entertainment industry serves as an important driver for both broadband adoption and utilization, and I expect that content to be the focus of most of my fellow panelists. In my comments, I hope to first provide a different perspective--one that highlights the important work being done by innovators, independent and amateur creators — and to then sound a note of caution. Any measures taken to protect content online must take into account all creators in the broadband ecosystem, must not obstruct the free flow of information online and must protect the privacy of end users.

Online content extends far beyond the boundaries of Hollywood. User generated content, in its many forms, has enriched the lives of Americans and rivals studio content in terms of popularity. In the online video market, for example, nearly 42 percent of all videos viewed online are hosted by YouTube.[1] Of the tens of billions of videos that YouTube streams each month, nearly one fourth of the top 100 most viewed videos are clips created by end users, not studios.[2]

As you might imagine, not only is user-generated content popular--it also serves a diverse set of needs. Amateur podcasters inform the public about events that are ignored or

[1] "U.S. Online Video Market Soars in July as Summer Vacation Drives Pickup in Entertainment and Leisure Activities Online," comScore, August 27, 2009 (http://www.comscore.com/Press_Events/Press_Releases/2009/8/U.S._Online_Video_Market_Soars_in_July_as_Summer_Vacation_Drives_Pickup_in_Entertainment_and_Leisure_Activities_Online).
[2] "YouTube's Most Popular Clips: Still Mostly Free," *All Things Digital*, August 12, 2009 (http://mediamemo.allthingsd.com/20090812/youtubes-most-popular-clips-still-mostly-ad-free/).

underreported on by the national news media.[3] Universities showcase their course materials and lectures via "OpenCourseWare" to members of the general public, many of whom would normally not possess the means to pursue higher education in a traditional setting.[4] Even Presidential candidates have harnessed the power of user-generated content, encouraging users to submit debate questions via YouTube in 2008.[5] And yes, many users, "vidders" and mashup artists use copyrighted clips to express themselves through commentary, criticism, and parody under the legal doctrine of fair use.

Meanwhile, innovative technology companies are empowering users to access and create the content of their choice in exciting new ways. Apple's iTunes provides easy access to the best of big studio content as well as user-created podcasts.[6] Boxee's "app box" allows users to automatically promote the media they enjoy to friends, putting "traditional" web video sources like Hulu, CNN, CBS, and Comedy Central on equal footing with user generated sources like YouTube, BitTorrent, and online video networks like TWIT.tv and Revision3.[7] Hardware manufacturers are embedding web interfaces into new HDTVs to allow easy-to-use online video streaming from the comfort of the consumer's couch.[8] Clearly, a great deal of creativity and innovation is taking place on the Internet, and up until now, it's been a level playing field. This innovation must be allowed to continue without undue restraints so that users can reap the

[3] "News Unfiltered: YouTube Embraces Citizen Journalism," *Ars Technica*, May 20, 2008 (http://arstechnica.com/old/content/2008/05/news-unfiltered-youtube-embraces-citizen-journalism.ars).
[4] "MIT's OpenCourseWare Project Continues Apace," *The Chronicle of Higher Education*, March 23, 2007 (http://chronicle.com/article/MIT-s-OpenCourseWare-Project/15958).
[5] "In Obama-McCain Race, YouTube Became a Serious Battleground for Presidential Politics," *U.S. News and World Report*, November 7, 2008 (http://www.usnews.com/articles/news/campaign-2008/2008/11/07/in-obama-mccain-race-youtube-became-a-serious-battleground-for-presidential-politics.html).
[6] *See* Apple Inc. website (http://www.apple.com/itunes/whats-on/).
[7] "Boxee Launches Update, 'App Box,' and Hulu RSS Support," *TUAW*, March 6, 2009 (http://www.tuaw.com/2009/03/06/boxee-launches-update-app-box-and-hulu-rss-support/).
[8] "Internet-Ready TVs Usher Web Into Living Room," *Wall Street Journal*, January 5, 2009 (http://online.wsj.com/article/SB123111603391052641.html).

benefits of increased choice, affordability and greater convenience and exercise their right to free speech and public discourse.

Before Taking Action, the FCC Must Collect Empirical Data About Online Infringement and its Actual Effect on the Entertainment Industries

Since assuming the office of Chairman of the Commission, Julius Genachowski has stated numerous times that he will ensure that decisions made by the FCC are data-driven.[9] With this in mind, the Commission should seek out independent empirical studies that quantify the extent of, and actual harm caused by, online copyright infringement. No policymaker can deliberate a solution unless she fully understands the problem. Though the content industry has cited studies in the past that purportedly address the financial losses resulting from online infringement, many of these industry-funded studies have proven unreliable. One widely cited study produced by research firm L.E.K. for the Motion Picture Association of America (MPAA), claimed that 44 percent of all Internet traffic on University networks was infringing in nature.[10] Three years after releasing the study, the MPAA was forced to admit that the 44 percent figure was inaccurate--the correct figure, according to the MPAA, was 15 percent.[11] EDUCAUSE, however, asserts that the correct figure is, in fact, 3 percent.[12] And while Congressman Arlen Specter asked the MPAA to provide its methodology for the study to Congress in 2006, the MPAA has yet to make this information available.[13]

---

[9] "The GigaOm Interview: FCC Chairman Julius Genachowski on Mobile, Broadband, iPhone and Innovation," *GigaOm*, August 3, 2009 (http://gigaom.com/2009/08/03/the-gigaom-interview-fcc-chair-julius-genachowski/).
[10] "MPAA Statement on Motion Picture Industry Losses Due to Piracy Among College Students," Motion Picture Association of America, January 22, 2008 (http://www.mpaa.org/press_releases/lek%20college%20student%20data_f.pdf).
[11] "Movie Industry Admits Error in Study on Campus Piracy," *The Chronicle of Higher Education*, January 22, 2008 (http://chronicle.com/blogPost/Movie-Industry-Admits-Error-in/3632).
[12] "MPAA Admits Piracy Study Flawed," *Copyright Advisory Network*, February 13, 2008 (http://librarycopyright.net/wordpress/?p=75).
[13] *Ibid.*

If the FCC is to address the matter of online copyright infringement, it must first seek out reliable, empirical studies produced by impartial third parties. In fact, many such studies are already available. In a Harvard Business School working paper on the topic of file-sharing and copyright, Felix Oberholzer-Gee and Koleman Strumpf embarked upon a literature review of studies published on the topic of online copyright infringement.[14] The study is attached to this statement. Among the studies they list, many differ from those cited by the entertainment industry and assert that the impact of infringement on sales of entertainment products is far less negative than is often acknowledged. Some of these studies even suggest that online file-sharing does not effect sales, or that online file-sharers are more likely to buy entertainment products than those who do not engage in file-sharing, which suggests a positive, rather than negative, relationship between file-sharing and sales.[15] The Commission should consider all available data when researching the impact that file-sharing has on the entertainment industry, including those studies cited by Oberholzer-Gee and Strumpf.

Copyright Filters Are Not Suitable for Use on ISP Networks

As the comments submitted in this proceeding attest, many in the entertainment industry are looking to automated copyright filtering technologies to solve the problems caused by the unlawful exchange of copyrighted content online.[16] While such technologies are being heralded as a fail-safe solution, they are, in fact, critically flawed. Simply put, a copyright filter is a blunt

[14] *See* "File Sharing and Copyright," Harvard Business School, May 15, 2009, p. 35-37 (attached) and "The Effect of File Sharing on Record Sales: An Emprical Analysis," Felix Oberholzer-Gee and Koleman Strumpf, December 12, 2006 (attached).
[15] *Ibid.*
[16] *See Joint Comments of American Federation of Television and Radio Artists, AFL-CIO, MPAA et al.*, In the Matter of A National Broadband Plan for Our Future, GN Docket No. 09-51 (all referenced comments below can be found in this docket unless otherwise noted); *Comments of Songwriters Guild of America*; *Comments of Arts+Labs* at 6 (urging the commission to "stand by its existing four principles, including the right to reasonably manage the networks" to among other things prevent copyright infringement); *Comments of the Entertainment Software Association* at 4 (urging the Commission to clarify that the Internet Policy Statement permits network operators to use "reasonable network management to combat online piracy"); and *Comments of Walt Disney Company*.

instrument and is not suitable for use on an ISP network. Public Knowledge has already submitted for the record a whitepaper that these flaws in great detail. I'd like to briefly address a few of the reasons why copyright filters are not a viable solution.[17]

A close analysis of copyright filtering technology reveals that ISP-level filtering will be ineffective at best and at worst, harmful to the network, end users and the goals of the National Broadband Plan. By virtue of their design, filters will be both overinclusive and underinclusive--that is to say that they will fail to identify all unlawful uses of content while blocking lawful uses.[18] As a result, automated filtering will block, delay or degrade lawful content, including but not limited to content that makes fair use of copyrighted content for the purposes of parody, satire and critical commentary. In so doing, copyright filters will discourage citizens from exercising their free speech rights online. While big studio content is important, we should recognize that user-generated content is an equally, important part of the broadband ecosystem and as such, it deserves an equal degree of protection. For this reason, copyright filters, which would automatically block all uses of copyrighted content online, including those aforementioned fair uses, are not appropriate tools to curb infringement.

A. Copyright Filters Will Harm the Network and User Privacy

Copyright filtering will also alter the behavior of data networks on a fundamental level, slowing down traffic, impeding the operation of high-latency applications and compromising the privacy of all Internet users.[19] In so doing, copyright filters will discourage investment in the Internet ecosystem, prevent innovators from developing exciting new applications, dissuade

[17] *See* "Forcing the Net Through a Sieve: Why Copyright Filtering is not a Viable Solution for U.S. ISPs," Public Knowledge, July 22, 2009 (http://www.publicknowledge.org/pdf/pk-filtering-whitepaper-200907.pdf).
[18] *Ibid.*, 7-24.
[19] *Ibid*., 25-37.

users from fully utilizing their broadband connections and raise the cost of access for consumers--all the while undermining some of the most important goals of the National Broadband Plan.[20]

B. Copyright Filtering is Likely to Result in Unintended Consequences

Copyright filtering at the ISP level could also result in unintended consequences that will likely result in greater harm than good to its proponents. Users will likely devise methods for circumventing the filter, in order to access the content of their choice without interference.[21] The methods that will likely be used--encryption and protocol obfuscation--will decrease the efficiency and speed of the network.[22] What's more, the architects of the filter will be forced to pour a tremendous amount of resources into research and development in order to thwart those users who seek to circumvent the filter, resulting in increased costs for both the user and provider.[23]

In addition, the core technology behind copyright filtering also holds the potential to be misused for purposes of censorship. The timeliest example of this sort of misuse comes from Iran, where content filtering technologies were used to block access to popular websites like Twitter, YouTube and Facebook, in the wake of widespread political unrest.[24] Apparently, the Iranian government first installed this hardware for the purported purpose of blocking pornography, citing "lawful intercept"—an internationally-recognized concept that "relates to intercepting data for the purposes of combating terrorism, child pornography, drug trafficking and other criminal activities carried out online."[25] This example illustrates that the act of filtering

[20] *See* "FCC Launches Development of National Broadband Plan," April, 8, 2009 (http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-289900A1.pdf).
[21] *Ibid*., 29-37.
[22] *Ibid*.
[23] *Ibid.*
[24] "Iran's Web Spying Aided by Western Technology," *The Wall Street Journal*, June 22, 2009 (http://online.wsj.com/article/SB124562668777335653.html).
[25] *Ibid.*

is a slippery slope. While filtering technology might be deployed to serve a legitimate purpose—be it to stem the flow of child pornography or illegally obtained copyrighted content—when placed in the wrong hands, that same technology can become a highly effective instrument of private or governmental censorship. Once such technology is installed at the ISP network level, it is all too easy to use that hardware for undemocratic ends.

C. Copyright Filtering is Content Management, Not Network Management

In the context of the National Broadband Plan, the entertainment industry is urging that copyright filters be considered a form of reasonable network management, and as such would not violate the FCC's broadband principles or other principles of openness. This characterization is misleading and obfuscates the manner in which filters operate. Copyright filters decide how data packets should be treated based on the content that they carry. As such, copyright filtering is a form of content management rather than network management and has no place in the National Broadband Plan.[26] Notwithstanding the question of whether the FCC has the authority to mandate copyright filters (discussed below), the Commission should not unduly burden ISPs with technology mandates and to the extent that the Commission addresses network management in the context of the National Broadband Plan, only those methods that are neutral and nondiscriminatory should be permitted.

A Three Strikes Regime Would Allow Private Companies to Decide Who Should Have Access to the Internet

Another blunt instrument that is being recommended by some in the entertainment industry is the so-called "graduated response" or "three strikes" regime.[27] Under this system, ISPs would be compelled to kick users off of their networks after receiving three notices from

[26] *Ibid.*, 5.
[27] "Digital Piracy Spreads, and Defies a Fix," *the New York Times*, April 6, 2009 (http://www.nytimes.com/2009/04/07/business/media/07piracy.html).

copyright holders indicating that a user was suspected of sharing content in an unlawful manner.[28] Never mind the industry's poor track record with regard to the accuracy of such notices:[29] three strikes lets the entertainment industry decide who gets kicked off the Internet based on a mere accusation of unlawful activity--without any judicial finding that infringement actually took place. Three strikes does not contemplate any judicial review and users would have no legal recourse or ability to challenge the accusations made. If we implemented such a regime devoid of due process in this country, we would be allowing private companies to cut citizens off from the most effective conduit for civic engagement, economic opportunity and education that is currently available, based purely on allegation. Indeed, the French Constitutional Council recently struck down a three strikes regime in that country, citing the Declaration of the Rights of Man and of the Citizen, a founding document of the French Revolution.[30] "…[W]hereas under section nine of the Declaration of 1789, every man is presumed innocent until he has been proven guilty, it follows that in principle the legislature does not establish a presumption of guilt in criminal matters," the Council wrote in its ruling.[31]

Needless to say, such a mandate, if instituted in this country, would be contrary to the goals of the National Broadband Plan.

[28] "Music Industry to Abandon Mass Suits," *Wall Street Journal*, December 19, 2008 (http://online.wsj.com/article/SB122966038836021137.html).
[29] "Flunking File Swappers: Inside the RIAA's Anti-P2P Machine," *Ars Technica*, May 14, 2008 (http://arstechnica.com/old/content/2008/05/flunking-file-swappers-inside-the-riaas-anti-p2p-machine.ars).
[30] "Top Legal Body Strikes Down Anti-Piracy Law," *France 24*, June 10, 2009 (http://www.france24.com/en/20090610-top-legal-body-strikes-down-anti-piracy-law-hadopi-constitutional-council-internet-france).
[31] *See Decision n° 2009-580 of June 10th 2009*, The Constitutional Council, June 10, 2009 (http://www.conseil-constitutionnel.fr/conseil-constitutionnel/root/bank/download/2009-580DC-2009_580dc.pdf).

The FCC Does Not Possess the Authority to Mandate Either Copyright Filtering or a Three Strikes Regime

The FCC has tried to step into the business of copyright regulation before. In its Broadcast Flag ruling,[32] the Commission required that hardware that connects to public networks contain technology designed to prevent some kinds of copying. The FCC's ruling was reversed by the D.C. Circuit, which held that the agency exceeded its authority when it adopted rules requiring digital media devices to prevent some unauthorized (but not necessarily illegal) copying.[33] Before that, the Commission was reversed when it tried to mandate that all broadcasters add video description information to their programming.[34] In both cases, the D.C. Circuit held that the FCC had overstepped the authority given to it by Congress.[35]

If the FCC were to mandate the use of copyright filters on ISP networks or a "three strikes regime," it would once again be stepping outside of its jurisdiction. Neither of these mandates would constitute the regulation of "communication by wire or radio".[36] Instead, both of these scenarios would transform the FCC into a copyright agency and would, in practice, limit rights of fair use. The FCC's authority to regulate "communication" does not give it the general

[32] Digital Broadcast Content Protection, *Report & Order & Further Notice of Proposed Rulemaking*, 18 FCC Rcd. 23,550 (2003).
[33] *American Library Ass'n. v. F.C.C.*, 406 F.3d 689 (D.C. Cir. 2005) (Commission exceeded its authority by requiring that devices respect the "broadcast flag").
[34] *Motion Picture Ass'n of Am. v. F.C.C.*, 309 F.3d 796 (D.C. Cir. 2002) (Commission exceeded its authority by implementing video description rules).
[35] *Motion Picture Ass'n.*, 309 F.3d at 801 ("An agency may not promulgate even reasonable regulations that claim a force of law without delegated authority from Congress."). Also, in *American Library Ass'n.*, 406 F.3d at 698, the court explains,

> *The FCC, like other federal agencies, 'literally has no power to act … unless and until Congress confers power upon it.' La. Pub. Serv. Comm'n v. FCC, 476 U.S. 355, 374 … (1986). The Commission 'has no constitutional or common law existence or authority, but only those authorities conferred upon it by Congress.' Michigan v. EPA, 268 F.3d 1075, 1081 (D.C. Cir. 2001). Hence, the FCC's power to promulgate legislative regulations is limited to the scope of the authority Congress has delegated to it. Id. (citing Bowen v. Georgetown Univ. Hosp., 488 U.S. 204, 208 … (1988)).*

[36] *American Library Ass'n.*, 406 F.3d at 703 ("The Federal Communications Commission may not lawfully exercise jurisdiction over activities that do not constitute communication by wire or radio.").

authority, absent an express delegation of power from Congress, to regulate content.[37] Additionally, under any mandated scheme involving copyright filtering or three strikes, disputes would undoubtedly arise hinging on whether a particular unauthorized use is in fact an illegal one. The FCC has no authority to hear or decide this kind of dispute, and cannot require that network operators become "copyright cops."[38] It has no power at all to regulate copyright absent an express delegation of power. As Sen. Patrick J. Leahy, Chairman of the Senate Judiciary Committee has written, Title 47 "grants…the FCC…no express authority…to address the complex issues of intellectual property matters[.]"[39]The Commission should not heed calls for it to again exceed its authority by instituting policies that would, in practice, constitute copyright law.

Solutions and Remedies

If we wish to prevent content from being shared unlawfully online, we should first look to those existing strategies and remedies that have proven successful. The most effective solution will always be the simplest one: for the entertainment industries to make content widely available online at a fair price and to allow users to access that content at the time of their choosing and on their device of choice. The current state of the market serves as a testament to this fact. Movie studios have found great success selling and renting films through services like Netflix,[40] music labels have seen consistent growth in the sales of online music via services like iTunes[41] and literary publishers have generated great excitement through their support for

[37] *Motion Picture Ass'n.*, 309 F.3d at 801.
[38] *Am. Library Ass'n*, 406 F.3d at 702 ("[T]he Commission may not invoke its ancillary jurisdiction under Title I to regulate matters outside of the compass of communication by wire or radio.").
[39] Letter from Sen. Patrick J. Leahy, Chairman, Senate Judiciary Committee and Rep. F. James Sensenbrenner, Jr., Chairman, House Committee on the Judiciary, *et al*. to Michael K. Powell, Chairman, FCC (Sept. 9, 2002).
[40] "Netflix Boss Plots Life After the DVD," *Wall Street Journal*, June 23, 2009 (http://online.wsj.com/article/SB124570665631638633.html).
[41] "iTunes Sells 25% of All Music in the US, 69% of Digital," *Ars Technica*, August 18, 2009 (http://arstechnica.com/apple/news/2009/08/itunes-sells-25-of-all-music-in-the-us-69-of-digital.ars).

devices like the Amazon Kindle.[42] Those studios who have threatened to hold back content unless certain, unproven protection mechanisms are put in place, as some did in the Broadcast Flag proceeding, only encourage users who seek that content to obtain it through unlawful means.[43] I applaud those content providers who have made their products available online at a fair price and hope that they will continue to expand their offerings as others follow suit. Within existing law, content providers also have access to a number of legal remedies for combating the unlawful use of their content online. The Digital Millennium Copyright Act's notice-and-takedown system provides a framework for removing infringing content that is hosted online. Some ISPs have agreed to pass infringement notices on to end users; a practice that the entertainment industry itself has found discourages users from engaging in unlawful conduct in the vast majority of cases.[44] And as always, the entertainment industry has the ability to combat the unlawful sharing of content by cutting that content off at the source, by targeting large-scale infringers, hard-goods counterfeiters and other unlawful providers who make infringing content widely available.

Conclusion

Ultimately, our goal should be to encourage the creation, sale and use of content online, which in turn, will further the goals of the National Broadband Plan. Digital entertainment is a valuable driver for both broadband adoption and utilization. By fully harnessing its potential, we can increase access to broadband for all Americans, educate users on how to make full use of

[42] "The Lessons From the Kindle's Success," *the New York Times*, August 12, 2008 (http://bits.blogs.nytimes.com/2008/08/12/the-lessons-from-the-kindles-success/).
[43] *Comments of Viacom*, FCC Docket 02-230, at 1.
[44] "RIAA President: No Talk of Blacklisting File Sharers," *Cnet*, December 19, 2008 (http://news.cnet.com/8301-1023_3-10127313-93.html).

their connections, and improve the lives of citizens by delivering economic, civic and educational opportunities directly to their personal computers.

ATTACHMENT A

FILE SHARING AND COPYRIGHT

FELIX OBERHOLZER-GEE AND KOLEMAN STRUMPF

HARVARD | BUSINESS | SCHOOL

File-Sharing and Copyright

Felix Oberholzer-Gee
Koleman Strumpf

Working Paper

09-132

File-Sharing and Copyright[1]

Felix Oberholzer-Gee
Harvard University
foberholzer@hbs.edu

Koleman Strumpf
University of Kansas
cigar@ku.edu

May 15, 2009

1. Introduction

The advent of file-sharing technology has allowed consumers to copy music, books, video games and other protected works on an unprecedented scale at minimal cost. In this essay, we ask whether the new technology has undermined the incentives of authors and entertainment companies to create, market and distribute new works. While the empirical evidence of the effect of file sharing on sales is mixed, many studies conclude that music piracy can perhaps explain as much as one fifth of the recent decline in industry sales. A displacement of sales alone, however, is not sufficient to conclude that authors have weaker incentives to create new works. File sharing also influences the markets for concerts, electronics and communications infrastructure. For example, the technology increased concert prices, enticing artists to tour more often and, ultimately, raising their overall income.

Data on the supply of new works are consistent with our argument that file sharing did not discourage authors and publishers.[2] The publication of new books rose by 66% over the 2002-2007 period. Since 2000, the annual release of new music albums has more than doubled, and worldwide feature film production is up by more than 30%

[1] We would like to thank Josh Lerner, Scott Stern, Amitay Alter and participants in the NBER's 2009 Innovation Policy and the Economy Conference in Washington, D.C., for helpful comments.
[2] Copyright refers to a complex bundle of rights that includes the rights of authors (composers, lyricists) and publishers (for a detailed description of these contracts, see Towse 1999; Passman 2000). Throughout this essay, we use the term somewhat loosely, referring to all legal protections – including, for instance, the "neighboring rights" of performers – that encourage the creation, production, marketing, and distribution of works. Also, we neglect the tensions that exist in copyright between artist and publisher interests (see Towse, 1999; Gayer and Shy, 2006.)

since 2003. At the same time, empirical research in file sharing documents that consumer welfare increased substantially due to the new technology.

Over the past 200 years, most countries evolved their copyright regimes in one direction only: lawmakers repeatedly strengthened the legal protections of authors and publishers, raising prices for the general public and discouraging consumption.[3] Seen against this backdrop, file sharing is a unique experiment that considerably weakened copyright protections. While file sharing disrupted some traditional business models in the creative industries, foremost in music, in our reading of the evidence there is little to suggest that the new technology has discouraged artistic production. Weaker copyright protection, it seems, has benefited society.

In this essay, we discuss the currently available research that sheds light on the effects of file sharing, particularly in music where its effects have been most pronounced. We start by describing the new technology and how consumers are using it. Section 4 reviews the evidence that file sharing reduces the profitability of creating and selling new works. We discuss the importance of complements to original works in Section 5 and describe the artistic and corporate response to file sharing in section 6. The concluding section offers policy implications.

2. File-Sharing and Copyright

In setting copyright terms, lawmakers trade off the increased incentives to create protected works and the higher prices that consumers face when books, movies, and recordings must not be copied freely (Landes and Posner, 1989). As this description suggests, the lawmakers' task is a challenging one. Setting copyright terms in a manner that benefits society requires an answer to two questions. First, we need to know how much weaker the incentives to create new works would be in a regime with more

[3] In the United States, as elsewhere, the degree of protection has steadily expanded, from the modest Copyright Act of 1790, which offered 14 years of protection with a renewal period of 14 years, to the legislation passed in 1831 (28 years), 1909 (renewal extended to 28 years), 1976 (50 years after the author's death), 1992 (automatic renewal), and 1998 (70 years).

constrained copyright. Second, and equally important, is the question how producers would respond to weaker incentives. Would they offer fewer works? Or perhaps works of lesser quality? In this essay, we discuss what we know about these questions, using the advent of file-sharing as our example for a technology that considerably weakened copyright protection for music, movies, books and video games.

Weaker copyright is unambiguously desirable if it does not lessen the incentives of artists and entertainment companies to produce new works. To appreciate the impact of file sharing, we first need to know whether the technology did in fact reduce the profitability of creating, marketing, and distributing new works. Of course, we know that millions of consumers share billions of files without compensating artists or entertainment companies. But the fact that file sharing is popular tells us little about the impact of the technology on industry profits. At a price close to zero, many consumers will download music and movies that they would not have bought at current prices. This issue is likely to be important. In a sample of 5,600 consumers who were willing to share their iPod listening statistics, the average player held a collection of over 3,500 songs (Lamere, 2006). A full 64% of these songs had never been played, making it unlikely that these consumers would have paid much for a good portion of the music they owned. While it is difficult to say how representative this sample is, there is no doubt that trade groups such as the Business Software Alliance vastly exaggerate the impact of file sharing on industry profitability when they treat every pirated copy as a lost sale (Economist, 2005). The demand for titles is not completely price inelastic.

Weaker property rights can undermine industry profitability if consumers who would have purchased a recording obtain a free copy instead. The critical question is then whether consumers perceive protected and freely shared works as close substitutes. As the name suggests, substitutes are products that meet similar consumer demands. For two substitute goods, a price decline for one leads to a decline in the demand for the other.[4] For example, if we allowed mash-up artists to freely copy parts of an original song, consumers who regard the derivative work as a close substitute would be less likely

[4] A classic example is butter and margarine.

to buy the original.[5] However, if consumers learned to better appreciate the original through the mash-up, demand for the original work might actually increase. In this case, the two versions of the song are complements, two goods for which a decrease in the price of one leads to an increase in the demand for the other. A well-known example for two complements is music and iPods. As file-sharing eroded the effective price of music for a large group of consumers, demand for mp3-players soared, allowing Apple to benefit from consumers' increased willingness-to-pay for its line of products.[6]

In practice, it is often surprisingly difficult to predict whether new products and technologies are complements or substitutes. As a result, we can often not be sure how changes in copyright will influence demand and industry profitability. The entertainment industry's history provides many examples of the difficulties involved in distinguishing substitutes, unrelated products, and complements. Music companies fought the introduction of radio in the 1920s, fearing the new medium would provide close substitutes to buying records. Since that time, the numerous attempts to bribe radio stations in the hopes of influencing playlists suggest the industry has come to see radio as an important complement to recordings (Coase, 1979). Similarly, the entertainment industry battled home taping[7] and the introduction of the VCR, arguing the new technology "is to the American film producer and the American public as the Boston strangler is to the woman home alone" (Valenti, 1982). Once the Supreme Court decided to protect technologies like the VCR, it did not take the industry long to discover that selling videotapes (and now DVDs) presents a major business opportunity.

Similar uncertainty surrounds file-sharing technology today. Some argue that protected works and copies on file-sharing networks are substitutes because consumers who would have bought the copyrighted version now choose to download a free copy instead. Others see protected works and copies on file-sharing networks as largely

[5] A mash-up is a song created out of pieces of two or more songs, usually by overlaying the vocal track of one song over the music track of another.
[6] Leung (2008) estimates that piracy contributes 20% to iPod sales.
[7] Stanley M. Gortikov, president of the Recording Industry Association of America (RIAA), explained in hearings before a House committee on 14 April 1982: "I'm scared, and so is my industry. Changing technology today is threatening to destroy the value of our copyrights and the vitality of the music industry. Our nemesis is home taping."

unrelated because they believe that file sharers are mostly consumers who are not willing to pay $10 for Taylor Swift's latest release. Finally, protected works and copies on file sharing networks are complements if consumers rely on the new technology to discover CDs or DVDs they want to purchase. These views need not be mutually exclusive. In a recent survey among file sharers, we found some support for all three conjectures (Oberholzer-Gee and Strumpf, 2005). 65% of respondents acknowledged they did not buy an album because they had downloaded it. An even larger group (80%) claimed they bought at least one album because they sampled it first on a file-sharing network. Fortunately, there is now a body of research that studies in a more systematic manner whether copyright protected works and copies on file-sharing networks are complements or substitutes. We will discuss this literature in section 4 of this essay.

Even if a weakened copyright regime turned out to reduce industry profitability, it is not obvious whether a decline in profits would undermine the incentives to create, market and distribute artistic works. Two considerations seem particularly important. First, as copyright weakens, the effective price of music, movies, and books falls and consumer willingness-to-pay for complements increases. If artists derive income from these complements as well, the overall incentives to produce new works might not decline. For instance, as music becomes effectively available for free, the price of concerts, a complement to music, is likely to rise, and artists who earn income from concerts might not be hurt by a decline in music sales (Krueger, 2005; Mortimer and Sorensen, 2005). Similarly, authors might be better able to supplement their income from books through speaking tours if many more readers are familiar with their writings.[8]

A second reason that a decline in industry profitability might not hurt artistic production has to do with artist motivations. The remuneration of artistic talent differs from other types of labor in at least two important respects. On the one hand, artists often enjoy what they do, suggesting they might continue being creative even when the monetary incentives to do so become weaker. In addition, artists receive a significant portion of their remuneration not in monetary form – many of them enjoy fame,

[8] Author Cory Doctorow, for instance, says:"I really feel like my problem isn't piracy. It's obscurity." (Rich, 2009).

admiration, social status, and free beer in bars – suggesting a reduction in monetary incentives might possibly have a reduced impact on the quantity and quality of artistic production.

There is no doubt that file sharing substantially weakened the protection of copyrighted works. Yet, as our discussion shows, the outcome of this experiment is far from certain. Three conditions need to hold for less-certain rights to undermine the incentives for artistic production: original works and copies on file-sharing networks must be reasonably close substitutes; artists and the entertainment industry must not be able to shift from previous sources of income to the (similarly profitable) sale of complements; and falling incomes must be an important-enough motivator for artists to reduce production. Only if all three conditions hold will file sharing hurt social welfare.

It might seem curious to some of our readers that we do not consider the welfare of artists and entertainment companies in our calculus. Our approach, however, reflects the original intent of copyright protection, which was conceived not as a welfare program for authors but to encourage the creation of new works. We know that stronger copyright protection can increase the market value of companies.[9] But these gains are a mechanism to raise social welfare, not the intended consequence.[10]

3. A Brief History of File-Sharing

To better understand the impact of file-sharing technology on copyright protection, it is useful to review the basics of file-sharing. In this section, we will also describe recent changes in technology and review the most significant legal challenges that companies providing file-sharing software faced to date.

File sharing relies on computers forming networks to allow the transfer of data. Each computer (or node) may agree to share some files, and file-sharing software allows

[9] Baker and Cunningham (2006), for example, estimate that a statue broadening copyright adds up to $39 million to the market capitalization of a typical firm.
[10] To frame our discussion in terms of efficiency (Pareto improvements), we argue that the relevant benchmark is the welfare of groups in a situation without copyright.

users to search for and download files from other computers in the network. Individual nodes are called clients if they request information, servers if they fulfill requests, and peers if they do both.

Shawn Fanning, an 18-year-old student at Boston's Northeastern University, started the file-sharing revolution when he released Napster in June of 1999 (table 1 provides a timeline). The software first allowed the freshman to trade music with his dorm mates. Prior to Napster, fans used search engines such as Lycos and music websites to download music. However, searching for files was cumbersome because the available music indices were often out of date. Many sites offered more broken links than hits. Napster was novel in that it maintained a central, dynamic index of all available files. This index was updated every time a user logged on or off. Thanks to its user-friendly interface and seemingly unlimited supply of music, the service gained 30 million users in its first year.

Napster's legal difficulties started not long after its initial release. In December 1999, the Recording Industry Association of America (RIAA) sued Napster for contributory and vicarious copyright infringement (*A&M Records, Inc. v. Napster, Inc.*, 239 F.3d 1004 (9th Cir. 2001).[11] Two years and one appeal later, the Ninth Circuit Court of Appeals ruled against Napster, arguing the service's central directory of files gave its makers knowledge of and the ability to control user infringement. Unable to filter files from the network, Napster shut down. However, putting Napster out of business proved easier than ending file sharing. Most Napster users simply switched to second-generation peer-to-peer services, and they were joined by millions of file-sharing novices. Three major networks eventually developed: eDonkey; FastTrack, a network used by KaZaA and Grokster; and Gnutella, an open-source network for clients such as Bearshare, Gnucleus, LimeWire, and Morpheus.

The Circuit Court decision also proved influential for the further technological development of file-sharing services. If peer-to-peer companies had no direct knowledge

[11] A party is liable for contributory infringement if it knows of the infringing activity and materially contributes to it. Vicarious infringement occurs when the indirect infringer benefits financially from the infringement.

of and control over infringing activities, many in the industry believed, file-sharing services might be protected by the Supreme Court's Betamax decision (*Sony Corp. of America v. Universal City Studios, Inc.*, 464 U.S. 417 (1984). The decision holds that companies are not liable for customers' acts of copyright infringement if their technology is capable of substantial non-infringing uses. In the Sony case, the Court estimated that about 9% of VCR recordings were of TV shows that consumers had taped to watch at a later time and that the producers of these shows did not object to time shifting. This was sufficient to shield Sony from liability.

Convinced that peer-peer technology had substantial legal uses – for example the exchange of files that were in the public domain or the sharing of documents within a company – second-generation file-sharing services eliminated centralized indices (Oberholzer-Gee, 2006). In these systems, users first connect to a single peer using a specific internet protocol. The peer then tells the software about other peers in the network, in effect decentralizing the search and download processes and making it impossible for peer-to-peer companies to know whether users trade copyrighted materials. At first, this strategy appeared to work. When the RIAA sued the makers of Grokster, a branded version of KaZaA, and Morpheus for contributory and vicarious copyright infringement, District Court Judge Stephen V. Wilson ruled that the two companies could not be held liable (*MGM Studios, Inc. v. Grokster, Ltd.,* 259 F. Supp. 2d 1029 (D. Cal. 2003): "All Napster search traffic went through, and relied upon, Napster… [But] when users search for and initiate transfers of files using the Grokster client, they do so without any information being transmitted to or through any computers owned or controlled by Grokster… If either defendant closed their doors and deactivated all computers within their control, users of their products could continue sharing files with little or no interruption."

The entertainment companies appealed the case, but the circuit court upheld the earlier decision, affirming that decentralized peer-to-peer systems met the standard set in Sony. On June 27, 2005, however, the Supreme Court overturned the Ninth Circuit, sending the case back to the district court for further consideration (*MGM Studios, Inc. v. Grokster, Ltd.*, 545 U.S. 913 (2005): "Because substantial evidence supports MGM on all

elements, summary judgment for the respondents was in error. On remand, reconsideration of MGM's summary judgment motion will be in order." The justices ruled that a company that distributed a device "with the object of promoting its use to infringe copyright" could be liable for the resulting illegal acts. The Court argued that Grokster and Morpheus had wanted to be the next Napster, showing their goal was to induce copyright infringement.

The Supreme Court's decision led most peer-to-peer companies to settle with the entertainment industry. An exception was LimeWire, a service that continues to operate to this day. LimeWire argues that its software provides substantial legal uses. For example, the company operates a digital music store that offers 500,000 songs, many of them from independent bands. And LimeWire insists that it does not induce consumers to infringe copyright. The RIAA filed a lawsuit against LimeWire in April 2006. At the time of this writing, no decision has been reached, leaving open the question whether services such as LimeWire are protected by the standard set in Sony. At the same time, several second-generation file-sharing programs such as Ares Galaxy and eMule, the former eDonkey, continue to be available as open-source software.

While pursuing the developers of peer-to-peer software in the courts, the RIAA also started suing P2P users who shared a large number of files—typically more than 1,000 tracks—starting in 2003. The association hoped its actions would help reverse the common view that file sharing was a legitimate activity. In a Pew Internet & American Life Project survey in 2000, 78% of internet users who downloaded music did not think they were stealing. A majority of the general internet population held the same view (Lenhart and Fox, 2000). By the end of 2008, the industry had brought suits against more than 35,000 file sharers. Most cases were settled, typically for a few thousand dollars.

In a surprising shift in legal tactics, however, the RIAA announced in December 2008 that it had decided to drop its campaign against individual file sharers. Instead, the industry hoped to collaborate with internet service providers (ISPs) to stop the transfer of copyrighted materials. The trade group has worked out preliminary agreements with

major ISPs under which it will send an email to the provider when it finds that customers share copyright-protected files (McBride and Smith, 2008).

While the RIAA had some success putting peer-to-peer companies out of business, file-sharing technology continued to evolve. The most important technical advance was the emergence of BitTorrent. BitTorrent file requests differ from classic full-file HTTP requests in that the client makes many small data requests, similar to internet telephony which breaks voices into small packets of data. In addition, BitTorrent downloads follow a "rarest-first" order which ensures high availability of files across the network. To start the downloading process, users first obtain a torrent, a small file that contains metadata about the file to be downloaded and information about the tracker, the computer that coordinates the file distribution. Torrents are hosted by a fairly small number of websites. The Pirate Bay is probably the best-known among them. The torrent allows the client to connect to the tracker, from which it receives a list of peers that currently transfer pieces of the file. As more peers connect to a tracker, they form a swarm and begin to trade pieces with one another.

The advent of BitTorrent is significant for a number of reasons. First, the improved technology significantly reduces download times. While the user experience varies significantly, it has now become possible to download a feature film in less than two hours. Second, the technology forces users to share the parts of files that they already own while they download the remaining bits. This procedure reduces the opportunity to free-ride that plagued older P2P systems. The protocol also rewards users who contribute more generously, for instance by allowing faster downloads for those with greater upload capacity. Sharing digital files was always non-rivalrous because the original owner of a file retained his copy. But more efficient file distribution systems such as BitTorrent have now also succeeded in reducing the negative externalities that users impose on one another when they transfer files.

a. Size of File-sharing Activity

Measuring the extent of file sharing is challenging (Karagiannis, 2003; Pasick, 2004). Initial studies relied on surveys to determine the number of users, but this

approach is flawed because respondents are likely to understate their participation in a potentially illegal activity. More worrisome, the level of understatement likely varies over time based on the legal climate and peer effects among teens. Surveys are also unreliable because it is difficult to survey a representative population of file sharers and due to recall issues.

A better approach involves identifying the packets traversing computer networks. These studies use special hardware to classify messages that are sent along networks by source, such as web (http) traffic, email, or file sharing. This approach is taxing because of the scale of the activity (ISPs typically handle many gigabits per second), the changes in the predominant protocol file-sharing protocol, and the recent move to encryption, which makes packets unreadable to unauthorized observers. Measurement studies employ three basic approaches to deal with these technical issues: flow monitors, deep-packet inspection, and direct interface with file sharing users.

Flow monitoring analyzes unidirectional sequences of packets from one IP address to another at the router level (Shalunov and Teitelbaum, 2001). This approach inspects packets in a rather shallow way, relying primarily on header information such as IP protocol and an examination of ports. Flow monitoring can analyze a large amount of traffic, at the risk of misclassifying some of it. A detailed flow analysis of Internet2, the U.S. high-speed network which primarily connects universities, is available at the weekly level back to 2003 (Internet2 Netflow Statistics, 2009). Figure 1 shows that file sharing traffic on Internet2 has roughly grown by a factor of ten – from about 1 terabyte to about 10 terabytes – from 2003 through 2009.[12] While this growth has been fairly steady, during 2003-2005 there were large traffic dips during late spring and early summer as well as smaller drops during Christmas. These drops in file-sharing activity reflect school vacations, periods during which college students, who are among the highest file sharing users, leave their high-speed campus internet connections.

The second type of evidence comes from deep packet inspection. Rather than relying just on the packet header, this approach considers characteristics of the payload

[12]Karagainni, et al (2004) employ a similar methodology in studying Tier 1 ISP traffic. They conclude that file sharing did not decline over the period 2003-2004.

itself (Allot Communications, 2007). Packet inspection is the most accurate method of identifying file sharing, but the technique requires extremely sophisticated equipment since huge amounts of data must be analyzed. The deep-packet inspection company Sandvine has been monitoring file-sharing trends for several years. The company's reports show that file sharing accounted for between forty and sixty percent of all bandwidth usage over 2002-2008 (Sandvine, 2002-2007 and 2008ab). CacheLogic, another deep-packet inspection company, finds similar trends in global file activity (Ferguson, 2006). Figure 2 shows the growing role of file sharing over 1999-2006. By 2006 sixty percent of all consumer internet traffic was due to file sharing, a majority of which was composed of video files.

The final approach to measuring file sharing comes from studying peer-to-peer networks directly. Observers use a modified version of file-sharing software to connect to a large number of users on the network. Direct observation can provide fine-grained information such as the identity of files. A difficulty with this approach is that direct observers need to monitor an ever-changing representative sample of networks. The leading practitioner is BigChampagne, a company which monitors individual search requests as well as the content of folders that users share. Figure 3 shows BigChampagne's count of the monthly number of U.S. file-sharing users from mid-2002 through mid-2006.[13] By the end of this period there were about seven million simultaneous users in the U.S. Unfortunately, more recent figures are not publicly available. As with the earlier data on file sharing traffic, there is evidence of secular growth as well as reductions, or least a lack of growth, during summer months. The data also suggest one reason why the RIAA has abandoned its approach of suing individual file sharers. In figure 3, it is difficult to ascertain an effect of the beginning of the 2003 lawsuit campaign (Manuse, 2003). While the overall campaign may have been disappointing from the RIAA's perspective, research has documented a short-run decline in the number of files shared and in downloading activity in response to the first round of

[13] User counts from the independent file-sharing site slyck.com largely mirror these numbers.

lawsuits (Bhattacharjee et al., 2006). In contrast, the *Grokster* Supreme Court decision in 2005 does not appear to have had much impact on the user-base.[14]

The data from these disparate sources paint a similar picture for trends in U.S. file sharing. There has been secular growth in both the amount of file sharing and the number of users. This upward trend has largely been unaffected by shifts in technology and the legal environment. At the same time, figure 1 shows that the intra-year cycle in file sharing observed in the early years has started to disappear. As broadband has proliferated outside of universities and to the home, young file-sharing users no longer rely on their university connections during the school year to download files.

b. Consumer Behavior

Three facts about consumer behavior on file-sharing networks strike us as particularly interesting: the narrow focus on a limited set of files; the truly global nature of file sharing; and the continued importance of industry marketing efforts. We discuss each of these in turn.

Users share a wide variety of files on P2P networks. Table 2 shows the distribution of a selected list of genres on a popular P2P network and compares it to store sales of these albums and downloads of songs (for a detailed description of the sample, see Oberholzer-Gee and Strumpf, 2007). Genres such as R&B, Rap and New Artists are overrepresented, while there is comparatively little country music. Looking at what users actually download, it is striking to see how dominant the Current Alternative category is. Almost one half of all downloads are transfers of songs in this genre. The data in Table 2 reflect the supply of music files in 2002, the stone age of file sharing. We don't know of any study that has systematically compared changes in content over time.

While the supply of files is vast, peer-to-peer users download only a small share of the files that are available. In our sample of 10,271 different music tracks, 60% are never downloaded over a period of 17 weeks, and 81% are downloaded less than 5 times,

[14]Similarly, Ferguson (2006) shows that eDonkey traffic levels were largely unaffected in 2006 when legal authorities forced the closure of a large network of servers.

a number that is just slightly above the mean.[15] Even in movies, where the number of available titles is far smaller, there is a notable focus on the most popular titles. Table 3 shows the availability of and the demand for movies on Mininova, a popular BitTorrent index site. Not surprisingly, the top DVD rentals are all in high demand (column 2). But demand trails off markedly for older titles, many of which are not even available. A point in case is Malin Akerman, a Swedish actress voted number one on IMDB's starmeter in early 2009. Akerman was one of the stars of the then popular movie Watchmen. As the last column in Table 3 shows, there was in fact significant demand for that release. But movie buffs with an interest in Akerman's previous films faced rather slim pickings. At the height of the popularity of Akerman, four of her last ten movies were unavailable and there was no demand for two additional films.[16] As in music, downloading activity for movies is heavily concentrated on current releases and the supply of titles is substantially broader than the demand.

A second interesting fact about consumer behavior on peer-to-peer networks is the truly global nature of file-sharing. Table 4 shows the top countries for users and downloads (from Oberholzer-Gee and Strumpf, 2007). Interactions among file sharers transcend geography and language. U.S. users download only 45.1% of their files from other U.S. users, with the remainder coming from a diverse range of countries including Germany (16.5%), Canada (6.9%) and Italy (6.1%). One implication of these interactions is that national regulations of file sharing will only have limited bite. For instance, if the RIAA and domestic ISPs discouraged U.S. users from making files available, as they currently hope to do, users in the U.S. could simply download files from other countries.

A final observation concerns the marketing efforts of the entertainment industry. In view of the vast supply of music and videos on the internet and the many electronic networks connecting individuals, it might seem reasonable to expect that the industry's

[15] Our sample is drawn from SoundScan charts, which include all commercially relevant albums. Though some of the albums in the sample had low sales, many in fact were very high sellers.
[16] The concentration of movie downloads in part reflects the current BitTorrent technology. Index sites, which list the files available for download, typically de-list a title when no one is sharing a complete copy for some length of time. As a result, less popular movies become often unavailable, as are older movies since the number of shared copies tends to decline over time.

ability to draw attention to particular products has been greatly diminished. But the data in figure 4 tell a different story. The graph shows downloads and sales of the popular Eight Mile soundtrack, a commercial success directed by Curtis Hanson, starring rapper Eminem. Note that the recording leaked about 6 weeks prior to the official album release, with Eight Mile songs becoming available on peer-to-peer networks. But, interestingly, the level of downloads remained small until the industry marketing campaign began. Unless the industry drums up support for a new release, it is apparently difficult to give it away for free. This pattern of downloads and sales is fairly typical in our data. Contrary to the view that the entertainment industry has lost its ability to create value in a networked world, these data suggest the recording industry remains unrivaled in its ability to steer consumer attention.

4. Does File-Sharing Reduce the Sale of Copyrighted Materials?

The sharing of information goods such as music, movies, and books has been the subject of a substantial literature, both theoretical and empirical. Theory has most often focused on two competing intuitions about the effects of file sharing. A first is obvious: copying hurts producers because consumers who would have purchased a product now obtain it for free. But there is a second effect that runs counter to this idea. Because consumers anticipate sharing products, their willingness to pay (and hence producer profits) might actually increase. For example, a family might be willing to buy an expensive videogame because the parents know that several children will enjoy playing it. The theoretical literature has successfully identified a number of factors that influence the balance of these two effects, including the relative cost of producing information goods and sharing, the variation in the size of groups that share protected works, as well as the diversity in consumer valuations and the correlation of valuations within a sharing group (Novos and Waldman, 1984; Johnson, 1985; Liebowitz, 1985; Besen and Kirby, 1989; Bakos, Brynjolfsson and Lichtman, 1999; Varian, 2000). Depending on the importance of the relevant parameters, theoretical modeling predicts that file-sharing can

either hurt or help producers (for a review of theory papers, see Peitz and Waelbroeck, 2003).

Because the theoretical results are inconclusive, the effect of file sharing on industry profitability is largely an empirical question. We summarize the findings of some of the major studies in table 5. As the list shows, the results are decidedly mixed. There are two studies that document a positive effect of file-sharing on sales: Andersen and Franz (2008) for a representative sample of Canadian consumers and, more narrowly, Gopal et al. (2006) for the effect of sampling on CD sales.[17] The majority of studies finds that file sharing reduces sales, with estimated displacement rates ranging 3.5% for movies (Rob and Waldfogel, 2007) to rates as high as 30% for music (Zentner, 2006).[18] A typical estimate is a displacement rate of about 20%. One implication of these results is that developments other than file sharing must have had a profound impact on sales. For music, the popularity of new types of (internet-based) entertainment and the end of the transition from LPs to CDs are leading explanations for the overall decline in sales (Hong, 2004; Oberholzer-Gee and Strumpf, 2007). While many studies find some displacement, an important group of papers reports that file-sharing does not hurt sales at all (Tanaka, 2004; Bhattacharjee et al., 2007; Oberholzer-Gee and Strumpf, 2007; Smith and Telang, 2008). And even among the studies that show some displacement, there tend to be important subsamples that were not affected. For example, Rob and Waldfogel (2006) find an average displacement effect of 20% but report that file sharing had no impact on hit albums.

In order to better understand why file-sharing studies come to varying conclusions, it is instructive to consider a number of challenges in the empirical literature.

Choice of Sample – Researchers frequently rely on convenience samples, typically students, to estimate the effect of file sharing on sales. This is problematic because surveys show high school and college students to be among the most active file

[17] Gopal et al.'s (2006) results are consistent with the theoretical findings in Peitz and Waelbroeck (2006).
[18] An outlier is Liebowitz (2008) who reports a displacement rate of more than 100% for a selection of U.S. music markets.

sharers (Pew Internet Project, 2003). As a result, the displacement rates documented in these studies are likely to lie above the true population rates. Convenience aside, we suspect that many scholars rely on unrepresentative samples of students because it used to be almost impossible, and remains often expensive, to gain access to representative sales data. For instance, U.S. sales data for music, traditionally shared among record companies, has only become available to researchers in the most recent years. And even today, short-term subscriptions to industry databases can cost thousands of dollars, excluding scholars with more limited research budgets.[19] To arrive at a more complete understanding of file sharing, increased collaboration between industry and academia – and the employment of representative samples – appears essential to us.

Measures of piracy – A key difficulty in interpreting the findings of many studies is that they rely on self-reported data or poor proxies for actual file sharing. As table 5 indicates, surveys with self-reported measures of piracy play a significant role in the literature. Unfortunately, we do not know much about the accuracy of survey data in the context of file sharing. As Zentner (2006) points out, some individuals might play down their file sharing because they understand it is illegal. On the other hand, if file sharing is hip, as is the case on many college campuses, students might exaggerate the activity. In Andersen and Frenz (2008), more than 10% of respondents who report having downloaded music do not provide the number of downloaded files, suggesting recall or perhaps response bias might also be an issue. In view of the popularity of survey-based measures of piracy, we consider it important for future research to establish their accuracy. If these data turn out to be reliable, they could play a major role in future research because survey data are simple and inexpensive to obtain.

Where survey data on piracy is unavailable, researchers tend to rely on crude proxies for file sharing such as internet penetration. In a number of studies, internet-related measures (penetration, user sophistication) also serve as an instrument for downloading. In our view, both usages are inappropriate. Internet penetration proxies

[19] Nielsen SoundScan, the dominant provider of record sales, offers an academic subscription for $10,000 a year. Nielsen VideoScan is even more expensive. Box office numbers for theatrical releases are freely available from Box Office Mojo, but learning about geographic variation in sales is more difficult. Fortunately, Nielsen Bookscan data are available at a reasonable cost.

for new forms of entertainment – think YouTube and World of Warcraft – that compete directly with music and traditional film consumption, yielding a negative bias in displacement studies. Given these fairly obvious shortcomings, why are there so few papers that use actual data on file sharing to measure its effect on sales? One reason, we believe, is that collecting data on file-sharing networks is labor intensive and often cumbersome. Sometimes it is necessary to gain the trust of individuals operating file-sharing servers. And automated measurement studies require considerable programming skills and knowledge of file-sharing software. These hurdles notwithstanding, it is disappointing to see how few social scientists have made the effort to collect data on actual behavior. Many scholars prefer to use widely available, but in our view inappropriate, proxies for file sharing. The resulting research is poorer for it. The situation in the social sciences is in marked contrast to the research in computer science where many studies carefully measure individual file-sharing activity (e.g. Leibowitz et al. 2002; Gummadi et al. 2003; Pouwelse et al. 2005; Liang et al. 2005a, 2005b; Dhungel, et al. 2008).

We emphasize these issues because the results in table 5 seem to suggest that measurement choices have a systematic impact on results. While the majority of papers reports some sales displacement, the four studies using actual measures of file sharing (Tanaka, 2004; Bhattacharjee et al., 2007; Oberholzer-Gee and Strumpf, 2007; Smith and Telang, 2008) find that file sharing is unrelated to changes in sales.

Unobserved heterogeneity – A common difficulty in studying the link between downloads and sales is that file sharing is endogenous. That is, there are factors, some of them unobserved by the econometrician, that influence both downloads and sales. For example, music lovers are likely to download more songs and they also buy a larger number of albums, making it look like there was a positive relation between file sharing and sales. To see this, consider figure 5, taken from Oberholzer-Gee and Strumpf (2005). In this graph, downloads (horizontal axis) appear to increase sales (vertical axis). But an alternative explanation is that the popularity of a release increases both file-sharing activity and sales: popular recordings are in high demand on the internet and in the store.

Difference-in-difference (DD) estimates and instrumental variable techniques are popular means by which scholars hope to break the link between unobserved factors and the estimated impact of piracy on sales. DD models yield unbiased estimates if the unobserved heterogeneity is time invariant. Unfortunately, time-varying unobserved factors appear to play a major role in file sharing. Comparing DD estimates with results that take into account how cohort characteristics change over time, Hong (2008) finds that DD estimates attribute the entire 2002 decline in record sales to Napster. Once changes in unobserved heterogeneity are taken into account, the sales displacement rate drops from 100% to 20%. Similarly, Oberholzer-Gee and Strumpf (2007) show that the combination of album and week fixed effects is insufficient to control for unobserved heterogeneity.

Instrumental variable techniques provide a potentially more promising way to identify the effect of file sharing on sales. As noted above, we are skeptical of attempts to use measures of broadband adoption or user internet sophistication as instruments. More promising identification strategies exploit technical aspects of file-sharing systems – the availability of BitTorrent indexing sites, for instance, fluctuates considerably over time for largely technical reasons – and shocks to the global supply of content. For example, Oberholzer-Gee and Strumpf (2007) exploit the fact that many files downloaded in the US come from Germany. During German school holidays, file sharing in the US becomes easier: download times are shorter, a greater fraction of searches lead to a successful download, and fewer download requests remain incomplete. Because German holidays are unrelated to U.S. music sales, the holiday shock makes a promising instrument. More generally, because file sharing is a truly global phenomenon there are many shocks that spread from country to country. Some of these will be unrelated to the domestic demand for entertainment, making them promising prospects in the quest for proper identification.

5. How Important Are Complementary Sources of Income?

Even if file sharing displaces sales, the weaker copyright regime need not undermine the incentives to produce new works if artists and entertainment companies can shift their earnings from selling music, games and movies to selling complements to these products. An interesting example is concerts. As Table 6 shows, concerts and merchandising have become an important source of income for major artists (Connolly and Krueger, 2006). Concerts and new recordings are complements. A recording becomes more enjoyable if one can reminisce about the time at the concert, and knowing the songs in advance might make the concert more enjoyable. In the presence of complementary goods, file sharing will have two opposing effects (for a formal model, see Mortimer and Sorenson, 2005). As the effective price of music falls close to zero, a larger number of consumers will be familiar with an album, driving up the demand for concerts. At the same time, artists have weaker incentives to tour because concerts are a less effective way to increase revenues from a new recording if a large fraction of the audience shares files. Which of these effects is more important? Figure 6 shows that concert prices rose much more quickly than the CPI, and the difference appears to have widened since the advent of file sharing (Krueger, 2005). More detailed evidence on the link between file sharing and concerts comes from Mortimer and Sorenson (2005). Studying 2,135 artists over a ten-year period, they also conclude that the demand for concerts increased due to file sharing. One way to see this is to ask how many CDs an artist needs to sell to produce $20 of concert revenue. This number fell from 8.47 in the pre-Napster era to 6.36 in the 1999 to 2002 period. Not surprisingly, artists responded to these incentives by touring more frequently. Overall, the shift in relative prices and activities led to a sharp increase in income for the typical artist included in the authors' dataset.

As these results show, income from the sale of complements can more than compensate artists for any harm that file sharing might do to their primary activity. We are not aware of empirical work that has looked at these effects in industries other than music. But the potential of complements to provide ancillary income is certainly not unique to the music industry. In film, for instance, the International Licensing Industry Merchandisers' Association (LIMA) estimates that Hollywood derives $16 billion

annually from sales of entertainment merchandise, a figure that exceeds the value of ticket sales (Film Encyclopedia, 2008).

The role of complements makes it necessary to adopt a broad view of markets when considering the impact of file sharing on the creative industries. Unfortunately, the popular press – and a good number of policy experts – often evaluate file sharing looking at a single product market. Analyzing trends in CD sales, for example, they conclude that piracy has wrecked havoc on the music business. This view confuses value creation and value capture. Record companies may find it more difficult to profitably sell CDs, but the broader industry is in a far better position. In fact, it is easy to make an argument that the business has grown considerably. Figure 7 shows spending on CDs, concerts and iPods. The decline in music sales – they fell by 15% from 1997 to 2007 – is the focus of much discussion. However, adding in concerts alone shows the industry has grown by 5% over this period. If we also consider the sale of iPods as a revenue stream, the industry is now 66% larger than in 1997. Obviously, these numbers are no more than a rough back-of-the-envelope calculation. A more serious investigation would take into account differences in profitability across music and concert sales as well as the decreased spending in other electronics categories (CD players, speakers, etc.) The point of the graph, however, remains: technological change will often lead to changes in relative prices and shifts in business opportunities. Focusing exclusively on traditional streams of revenue to arrive at a sense of how new technology changes welfare will typically be misleading.

6. **Does File-Sharing Undermine Artistic Production?**

In any evaluation of file sharing, a key question is whether financial incentives are needed to encourage artistic output.[20] While this is in large part an open question, several indirect pieces of evidence suggest that financial incentives play a smaller role in the

[20] In this respect, the arts are similar to the production of open source software where many programmers appear to work for little monetary gain (Lerner and Tirole, 2005).

creative industries than elsewhere in the economy.[21] For concreteness we will focus our discussion on popular music, but many ideas discussed here carry over to film, visual arts, writing, and high culture music (see Caves, 2000).

The economic prospects for the group of popular musicians as a whole are quite poor. An album selling a half million copies or more (a Gold Album) is considered successful. Typically, a few hundred albums reach this level each year. Yet over 50,000 albums are released annually, suggesting the chance of success is less than one in a hundred. Perhaps more strikingly, only 950 new albums sold more than 25,000 copies in 2007.

Moreover, it is difficult for musicians to earn substantial income from recorded music sales, regardless of the success of their album. This is in part due to the nature of recorded music contracts (Passman, 2000). Recording musicians are paid for album sales based on the product of a royalty rate and album sales. The royalty rate is quite low (usually about a dollar or two per album) and musicians are not paid this money until they recoup all expenses, primarily the advance which is typically applied to the cost of recording the album. If an earlier album did not sell well enough to pay for the advance, music companies often deduct the difference from future album payments under a system called cross-collateralization. Putting all this together, even a Gold Album may not provide a musician with an economic windfall.[22]

Given these poor prospects, why are there so many musicians? One explanation is that musicians enjoy their profession. Under this view, musicians take pleasure from creating and performing music, as well as aspects of the lifestyle such as flexible hours and the lack of an immediate boss. If this theory is correct, the economic impact of file sharing is not likely to have a major impact on music creation.

An alternative explanation is that popular music is a tournament, where a few artists collect most of the economic rewards. This view is rooted in the theory of superstars (Rosen, 1981). Superstars develop in industries with low marginal cost of

[21] The broader critique of Boldrin and Levine (2008) implies that for innovation to take place more generally, copyright and patents are not needed.
[22] For specific dollar totals from insiders in the music industry, see Albini (1994) and Love (2000).

production, little relation between output and quality, and quality-conscious consumers. This seems to be a reasonable model of popular music: it is relatively cheap to produce CDs and even cheaper to make digital albums. Each album produced provides the same quality level, and most consumers would rather listen to one very good album than a few albums of lesser quality. Under the superstar theory musicians essentially consider their job to be a lottery. With some small chance they will become a star. In 2007, the top one percent of new releases accounted for 82% of new-release sales. In a superstar environment, file sharing has a muted effect on music output. Even if the new technology had a marked negative effect on the returns to stardom, it is not likely to have big effect on the chances of becoming a star.[23]

Survey evidence (as well as the long lines of contestants hoping to be part of talent shows like American Idol) support these theoretical arguments. In a Pew study of 2,755 musicians and songwriters (Madden, 2004), over three-fourths of respondents reported having a paying non-music job.[24] These second jobs are the primary source of income for most musicians. Only 16% reported that at least sixty percent of their income derived from their music job, while 66% said they earned less than twenty percent of their income from music. The small income share is not simply due to spending few hours on music. Even among those who spent at least thirty hours a week on music-related activities, only 22% derived at least four-fifths of their income from music.

Overall production figures for the creative industries appear to be consistent with this view that file sharing has not discouraged artists and publishers. While album sales have generally fallen since 2000, the number of albums being created has exploded. In 2000, 35,516 albums were released. Seven years later, 79,695 albums (including 25,159 digital albums) were published (Nielsen SoundScan, 2008). Even if file sharing were the

[23]Consider a model in which individuals must choose between being a musician and some outside reservation job. If p is the probability of being a star, S the income (and non-pecuniary benefits) of being a star, NS the income of a non-star, and R the income from the reservation jobs, than the person decides to be a musician when,

$$pU(S) + (1-p)U(NS) \geq U(R)$$

where U(.) is a utility function and S>>R>NS. Even if file sharing has a large negative effect on S, this will only have a limited impact on the left-hand side presuming S remains large and U''<0.

[24]The musicians surveyed come from a wide range of music genres including Pop, Folk, Country, Electronic, Blues, Rock, Jazz, Christian, Punk, Dance, Bluegrass, Latin, Reggae, and Hip Hop. This wide coverage suggests the responses should incorporate a range of viewpoints.

reason that sales have fallen, the new technology does not appear to have exacted a toll on the quantity of music produced.[25] Obviously, it would be nice to adjust output for differences in quality, but we are not aware of any research that has tackled this question.

Similar trends can be seen in other creative industries. For example, the worldwide number of feature films produced each year has increased from 3,807 in 2003 to 4,989 in 2007 (Screen Digest, 2004 and 2008). Countries where film piracy is rampant have typically increased production. This is true in South Korea (80 to 124), India (877 to 1164), and China (140 to 402). During this period, U.S. feature film production has increased from 459 feature films in 2003 to 590 in 2007 (MPAA, 2007).

7. Policy Implications and Conclusions

File-sharing technology considerably weakened copyright protection, first of music and software and increasingly of movies, games, and books. The policy discussion surrounding file sharing has largely focused on the legality of the new technology and the question whether or not declining sales in music are due to file sharing. While these are important questions, in our view, the debate has been overly narrow. Copyright exists to encourage innovation and the creation of new works; in other words to promote social welfare. The question to ask is thus whether the new technology has undermined the incentives to create, market, and distribute entertainment. Sales displacement is a necessary but not a sufficient condition for harm to occur. We also need to know whether income from complementary products offset the decline in income from copyrighted works. And even if income fell, welfare may not suffer if artists do not respond to weaker monetary incentives.

As our survey indicates, the empirical evidence on sales displacement is mixed. While some studies find evidence of a substitution effect, other findings, in particular the

[25] Similarly, recording contracts seem to remain appealing. In 2009, 1,900 acts performed at South-by-Southwest, a large music festival that attracts musicians looking to sign their first recording contract. The artists must typically pay their own travel and lodging expenses, in addition to any foregone wages from their secondary job. Clearly a large number of musicians thought attending the festival was a worthwhile investment (Pareles, 2009).

papers using actual file-sharing data, suggest that piracy and music sales are largely unrelated. In contrast, there is clear evidence that income from complements has risen in recent years. For example, concert sales have increased more than music sales have fallen. Similarly, a fraction of consumer electronics purchases and internet-related expenditures are due to file sharing. Unfortunately, we know little about the distribution of these impacts. How markets for complimentary goods have responded to file sharing remains an area of inquiry that is largely unexplored in academic research.

The same holds true for the question how artists would respond to weaker monetary incentives. Looking at aggregate output – the number of recordings, books, and movies produced every year – we see no evidence that file sharing has discouraged the production of artistic works. However, as with income from complementary goods, aggregate statistics need to be interpreted with some care. For example, digital formats not only encouraged file sharing; digital technology also lowered the cost of producing movies and music and they allowed artists to reach their audience in novel ways. The observed increase in output is in part due to these changes. The response of artists to technology-induced changes in income is a second area that we would like to single out as important for future research.

As this essay has made clear, we do not yet have a full understanding of the mechanisms by which file sharing may have altered the incentives to produce entertainment. However, in the industry with the largest purported impact – music – consumer access to recordings has vastly improved since the advent of file haring. Since 2000, the number of recordings produced has more than doubled. In our view, this makes it difficult to argue that weaker copyright protection has had a negative impact on artists' incentives to be creative.

Literature

Andersen, Birgitte and Marion Frenz (2008). The Impact of Music Downloads and P2P File-Sharing on the Purchase of Music: A Study for Industry Canada. University of London Working Paper.

Albini, Steve (1994). "The Problem with Music." *Maximum Rock 'n' Roll* Issue #133. http://www.arancidamoeba.com/mrr/problemwithmusic.html.

Allot Communications (2007). "Digging Deeper Into Deep Packet Inspection (DPI)." https://www.dpacket.org/articles/digging-deeper-deep-packet-inspection-dpi.

Baker, Matthew J. and Brendan M. Cunningham (2006). Court Decisions and Equity Markets: Estimating the Value of Copyright Protection. *Journal of Law and Economics* 49: 567-596.

Bakos, Yannis, Erik Brynjolfsson, and Douglas Lichtman (1999). Shared Information Goods. Journal of Law and Economics 42: 117-155.

Besen, Stanley M. and Sheila N. Kirby (1989). Private Copying, Appropriability, and Optimal Copyright Royalties. *Journal of Law and Economics* 32: 255-280.

Bhattacharjee, Sudip, Ram D. Gopal, Kaveepan Lertwachara, James R. Marsden and Rahul Telang (2007). The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts. *Management Science* 53(9): 1359-1374.

Boldrin, Michele and David K. Levine (2008). *Against Intellectual Monopoly*. Cambridge: Cambridge University Press.

Caves, Richard (2000). *Creative Industries: Contracts Between Art and Commerce*. Cambridge, MA: Harvard University Press.

Coase, Ronald H. (1979). Payola in Radio and Television Broadcasting. *Journal of Law & Economics* 22 (2): 269-328.

Connolly, Marie and Alan Krueger (2006). Rockonomics: The Economics of Popular Music. Handbook of Arts and Culture, Amsterdam: North Holland: 667-720.

Dhungel, P., D. Wu, B. Schonhorst, K.W. Ross (2008). A Measurement Study of Attacks on BitTorrent Leechers, IPTPS.

Economist (2005). BSA or just BS? 19 May:

Edison Media Research (2003). The National Record Buyers Study III. Sponsored by Radio & Records. http://www.edisonresearch.com.

Ferguson, David (2006). "Trends and Statistics in Peer-to-Peer." CacheLogic. Presented at Workshop on Technical and legal aspects of peer-to-peer television, the Netherlands.

Film Encyclopedia (2008). Historical Background and Definitions. At http://www.filmreference.com/encyclopedia

Forrester (2002). "Downloads Save the Music Business." http://www.forrester.com.

Forrester 2004. "US Antipiracy Bill Won't Stop File Sharing."

Gayer, A. and O. Shy (2006). Publishers, Artists, and Copyright Enforcement. Information Economics & Policy 18(4): 374-384.

Gopal, Ram D. and Sudip Bhattacharjee (2006). Do Artists Benefit from Online Music Sharing? *Journal of Business* 79(3): 1503-1533.

Gummadi, Krishna, Richard Dunn, Stefan Saroiu, Steven Gribble, Henry Levy and John Zahorjan (2003). Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload. Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP-19).

Hui, Kai-Lung and Ivan Png (2002). On the Supply of Creative Works: Evidence from the Movies. *American Economic Review* 92(2): 217-220.

Hui, Kai-Lung and Ivan Png (2003). Piracy and the Legitimate Demand for Recorded Music. *Contributions to Economic Analysis & Policy* 2(1): Article 11.

Hong, Seung-Hyun (2004). The Effect of Napster on Recorded Music Sales: Evidence from the Consumper Expenditure Survey. Working Paper, Stanford University.

Hong, Seung-Hyun (2008). Measuring the Effect of Napster on Recorded Music Sales. Working Paper, University of Illinois.

International Federation of the Phonographic Industry 2002. Recording Industry in Numbers 2001. International Federation of Phonographic Industry.

Internet2 Netflow Statistics (2009). Internet2 NetFlow: Weekly Reports. http://netflow.Internet2.edu/weekly/.

Johnson, William R. (1985). The Economics of Copying. *Journal of Political Economy* 93:

Jupiter Media Metrix (2002). "File Sharing: To Preserve Market Value Look Beyond Easy Scapegoats." http://www.jupiterresearch.com.

Karagiannis, Thomas, Andre Broido, Nevil Brownlee, k claffy, Michalis Faloutsos (2003). File-sharing in the Internet: A characterization of P2P traffic in the backbone. Technical Report, University of California, Riverside.

Karagiannis, Thomas, Andre Broido, Nevil Brownlee, k claffy, Michalis Faloutsos (2004). "Is P2P dying or just hiding?" Presented at Globecom 2004.

Khan, B. Zorina (2001). Do Property Rights Matter? Evidence from U.S. International Copyright Law, 1790-1910. Bowdoin College Working Paper.

Klein, Benjamin, Andres V. Lerner and Kevin M. Murphy (2002). The Economics of Copyright Fair Use in a Networked World. *American Economic Review* 92: 205-208.

Krueger, Alan (2005). The Economics of Real Superstars: The Market for Concerts in the Material World, *Journal of Labor Economics* 23(1): 1-30.

Lamere, Paul (2006). What's On Your iPod? At http://blogs.sun.com/plamere/entry/what_s_on_your_ipod.

Landes, William M. and Richard A. Posner (1989). An Economic Analysis of Copyright Law. *Journal of Legal Studies* 18: 325-363.

Leibowitz, Nathaniel, Aviv Bergman, Roy Ben-Shaul and Aviv Shavit (2002). Are File Swapping Networks Cacheable? Characterizing P2P Traffic. Expand Networks working paper. Presented at the 7th International Workshop on Web Content Caching and Distribution (WCW).

Lerner, Josh and Jean Tirole (2005). The Economics of Technology Sharing: Open Source and Beyond. *Journal of Economic Perspectives* 19(Spring): 99-120.

Liebowitz, Stanley J. (1985). Copying and Indirect Appropriability: Photocopying of Journals. *Journal of Political Economy* 93: 945-957.

Liebowitz, Stanley J. (2008). Testing File Sharing's Impact on Music Album Sales in Cities. Research Note. *Management Science* 54(4): 852-859.

Lenhart, Amanda and Susannah Fox (2000). Downloading Free Music: Internet Music Lovers Don't Think It's Stealing, The Pew Internet & American Life Project's Online Music Report, 28 September 2000, at: http://www.pewinternet.org/PPF/r/23/report_display.asp.

Leung, Tin Cheuk (2008). Should the Music Industry Sue Its Own Customers? Impacts of Music Piracy and Policy Suggestions. University of Minnesota Working Paper.

Liang, J., R. Kumar, and K.W. Ross (2005a). The Kazaa Overlay: A Measurement Study, Computer Networks (Special Issue on Overlays).

Liang, J., R. Kumar, Yongjian Xi, K.W. Ross (2005b). Pollution in P2P File Sharing Systems, Infocom 05, Miami.

Love, Courtney (2000). "Courtney Love does the math." *Salon*. http://dir.salon.com/tech/feature/2000/06/14/love/index.html

Madden, Mary (2004). "Artists, Musicians and the Internet." Pew Internet & American Life Project. http://www.pewinternet.org/Reports/2004/Artists-Musicians-and-the-Internet.aspx.

Manuse, Andrew J. (2003). Music Industry Threats Not Slowing File-Sharing Volume. Boston Herald, June 27.

McBride, Sarah and Ethan Smith (2008), "Music Industry to Abandon Mass Suits," The Wall Street Journal, 19 December 2008, at B1.

Mortimer, Julie Holland and Alan Sorensen (2005). Supply Responses to Digital Distribution: Recorded Music and Live Performances. Harvard University Working Paper.

MPAA (2007). Theatrical Market Statistics. At http://mpaa.org/researchstatistics.asp.

Novos, Ian E. and Michael Waldman (1984). The Effects of Increased Copyright Protection: An Analytic Approach. *Journal of Political Economy* 92: 236-246.

Oberholzer-Gee, Felix (2006). eDonkey – Deciding the Future of File Sharing. Harvard Business School Case 707-482.

Oberholzer-Gee, Felix and Koleman Strumpf 2005. "The Effect of File Sharing on Record Sales: An Empirical Analysis." Working Paper, Harvard Business School and the University of North Carolina at Chapel Hill.

Oberholzer-Gee, Felix and Koleman Strumpf (2007). The Effect of File Sharing on Record Sales: An Empirical Analysis. *Journal of Political Economy* 115(1): 1-42.

Pareles, Jon (2009). "Stoking Careers in Frenzy of South by Southwest," 22 March 2009, New York Times.

Pasick, Adam (2004). "LIVEWIRE — File-sharing network thrives beneath the radar," Yahoo! News at http://www.interesting-people.org/archives/interesting-people/200411/msg00078.html

Passman, Donald (2000). *All You Need to Know About the Music Business*. New York: Simon & Schuster.

Peitz, Martin and Patrick Waelbroeck (2003). Piracy of Digital Products: A Critical Review of the Economics Literature. CESIfo Working Paper 1071.

Peitz, Martin and Patrick Waelbroeck (2004). The Effect of Internet Piracy on Music Sales: Cross-section Evidence. *Review of Economic Research on Copyright Issues* 1(2): 71-79.

Peitz, Martin and Patrick Waelbroeck (2006). Why the Music Industry May Gain From Free Downloading: The Role of Sampling. *International Journal of Industrial Organization* 24(5): 907-913.

Pew Internet Project (2003). Music Downloading, File-Sharing and Copyright. http://www.pewtrusts.com/pubs/.

Png, Ivan and Qiu-hong Wang (2006). Copyright Duration and the Supply of Creative Work: Evidence from the Movies. National University of Singapore Working Paper.

Pouwelse, J.A., P. Garbacki, D.H.J. Epema, H.J. Sips (2005). The Bittorrent P2P File-sharing System: Measurements and Analysis", 4th International Workshop on Peer-to-Peer Systems (IPTPS'05).

Rich, Motoko (2009). Print Books Are Target of Pirates on the Web. *New York Times*, 11 May.

Rob, Rafael and Joel Waldfogel (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(1): 29-62.

Rosen, Sherwin (1981). "The Economics of Superstars." *American Economic Review*. 71: 845-858.

Sandvine (2002-2007). Occasional white papers and press reports on file sharing activity in North America and World-wide. http://www.sandvine.com.

Sandvine (2008a). "2008 Analysis of Traffic Demographics in North-American Broadband Networks." http://www.sandvine.com.

Sandvine (2008b). "2008 Global Broadband Phenomena." http://www.sandvine.com.

Shalunov, Stanislav and Benjamin Teitelbaum (2001). "TCP Use and Performance on Internet2." ACM SIGCOMM Internet Measurement Eorkshop 2001.

Smith, Michael D. and Rahul Telang (2006). Piracy or Promotion? The Impact of the Broadband Internet Penetration on DVD Sales. Carnegie Mellon University Working Paper.

Smith, Michael D. and Rahul Telang (2008). Competing With Free: The Impact of Movie Broadcasts on DVD Sales and Internet Piracy. Carnegie Mellon University Working Paper.

Tanaka, Tatsou (2004). Does File-sharing Reduce CD sales?: A Case of Japan, Conference Paper Prepared for Conference on IT Innovation. Hitotsubashi University, Tokyo.

Towse, Ruth (1999). Copyright and Economic Incentives: An Application to Performer's Rights in the Music Industry. *Kyklos* 52: 69-90.

Varian, Hal (2000). Buying, Sharing and Renting Information Goods. University of California at Berkeley Working Paper.

Zentner, Alejandro (2006). Measuring the Effect of Music Downloads on Music Purchases. *Journal of Law & Economics* 49(1): 63-90.

TABLE 1
KEY EVENTS IN FILE SHARING

Date	Event
Spring 1998	First mass-produced MP3 player
October 1998	RIAA files restraining order against leading MP3 player manufacturer
June 1999	Napster begins operations
December 1999	RIAA sues Napster for copyright damages
July 2000	US District Court rules against Napster and in favor of RIAA. Case moves to US Court of Appeals which affirms in February 2001 that Napster is liable for damages
Spring-Summer 2001	Several alternative file sharing protocols are released including FastTrack/KaZaA, WinMX, Limewire, and BitTorrent
July 2001	Napster effectively shut-down
November 2001	RIAA and MPAA sue file sharing software distributors Morpheus and Grokster in *MGM v. Grokster*
Spring 2003	FastTrack/KaZaA peaks at about 4m simultaneous users.
September 2003	RIAA begins suing file sharing users. About 35,000 lawsuits have been filed by the end of 2008.
November 2003	The Pirate Bay, a BitTorrent index and tracker site, is founded
Fall 2004	A leading BitTorrent tracker + indexer has over 1m visits per day
June 2005	Supreme Court upholds the content-holders position in *MGM v. Grokster*. By the end of the 2005 distribution companies eDonkey and WinMX shut-down after receiving cease and desist letters from the RIAA
May 2006	In part due to pressure from the MPAA, Swedish police shut down The Pirate Bay and confiscate its servers. Site was operational again in three days, and servers are now spread over several countries
November 2008	25m users on leading BitTorrent tracker The Pirate Bay

TABLE 2
FILES ON FILE-SHARING NETWORKS

	% songs on network	% store sales	% downloads
Full sample	100.0%	100.0%	100.0%
Catalogue	8.0%	9.8%	12.6%
Current Alternative	19.1%	24.8%	48.6%
Hard Music Top Overall	3.0%	5.9%	5.3%
Jazz Current	2.9%	4.6%	0.4%
Latin	3.5%	5.8%	0.7%
New artists	8.0%	3.3%	1.8%
R&B	25.2%	9.7%	14.9%
Rap	13.7%	8.2%	4.6%
Top Current Country	10.2%	18.4%	7.3%
Top Soundtrack	6.4%	9.4%	3.9%

Source: Oberholzer-Gee and Strumpf (2007)

TABLE 3
AVAILABILITY OF MOVIES ON MININOVA

RANK	TOP DVD RENTALS MARCH 2009	# DOWNLOADS	MALIN AKERMAN MOVIES	# DOWNLOADS
1	Role Models (2008)	10,482	Watchmen	53,476
2	Transporter 3 (2008)	11,225	Bye Bye Sally	NA
3	Australia (2008)	17,244	27 Dresses	367
4	Milk (2008/I)	2,833	Heavy Petting	0
5	Beverly Hills Chihuahua (2008)	3,050	The Heartbreak Kid	53
6	Rachel Getting Married (2008)	1,705	The Brothers Solomon	0
7	Body of Lies (2008)	10,394	The Invasion	NA
8	In the Electric Mist (2009)	1,885	Harold & Kumar	382
9	Changeling (2008)	11,149	The Utopian Society	NA
10	Nights in Rodanthe (2008)	1,290	The Circle	NA

Sources: Internet Movie Database (http://www.imdb.com/) and Mininova (http://www.mininova.org/), accessed on 14 March 2009

TABLE 4
THE GEORGRAPHY OF FILE SHARING

Country	Share of users	Share of downloads	Users in U.S. download from (%)	Users in U.S. upload to (%)	Share World Population	Share World Internet Users
United States	30.9	35.7	45.1	49.0	4.6	27.4
Germany	13.5	14.1	16.5	8.9	1.3	5.3
Italy	11.1	9.9	6.1	5.7	0.9	3.2
Japan	8.4	2.8	2.5	1.8	2.0	9.3
France	6.9	6.9	3.8	4.7	1.0	2.8
Canada	5.4	6.1	6.9	7.9	0.5	2.8
United Kingdom	4.1	4.0	4.2	4.2	1.0	5.7
Spain	2.5	2.6	1.8	2.0	0.6	1.3
Netherlands	2.1	2.1	1.9	1.6	0.3	1.6
Australia	1.6	1.9	0.8	2.2	0.3	1.8
Sweden	1.5	1.7	1.8	1.5	0.1	1.0
Switzerland	1.4	1.5	0.9	1.0	0.1	0.6
Brazil	1.3	1.4	1.2	1.3	2.9	2.3
Belgium	0.9	1.2	0.5	1.0	0.2	0.6
Austria	0.8	0.6	0.6	0.4	0.1	0.6
Poland	0.5	0.7	0.7	0.5	0.6	1.1

Source: Oberholzer-Gee and Strumpf (2007)

TABLE 5
STUDIES OF THE ECONOMIC IMPACT OF FILE SHARING

Study	Study Question, Data and Sample	Methodology	Key Findings
Music			
Hui and Png (2003)	Do country-level piracy rates explain the decline in music sales? Macro data, 28 countries, 1994-1998	Sales regressions with country fixed effects; uses piracy rates for music cassettes and business computer software as instruments	For every pirated CD, sales fall by 0.42 units. Estimated effect is not robust to including year fixed effects and estimating separate displacement effects for high- and low-income countries.
Peitz and Waelbroeck (2004)	Do country averages in the likelihood of having downloaded music at least once predict music sales? Macro data, 16 countries, 1998-2002	Cross-sectional analysis relating changes in sales to the level of file-sharing in 2002; no measure for the intensity of file sharing	Piracy reduced sales by 20%; effect is significant at 10% level
Tanaka (2004)	Do albums that are popular on file-sharing networks sell fewer copies? Observed piracy; 261 best-selling titles; 2004	Study relates actual downloads on Winny, a popular Japanese file-sharing software, to CD sales; uses music genres as instruments	File-sharing does not reduce sales.
Gopal et al. (2006)	Are students who sample music they don't know more likely to purchase the CD? Survey; 200 students	Students indicate interest in buying and sampling music in a hypothetical-choice setting with set prices.	Students with faster internet connections are more likely to sample music; sampling increases the propensity to buy.
Rob and Waldfogel (2006)	Do students who downloaded music purchase fewer albums? Survey; 412 students; 2003/2004	Students report purchases and downloads of 8,200 specific recordings; study uses access to broadband to instrument for downloads	For hit albums the authors find no relationship between downloading and sales. For a wider set of music, downloading five albums displaces the sale of one CD. Instrumenting for downloads results in estimates that are too imprecise to draw any firm conclusions. Using student valuations of albums, the authors conclude that file-sharing increases social welfare.

Zentner (2006)	Do individuals who downloaded at least once buy fewer CDs? Survey; 15,000 European consumers, 2001	Cross-sectional analysis; uses measures of Internet sophistication and access to broadband as instruments; no measure for the intensity of file sharing	Having shared files reduces the probability of purchasing music by 30%.
Bhattacharjee et al. (2007)	Do albums that are more frequently shared drop off the Billboard charts in a shorter period of time? Observed piracy; best-selling titles; 2002-2003	Relates the supply of files on file-sharing network (WinMx) to chart rankings; study uses RIAA announcement of lawsuits as instrument	Overall, file sharing has no statistically significant effect on survival on charts. The authors find a small negative effect for weaker releases.
Oberholzer-Gee and Strumpf (2007)	Do albums that are popular on file-sharing networks sell fewer copies? Observed piracy; representative sample of recordings; 2002	Relates downloads of files to CD sales; uses the supply shock due to German school holidays to instrument for downloads	File-sharing does not have a statistically significant impact on record sales.
Andersen and Frenz (2008)	Do individuals who obtain music for free buy fewer CDs? Survey; representative sample of Canadians, 2006	Authors have information on many forms of sharing, including P2P, ripping, promotional downloads, and copying of mp3 files; cross-sectional regressions without instruments	File sharing increases music purchases. 12 additional downloads lead to the sale of an additional 0.44 CDs.
Hong (2004, 2008)	Do households with internet access report lower music purchases post Napster? Survey; 2000	Two-variate propensity score matching; probability of using Napster is unobserved; needs to be imputed from UCLA survey using demographic information	The introduction of Napster explains 20% of the decline in music expenditures. 80% of the decline is due to changes in the prices of other entertainment goods and the ending of the transition from LPs to CDs (Hong 2004). Using a conventional difference-in-difference approach, the effect of Napster would be significantly overestimated, explaining the entire decline.
Leung (2008)	Do students who indicate they would download music intend to buy fewer songs? Conjoint survey; 884 (270) students	Students report past consumption of music and make hypothetical choices between legal music, iPods, and pirated music; the study uses an assumed probability of getting caught and the size of the fine as instruments	When students pirate 10% more music, they intend to buy 0.7% fewer iTunes songs and 0.4% fewer CDs.

Liebowitz (2008)	Do U.S. cities with greater internet penetration have lower record sales? Macro data; 89 markets, 1998-2003	Compares changes in city-wide internet penetration with changes in record sales, controlling for demographics	Using all markets, internet penetration is unrelated to changes in music sales; for a subset of markets (60) the internet reduces per-capita-sale by 1.55, indicating file sharing explains more than 100% of the decline in record sales.
Movies and TV			
Smith and Telang (2006)	Does broadband help or hurt DVD sales? Macro data; 2000-2003	Market fixed effects specification with autoregressive errors	Broadband penetration increases DVD sales. Almost 10% of the increase in DVD sales during the study period is attributable to advances in broadband penetration.
Rob and Waldfogel (2007)	Are students who watch a pirated copy of a movie subsequently less likely to purchase the DVD? Survey; 500 students; 2002-2005	Students report their viewing of 50 top movies; no instrumental variables; person fixed effects control for time-invariant unobserved heterogeneity	Illegal burning of DVDs and downloading make up 5.2% of movie viewing; unpaid consumption reduces paid consumption by 3.5%.
Waldfogel (2007)	Do students who watch a TV series on the web less likely to watch episodes on TV? Survey; 287 students; 2005-2007	Students report the consumption of TV series on TV, YouTube and network websites; no instruments; demand for TV is estimated in first differences	Web consumption (authorized and unauthorized) reduces the number of shows that students watch frequently on TV but it increases the number of shows they watch sometimes. Additional web viewing exceeds the reduction in traditional viewing; even network-controlled viewing (excluding YouTube) increases by 1.5 hours per week.
Smith and Telang (2008)	Do TV broadcasts of movies and piracy reduce the sale of DVDs? Observed piracy; 267 movies; 2005-2006	The study uses TV broadcasts as shocks to identify the effect of piracy on DVD sales	Free broadcasts of movies on TV increase DVD sales on Amazon by 118% during the first week after the broadcast. Piracy does not affect this increase in demand.

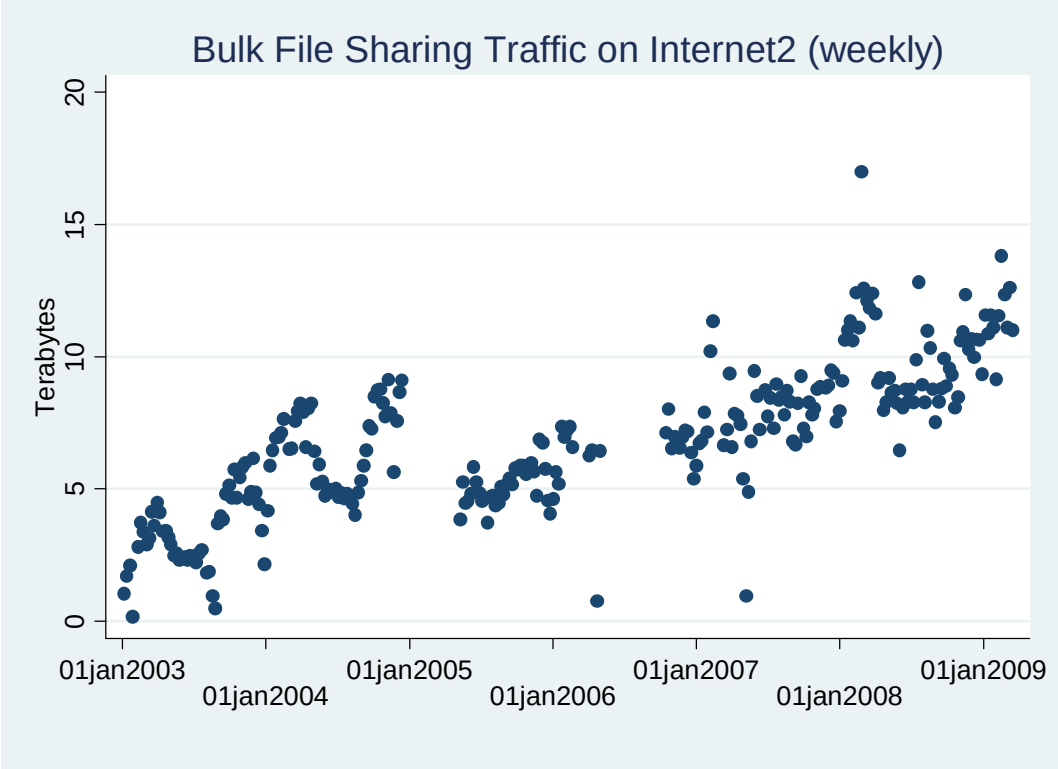
TABLE 6
ARTIST INCOMES (IN MILLIONS USD)

Rank	Artist	Concerts	Recordings	Publishing	Total
1	Paul McCartney	64.9	2.2	2.2	72.1
2	The Rolling Stones	39.6	0.9	2.2	44.0
3	Dave Matthews Band	27.9	0.0	2.5	31.3
4	Celine Dion	22.4	3.1	0.9	31.1
5	Eminem	5.5	10.4	3.8	28.9
6	Cher	26.2	0.5	0.0	26.7
7	Bruce Springsteen	17.9	2.2	4.5	24.8
8	Jay-Z	0.7	12.7	0.7	22.7
9	Ozzy Osbourne	3.8	0.2	0.5	22.5
10	Elton John	20.2	0.9	1.3	22.4
11	The Eagles	15.1	0.7	1.4	17.6
12	Jimmy Buffet	13.7	0.2	0.5	17.6
13	Billy Joel	16.0	0.0	1.0	17.0
14	Neil Diamond	16.5	0.0	0.3	16.8
15	Aerosmith	11.6	1.0	0.8	16.5
16	CSNY	15.7	0.0	0.3	16.0
17	Creed	10.9	1.1	1.6	13.4
18	Rush	13.4	0.0	0.0	13.4
19	Linkin Park	1.7	4.7	6.3	13.1
20	The Who	12.6	0.0	0.0	12.6
21	Red Hot Chili Peppers	6.1	3.4	2.7	12.1
22	Brian “Baby” Williams	0.2	2.7	0.9	11.8
23	Nsync	7.7	0.5	0.9	9.4
24	Barry Manilow	8.0	1.2	0.0	9.2
25	Britney Spears	5.5	1.8	1.0	9.1
26	Alan Jackson	4.6	3.0	1.4	9.0
27	Rod Stewart	6.6	1.4	0.8	8.8
28	Andrea Bocelli	8.1	0.2	0.4	8.7
29	Brooks and Dunn	6.7	0.4	1.4	8.1
30	Enrique Iglesias	4.4	1.5	1.7	7.6
31	Tom Petty	6.6	0.2	0.7	7.5
32	Tool	7.3	0.0	0.0	7.4
33	Kid Rock	3.4	0.8	1.3	7.0
34	Kenny Chesney	5.8	1.1	0.1	7.0
35	Santana	6.0	0.0	0.7	6.9
	Average	12.7	1.7	1.3	17.4

Note: Figures are estimates of pretax gross income in 2002.

Source: Connolly and Krueger (2006).

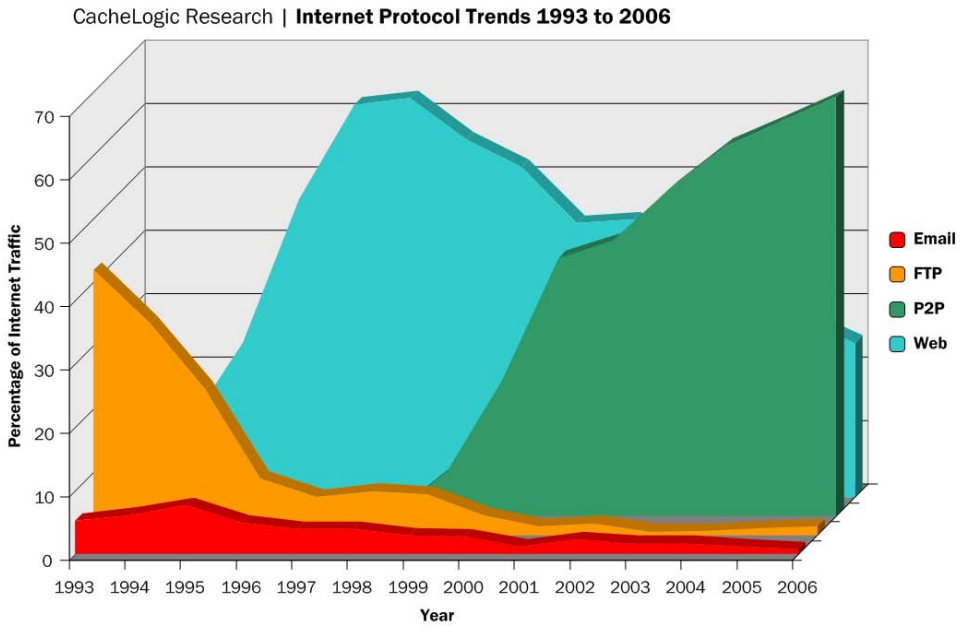
FIGURE 1
TRENDS IN U.S. FILE-SHARING ACTIVITY, 2003-2009



Notes: Bulk traffic is a TCP flow that transferred more than 10MB of data. No date is available for the following weeks: 2/3/03, 7/28/03, 2/23/04, 12/20/04-5/2/05, 7/11/05, 2/27/06-3/27/06, 4/17/06, 5/8/06-10/9/06, 2/19/07-3/5/07, 6/18/07, and 11/19/07.

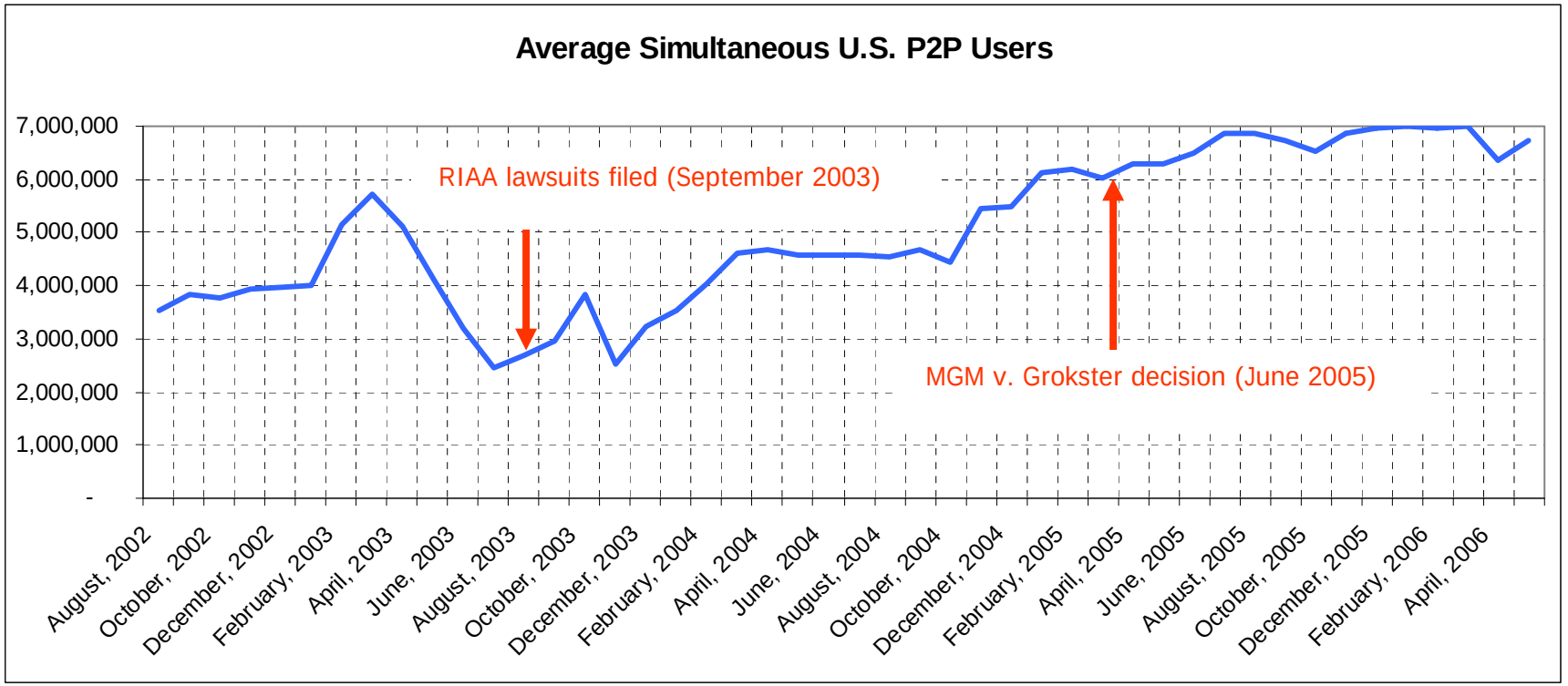
Source: Data from Internet2 Netflow Statistics (2009).

FIGURE 2
GLOBAL FILE SHARING, 1999-2006



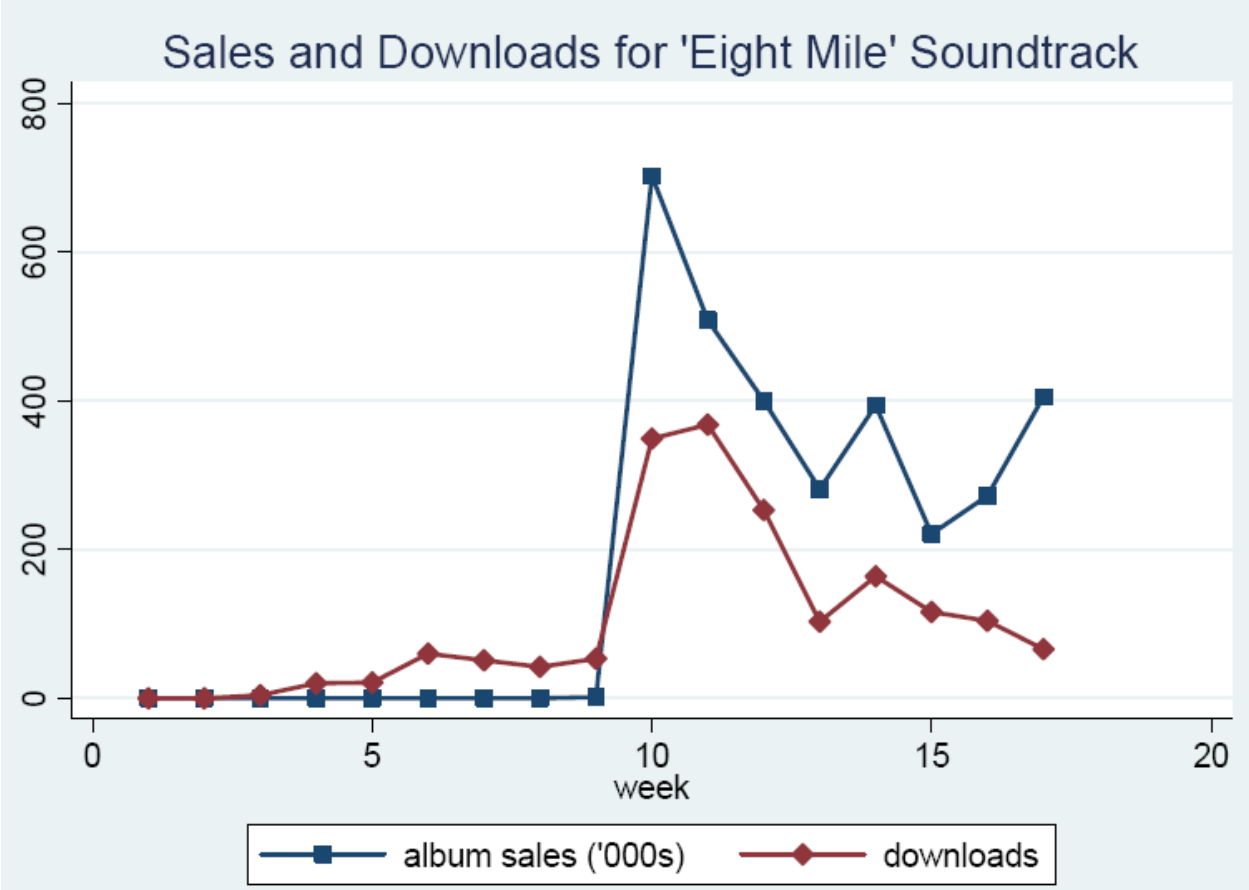
Source: Ferguson (2006)

FIGURE 3
TRENDS IN THE NUMBER OF U.S. FILE-SHARING USERS



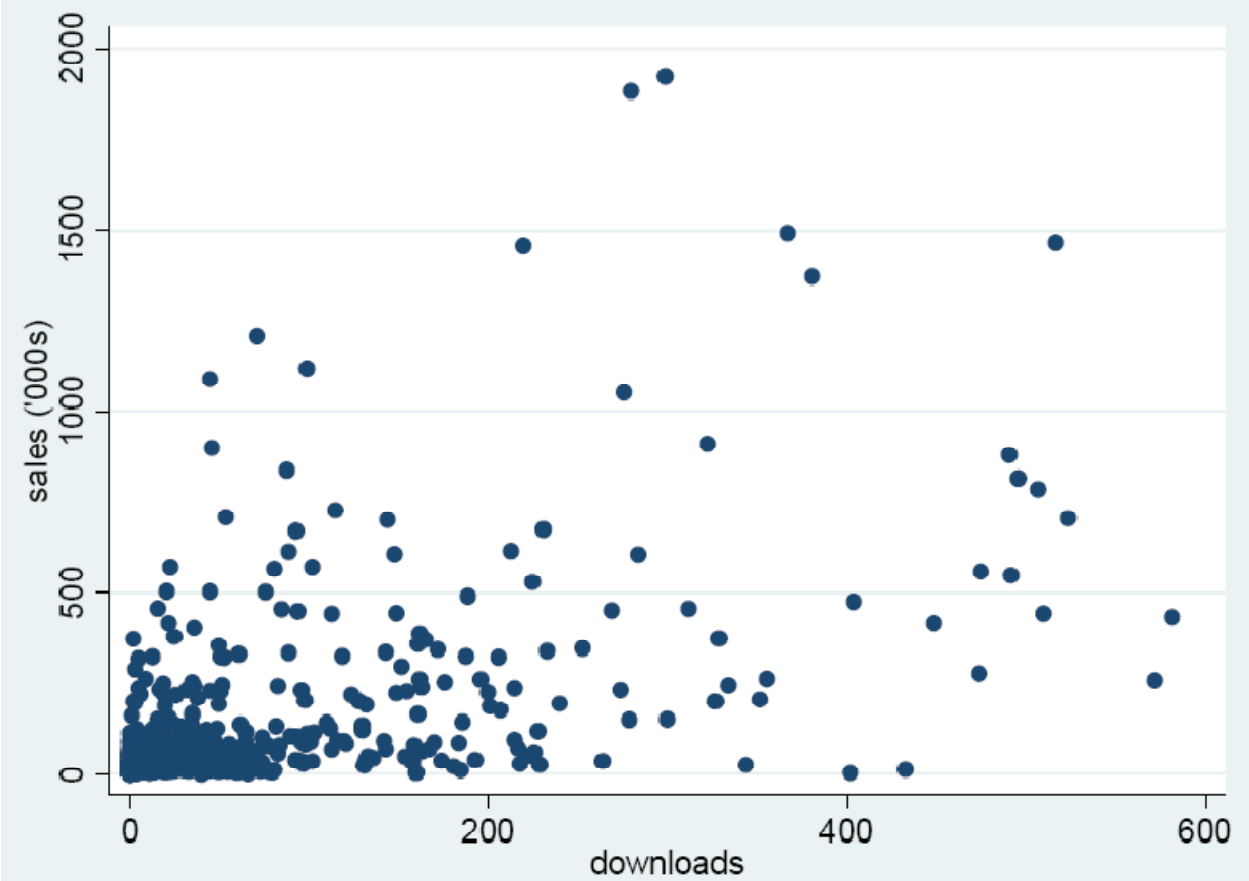
Source: BigChampagne.com

FIGURE 4
INDUSTRY MARKETING AND FILE-SHARING



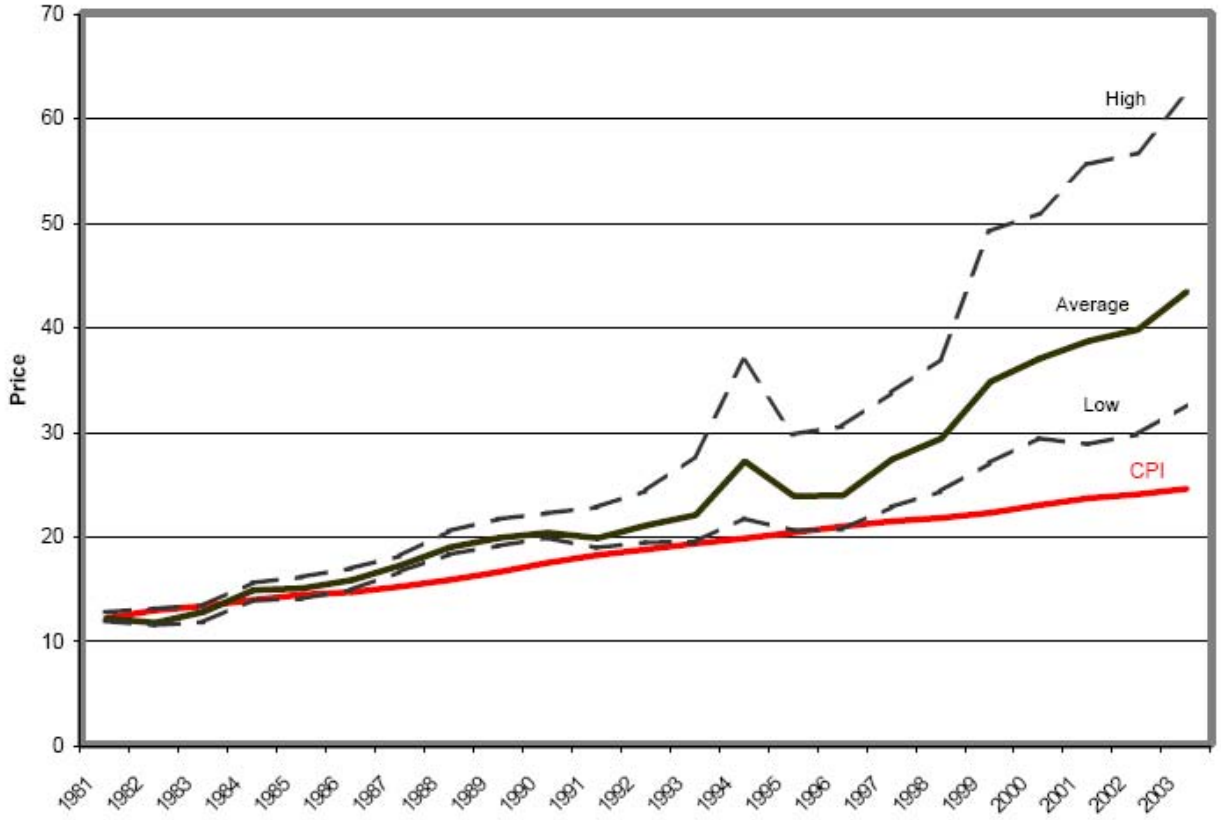
Data from Oberholzer-Gee and Strumpf (2007)

FIGURE 5
ENDOGENEITY OF FILE SHARING



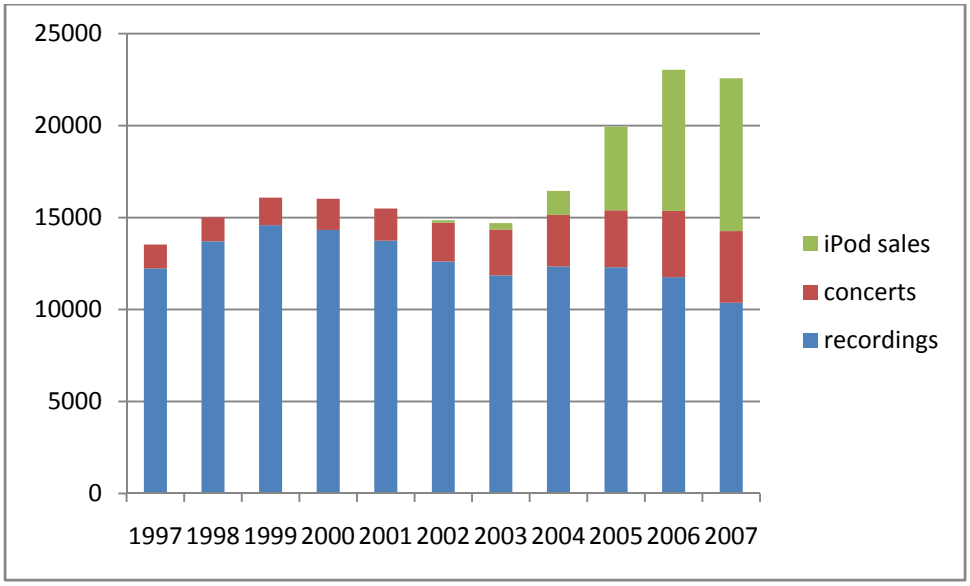
Data from Oberholzer-Gee and Strumpf (2007)

FIGURE 6
CONCERT PRICES 1981-2004



Source: Krueger, 2005

FIGURE 7
U.S. MUSIC INDUSTRY SALES TRENDS



Sources: Recording Industry Association of America, "2007 Year-End Shipment Statistics" (www.riaa.com), Pollstar (www.pollstar.com), Apple, Inc. Annual Reports (www.apple.com), accessed 18 March 2008.

ATTACHMENT B

THE EFFECT OF FILE SHARING ON RECORD SALES: AN EMPIRICAL ANALYSIS

FELIX OBERHOLZER-GEE AND KOLEMAN STRUMPF

The Effect of File Sharing on Record Sales
An Empirical Analysis[*]

Felix Oberholzer-Gee
Harvard University

Koleman Strumpf
University of Kansas

[*] We would like to thank Bharat Anand, Gary Becker, Bob Frank, Shane Greenstein, Austan Goolsbee, Alan Krueger, Steven Levitt, Tom Mroz, Alan Sorensen, Joel Waldfogel, Steven Wildman, Pai-Ling Yin, participants at numerous seminars, and two anonymous referees for helpful comments. This project would not have been possible without the assistance of several individuals and organizations. MixMasterFlame and the FlameNap network shared P2P data with us, and BigChampagne LLC, the CMJ Network, Nathaniel Leibowitz, and Nevil Brownlee generously provided auxiliary data. We thank Keith Ross and David Weekly for assistance in understanding the KaZaA, OpenNap, and WinMX search protocols and database indices. Sarah Woolverton and Christina Hsiung Chen provided superb research assistance. The financial support of the George F. Baker Foundation (Oberholzer-Gee) and the Kenan Faculty Fund (Strumpf) is gratefully acknowledged. We appreciated the aural support from Massive Attack, Sigur Ros and The Mountain Goats.

Abstract
For industries ranging from software to pharmaceuticals and entertainment, there is an intense debate about the appropriate level of protection for intellectual property. The Internet provides a natural crucible to assess the implications of reduced protection because it drastically lowers the cost of copying information. In this paper, we analyze whether file sharing has reduced the legal sales of music. While this question is receiving considerable attention in academia, industry and in Congress, we are the first to study the phenomenon employing data on actual downloads of music files. We match an extensive sample of downloads to U.S. sales data for a large number of albums. To establish causality, we instrument for downloads using data on international school holidays. Downloads have an effect on sales which is statistically indistinguishable from zero. Our estimates are inconsistent with claims that file sharing is the primary reason for the decline in music sales during our study period.

I. Introduction

File sharing is now one of the most common online activities. U.S. households swap more than 300 million files each month, a figure that has grown by over 50% in the last two years (Karagiannis, Broido, Brownlee, claffy and Faloutsos 2004; Billboard 2006). Sharing files is largely non-rivalrous because the original owner retains his copy of a downloaded file. The low cost of sharing and significant network externalities are key reasons for the dramatic growth in file-sharing. While few participated prior to 1999, the founding year of Napster, in 2006 there were about ten million simultaneous users on the major peer-to-peer (P2P) networks (BigChampagne 2006). Because physical distance is largely irrelevant in file sharing, individuals from virtually every country in the world participate.

There is great interest in understanding the economic effects of file sharing, in part because the music industry was quick to blame the phenomenon for the recent decline in sales. Between 2000 and 2005, the number of CDs shipped in the United States fell by 25% to 705 million units (RIAA 2006). Claiming that file sharing was the culprit, the recording industry started suing thousands of individuals who share files. The industry also asked the Supreme Court to rule on the legality of file-sharing services, a question which critically hinges on the "market harm" caused by the new technology. Congress is currently considering a number of measures designed to counter the perceived threat of file sharing.

While concerns about P2P are widespread, the theoretical effect of file sharing on record sales and industry profits is ambiguous (Bakos, Brynjolfsson and Lichtman 1999; Takeyama 1997; Varian 2000). Participants could substitute downloads for legal purchases, thus reducing sales. The inferior sound quality of downloads and the lack of features such as liner notes or cover art

perhaps limit such substitution. Alternatively, file sharing allows users to learn about music they would not otherwise be exposed to. In the file sharing community, it is common practice to browse the files of others and discuss music in file server chat rooms. This learning may promote new sales. Other mechanisms proposed in the theoretical literature have unclear effects on sales. Individuals can use file sharing to sample music, which will increase or decrease sales depending on whether users like what they hear (Shapiro and Varian 1999). The availability of file sharing could also change the willingness to pay for music – it could either decrease it due to the ever present option of downloading, or it could increase it through network effects and the greater ease of sharing (Takeyama 1994). Finally, it is possible there is little effect on sales. File sharing lowers the price of music, which draws in low-valuation individuals who would otherwise not have purchased albums. Rob and Waldfogel (2006) find in a recent survey that college students value albums they purchased in the store at $15.91. In contrast, respondents' willingness to pay for albums they downloaded was only $10.66, a value below the average purchase price of a CD.

With no clear theoretical prediction, the effect of file sharing on sales is an empirical question.[1] Most of what we know about the effects of file sharing is based on surveys. The evidence is mixed. File sharers generally acknowledge both sales displacement and learning effects, and it is unclear if either effect dominates. Rather than relying on surveys, this study is the first to use observations of actual file-sharing behavior of a large population to assess the impact of downloads on sales. Our dataset includes 0.01% of the world's downloads (1.75 million file

[1]The entertainment industry's opposition to file sharing is not a priori evidence that file sharing imposes economic damages. The industry has often blocked new technologies which later become sources of profit. For example, Motion Picture Association of America President Jack Valenti argued that "the VCR is to the American film producer as the Boston strangler is to the woman home alone" (Congressional Hearings on Home Recording, 12 April 1982). By 2004, 72% of domestic industry revenues came from VHS and DVD rentals or sales (DEG 2005; MPAA 2005). Other examples include the record industry's initial opposition to radio in the 1920s and 1930s and to home taping in the 1980s.

transfers) from the last third of 2002, a period of rapid growth in file sharing. We match audio downloads of users in the United States to a representative set of commercially relevant albums for which we have concurrent weekly sales, resulting in a database of over ten thousand album-weeks. This allows us to directly study the relationship between downloads and sales. To establish causality, we instrument for downloads using international school holidays, a supply shock that is plausibly exogenous to sales. Our instruments are relevant since they have a large impact on file transfer time, which in turn is a key determinant of the number of downloads.

We find that file sharing has only had a limited effect on record sales. After instrumenting for downloads, the estimated effect of file sharing on sales is not statistically distinguishable from zero. The economic effect of the point estimates is also small. When considering the policy implications of these results, it is important to take into account the precision of our estimates. Based on all specifications presented in this paper, even our least precise results, we can reject the hypothesis that file sharing cost the industry more than 24.1 million albums annually (3% of sales and less than one third of the observed decline in 2002). Models that consider the dynamics of file sharing allow us to make more precise statements. For example, if we account for the growth in file sharing during our study period we can reject a null that P2P displaced more than 6.6 million in CD sales or less than 10% of the 2002 decline. We arrive at similar conclusions if we allow the effect of international school holidays to vary by album. Our results continue to hold after permitting downloads to influence sales with a lag, omitting data from the holiday shopping season, and restricting our sample to popular titles. In total the estimates indicate that the sales decline over 2000-2002 was not primarily due to file sharing. While downloads occur on a vast scale, most users are likely individuals who in the absence of file sharing would not have bought the music they downloaded.

Our conclusion is supported by other data and methods of analysis. For instance, in the most recent Consumer Expenditure Survey (2004) for the U.S., households without a computer, who seem unlikely to engage in file sharing, report that they reduced their spending on CDs by 43% since 1999. Quasi-experimental evidence on the long-term effect of P2P on music sales also lead to similar results. For example, we document that the share of sales during the summer months when fewer students have access to high-speed campus Internet connections did not change as a result of P2P. Similarly, sales did not decline more precipitously in the Eastern Time Zone of the United States where P2P users can more conveniently download files provided Europeans. Using several years of data, we also show that the number of P2P users is not correlated with album sales. Finally we document that the recording industry often experiences sales reductions, including a recent episode with a sharper reduction than the current period. These experiments are an important complement to our micro-data results. While the main estimates focus on high-frequency variation over several months, the experiments focus on long-term trends using data spanning several years.

Our results have broader implications beyond the specific case of file sharing. A longstanding question in economics concerns the level of protection for intellectual property that is necessary to ensure innovation (Posner 2005). Economic research on the role of patents and copyrights likely began with the critique in Plant (1934) and continues today in the debate between Boldrin and Levine (2002) and Klein, Lerner and Murphy (2002). We provide specific evidence on the impact of weaker property rights for the case of a single industry, recorded music. The file-sharing technology available in 2002 had markedly lowered the protection that copyrighted music recordings enjoyed, so it is interesting to analyze to what extent this reduced protection adversely affected sales. For our study period, we do not detect a significant impact. The paper

also contributes to a growing literature which studies the interactions between the Internet and brick and mortar economies (Goolsbee 2000; Gentzkow forthcoming).

The outline of the paper is as follows. The next section provides an overview of the empirical literature. Section III describes the mechanics of file sharing, and we discuss our data in Section IV. Next we describe the econometric approach. Section VI presents the results, and the last section discusses the implications of this study.

II. The Literature

Empirical research on file sharing and record sales has been limited and inconclusive, primarily, we believe, due to shortcomings with the data. Most of what we know about the effect of file sharing on sales is based on surveys. There are numerous industry studies which arrive at a diverse range of conclusions. For instance, Forrester Research (2002) and Jupiter Media Metrix (2002) find neutral or positive effects, while the International Federation of the Phonographic Industry (2002), Edison Media Research (2003) and Forrester Research (2004) document a sales displacement. A general difficulty with these studies is that they compare the purchases of individuals who download files with the purchases of those who do not. While downloaders may in fact buy fewer records, this could simply reflect a selection effect. File sharing is attractive to those who are time-rich but cash-poor, and these individuals would purchase fewer CDs even in the absence of P2P networks.

A handful of academic studies rely on micro data to address the issue of unobserved heterogeneity among file sharers.[2] Rob and Waldfogel (2006) study the survey responses of a convenience sample of U.S. college students. For hit albums which sold more than 2 million

[2] The Journal of Law and Economics published additional papers in a symposium on file sharing in 2006. Oberholzer-Gee and Strumpf (2005) discusses these studies and additional work.

copies since 1999, they find no relationship between downloading and sales. Expanding the set of albums to include all music the students acquired in 2003, downloading five albums displaces the sale of one CD. These results could mean that piracy does not affect hit albums but hurts smaller artists, or it is also possible that file sharing had less of an effect on sales in earlier years. After instrumenting for downloads with the school the students attend – everyone at Penn has broadband access while this is not true for the other schools – the resulting estimates are too imprecise to draw any firm conclusions. Zentner (2006) employs European survey data to study the relation between file sharing and sales. Using measures of Internet sophistication and access to broadband as instruments, Zentner finds some displacement. Unfortunately, neither the Rob and Waldfogel study nor Zentner's work allow inferences about the total impact of file sharing on record sales because neither paper studies a representative sample of file sharers. Zentner also lacks information about the number of downloads and CD purchases.

Our approach differs from the current literature in that we directly observe file sharing. Our results are based on a large and representative sample of downloads, and individuals are generally unaware that their actions are being recorded.

III. File sharing Networks

File sharing relies on computers forming networks which allow the transfer of data. Each computer may agree to share some files and has the ability to search for and download files from other computers in the network. Our data come from the OpenNap network, an open-source descendant of Napster. OpenNap is an example of a centralized P2P network in which users log on to a central server that tracks all search requests and file downloads. During our study period in the fall of 2002, P2P networks were already quite large. FastTrack (which includes the

popular KaZaA service (see Liang, Kumar and Ross 2004) had grown to 3.5 million simultaneous users by December 2002. The second largest network was WinMX, which had about 1.5 million simultaneous users in 2002. Even the smaller networks were fairly large. OpenNap, the choice of about one percent of all P2P users, had at least 25,000 simultaneous users sharing over 10 million files. Napster no longer operated in the fall of 2002.

IV. Data

We use two main data sources for this study. Logs for two OpenNap servers allow us to observe what files users download. Weekly album-level sales data come from Nielsen SoundScan (2005). SoundScan tracks music purchases at over 14,000 retail, mass merchant and online stores in the United States. Nielsen SoundScan data are the source for the well-known Billboard music charts. To develop our instruments, we rely on a large number of additional data sources which we discuss in the next section.

File Sharing Data

Our data were collected from two OpenNap servers, which operated continuously for seventeen weeks from 8 September to 31 December 2002. The information on file transfers is collected as part of the log files which the servers generate, and most users are unaware their actions are being observed and recorded. An excerpt of a typical log file is:

```
[2:53:35 PM]: User evnormski "(XNap 2.2-pre3, 80.225.XX.XX)" logged in
[2:55:31 PM]: Search: evnormski "(XNap 2.2-pre3)": FILENAME CONTAINS "kid rock devil"
   MAX_RESULTS 200 BITRATE "EQUAL TO" "192" SIZE "EQUAL TO" "4600602" "(3 results)"
[3:02:15 PM]: Transfer: "C:\Program Files\KaZaA\My Shared Folder\Kid Rock –Devil
   Without A Cause.mp3" (evnormski from bobo-joe)
```

The last two lines in the log file show user "evnormski" downloading the song "Devil Without a Cause" by Kid Rock from user "bobo-joe". Information on downloads are the building blocks of our analysis. We focus on downloads because these are the files users actually obtain and they can potentially displace sales. Over the sample period we observe 1.75 million file downloads, or about 0.01% of all downloads in the world. We restrict the analysis to audio files by users in the U.S. The server logs include the I.P. address for each client which we use to identify our users' home country.

An important question is whether our sample is representative of data on all P2P networks.[3] While we are unaware of any database spanning the universe of music downloads, we were able to compare the data from our servers with a sample of more than 25,000 downloads from FastTrack/KaZaA, the leading network at the time. We find that the availability of titles is highly correlated on the two networks. Using a standard homogeneity test based on 1,789 unique songs, we cannot reject a null that the two download samples are drawn from the same population (Pearson χ^2 statistic is 1824.1). The resemblance of files is not surprising. Individuals in our data are similar to those on the most popular networks because the user experience is quite similar and many individuals employ software which allows them to simultaneously participate on several networks. For example, roughly one third of OpenNap participants uses the WinMX software, which allows them to simultaneously access the two largest networks during our study period. We also find that users on these larger networks and those on our servers have access to a comparable number of files and that network size has little effect on the *distribution* of downloads. Based on these tests, we conclude that our sample is representative of the file transfers on the major P2P networks during our study period.

[3] A more comprehensive discussion of this point is in Appendix A of Oberholzer-Gee and Strumpf (2005).

Sales Data and Album Sample

In this study, we focus on a *sample* of albums sold in U.S. stores in the second half of 2002. The sample is representative of all commercially relevant albums, allowing us to draw meaningful inferences about P2P's impact on overall music sales.[4] The sample is drawn from a *population* of albums on 11 charts produced by Nielsen SoundScan (2005): Alternative Albums (a chart with 50 positions), Hard Music Top Overall (100), Jazz Current (100), Latin Overall (50), R&B Current Albums (200), Rap Current Albums (100), Top Country Albums (75), Top Soundtracks (100), Top Current (200), New Artists (150), and Catalogue Albums (200). The charts are published on a weekly basis, and we include an album in the population if it appears on any chart in any week during the second half of 2002. The original population is extensive (2,282 albums) and includes many poorer-selling albums. For instance, our data include two albums which sold fewer than 100 copies during our study period, and the 25th percentile of sales in our data is only 12,493 copies.[5] While we study the commercially most relevant music, it would be incorrect to think of our population as a set of superstar albums. From this population, we draw a genre-based, stratified random sample of 680 releases. To reflect the popularity of different music styles, we set the sample share of a genre equal to its fraction of CD sales in 2002.[6] Within each genre, we randomly select individual titles.

The average album in the resulting sample sold 143,096 copies during our study period. Table 1 reports sales statistics for the full sample and for individual categories. Across all categories,

[4]The genre charts we sample from made up 81.8% of all CD sales in the United States in the last third of 2002. This is virtually identical to the 2002 share of 83.6% for the Big Five record companies, and 97% of the albums on the annual version of these charts were released on RIAA-associated labels.

[5]A typical measure of album success is gold certification which occurs at sales of half a million copies.

[6]Albums can appear on more than one chart because some charts (e.g., New Artists, Top Current) comprise many musical styles. For sampling purposes, we grouped all albums by style; a Rap album on the Top Current list is grouped with all other Rap albums during the sampling process. In the descriptive statistics, we classify albums by their original charts.

44% of population sales are represented in the sample. A two-sample Kolmogorov-Smirnov test comparing the distribution of sales on the original charts and in our sample is unable to reject the null that sample sales are representative of the population of all albums (p=0.991). We also reject this null comparing each of our 11 original charts with the sample sales for that particular chart (p>0.539 for all 11 charts.)

In order to compare sales and downloads, we match the 260,889 songs which U.S. users successfully transferred during our study period to the 10,271 songs on the 680 albums in our sample. The matching procedure is hierarchical in that we first parse each transfer line, identifying text strings that could be artist names. These text strings are then compared to the artist names in our set of albums. The list of artists contains the name on the cover and up to two other performing artists or producers that are associated with a particular song. For example, the song "Dog" on the B2K album "Pandemonium" is performed by Jhene featuring the rapping of Lil Fizz. For "Dog," B2K, Jhene and Lil Fizz are recognized as artists. Once an artist is identified, the program then matches strings of text to the set of songs associated with that particular artist. Using this algorithm, we match 47,709 downloads in the server log files to our list of songs, a matching rate of about 18%.

There are two reasons why this rate is less than 100%. First, a download may be for a song that is not in our sample. These transfers are not of any concern, they simply reflect the fact that we are working with a sample. A second reason for a match rate of less than 100% could be that our matching algorithm fails to recognize songs. To investigate this possibility, we hand-checked a file with 2,000 randomly chosen unmatched transfers, comparing these downloads against our sample. Only five of the unmatched songs were in our sample. As a result, we believe that the 18% match rate mostly reflects transfers of songs that are not in our sample.

Descriptive Statistics

As this is one of the few data sets that allow us to directly observe P2P users, we describe our data in some detail. A first stylized fact is that file sharing is truly global in nature. While over ninety percent of users are in developed countries, a total of 150 countries are represented in the data. U.S. users make up 31% of the sample. Table 2 shows the top countries for users and downloads. As the data indicate, there is only a loose correlation between user share and other country covariates such as Internet use or the software piracy rate. Column 3 in Table 2 confirms that interactions among file sharers transcend geography and language. U.S. users download only 45.1% of their files from other U.S. users, with the remainder coming from a diverse range of countries including Germany (16.5%), Canada (6.9%) and Italy (6.1%).

While file sharing activities are dispersed geographically, only a limited number of songs are transferred with any frequency. Table 3 shows the average song is downloaded 4.6 times over the study period, but the median number of downloads is zero.[7] Although our sample is representative of all commercially relevant music in the second half of 2002, it is striking to see that more than 60% of the songs in our sample are never downloaded. Aggregated up to the album level, users made 70 downloads from the average album in our sample. The most popular album among file sharers (and the second-best seller) has 1799 downloads, while the median number of downloads per album is 16, the 75th percentile is 63, the 90th percentile is 195, and the 95th percentile is 328. Both downloads and sales closely follow a power-law (pareto) distribution.

File sharing is limited to a select number of songs and most of these songs come from just a few charts. Table 3 shows that songs on the Top Current chart ("Billboard 200") are most frequently

[7]The 75th percentile of downloads per song is 2, the 90th percentile is 11, and the 95th percentile is 22.

downloaded. Downloads from this chart alone make up 48% of all file transfers. Another 25% come from the "Alternative" category. The remaining 9 charts are not particularly popular among file sharers. In view of the low cost of sharing and sampling music on P2P, one could expect users to seek out a great variety of songs representing many musical styles. But this is not the case. P2P downloads closely resemble the play lists of Top 40 radio stations. As a result, it is not surprising that songs from higher-selling albums are downloaded more frequently (Table 4). In the top quartile of sales, albums average 200 downloads. In the bottom category, the mean number of downloads is only 11. This suggests that common factors drive downloads and sales, which is a key concern for the development of our empirical strategy.

V. Empirical Strategy

Econometrics

Our goal is to measure the effect of file sharing on sales. We observe sales and downloads at the album-week level for seventeen weeks. These panel data allow us to estimate a model with album fixed effects,

$$S_{it} = X_{it}\beta + \gamma D_{it} + \omega_s t^s + \nu_i + \mu_{it}. \qquad (1)$$

i indicates the album, t denotes time in weeks, S_{it} is observed sales, X_{it} is a vector of time-varying album characteristics that includes a measure of the title's popularity in the U.S., D_{it} is the number of downloads for all songs on an album, and ω_s controls for time trends (a flexible polynomial or week fixed effects). The key concern in our empirical work is that the number of downloads is likely to be correlated with unobserved album-level heterogeneity. As the

descriptive statistics suggest, the popularity of an album is likely to drive both file sharing and sales, implying the parameter of interest γ will be estimated with a positive bias. The album fixed effects v_i control for some aspects of popularity, but only imperfectly so because the popularity of many releases in our sample changes quite dramatically during the study period.

We address this issue by instrumenting for D_{it} in a 2SLS model. Valid instruments Z_{it} predict file sharing but are uncorrelated with the second-stage error μ_{it}. As in the differentiated products literature, where the problem is correlation between prices and unobserved product quality, we use cost shifters to break the link between unobserved popularity, downloads and sales. An advantage of our instruments, which we discuss below, is that they do not rely on the common but potentially problematic assumption that product characteristics are exogenous (Nevo 2001).[8]

Instruments

Our most important instrument is the number of German secondary school kids who are on vacation in a given week. German users provide about one out of every six U.S. downloads, making Germany the most important foreign supplier of songs.[9] German school vacations produce an increase in the supply of files and make it easier for U.S. users to download music.[10] During holidays German teens can spend more time trading music online, since they do most of their file sharing at home (Niesyto 2002). School vacations also allow the German kids to stay up later, which means they can engage in file sharing during the peak U.S. trading hours (early evening, EST). Supporting this intuition, we find that the number of German kids on vacation is

[8]Appendix B of Oberholzer-Gee and Strumpf (2005) presents a formal model of purchase and download behavior which is the foundation for our econometric approach. In particular it shows why we can use linear demand equations rather than the more complicated transformations which are typical in this literature (Berry 1994; Bresnahan, Stern and Trajtenberg 1997).

[9]The important role of German file sharing users is documented in the authoritative BigChampagne database (OECD 2004). Oberholzer-Gee and Strumpf (2005) provides intuition on why this connection is so strong.

[10]Appendix C of Oberholzer-Gee and Strumpf (2005) shows German users are always net suppliers to file sharing networks, and this effect is accentuated during weeks when many kids are on vacation.

a significant predictor of the number of files uploaded from Germany to the United States (p=0.011). The effect is particularly large for music genres that are popular in Germany.

For German vacations to be a valid instrument, they must not be directly related to U.S. music demand. This seems likely because the vacation variable varies over time for reasons that are specific to Germany. The sixteen German Bundesländer (states) start their academic year at different points in time to smooth the demand for the German tourism industry and avoid traffic jams (Kultusministerkonferenz 2002). For example, Bavarian students were still on summer vacation during the first week of our study period while Rheinland-Pfälzer kids were already back in school (see Figure 1). A second difference to a typical U.S. vacation schedule is that many, but not all Bundesländer grant their students one or two weeks of fall vacation. In Rheinland-Pfalz, this happened in weeks 4 and 5. Bavaria, in contrast, did not schedule a longer fall recess. These länder-specific holidays move from year to year. A Bundesland with early summer vacations in one year is given a later slot in the following year (Agentur Lindner 2004). As we explain in greater detail below, there are additional reasons to believe this variable is exogenous. If file sharing were eliminated tomorrow, German school holidays would have no relation to U.S. record sales.

We create three additional instruments by interacting the German-kids-on-vacation variable with album-specific characteristics. These instruments are particularly useful because they vary across both time and albums and provide identification even if a full set of week and album fixed effects is included.

German-kids-on-vacation × band is on tour in Germany: Tours spur local interest and sales of an album, and they are likely to create a positive supply shock of downloadable files. This instrument is not directly related to U.S. sales because the promotional effect of tours will not

spill across the Atlantic and because the timing of fall and winter concerts in Germany typically reflects idiosyncratic features like venue availability and weather. We expect the effect of German vacations to be even larger if an artist happens to be on tour in Germany that week.

German-kids-on-vacation × indicator for misspellings in song titles: To download a song, a user's search query must match a shared file. At the time of our study, file sharing programs were rather rigid in determining matches.[11] Unless both the searcher and sharer agree on the naming convention, no match will occur. This two-sided search problem suggests that songs with unconventionally spelled titles may be more difficult to find. We use MS Word's spell checker to determine if an album has any song titles with an unconventional spelling. We expect misspellings to reduce the size of the positive supply shock coming from German vacations.

German-kids-on-vacation × rank of album on German charts: Songs from popular albums in Germany are easier to download because the supply of these files is larger. Our measure for German popularity is the rank of the album on the weekly German Top 100 chart (Musikmarkt 2002). Obviously, there is a concern that these chart positions might also measure U.S. popularity. However, the instrument is included along with album fixed effects, so it is the timing of the chart rankings in Germany that identifies downloads. There are important differences in the dynamics of song popularity in the two countries due to taste differences and differences in album release dates.

For all our instruments, we provide additional evidence for their exogeneity in the following sections. Summary statistics for the instruments are in Table 5. Each measure exhibits noticeable variation.

[11]For example, "lose yourself," the name of a popular song, would typically return over a thousand results, but mistyping even one character (such as "lose yourse;f") or omitting part of a word ("lose yours") returned zero results.

Mechanisms Underlying the Main Instruments

Our analysis presumes that each instrument influences download costs, and that these costs impact the number of file transfers. We test this idea by analyzing more detailed server log files which allow us to calculate the download time and success rate of download attempts. We construct five measures of download costs: the time between a download request and the successful initiation of the download (C_1), the time between a search request and a download request (C_2), the time between the initiation of the download and its successful completion (C_3), the ratio of search requests to the number of successful downloads (C_4), and the percentage of failed or canceled download requests (C_5). Each C_i term captures aspects of delay or frustration which a U.S. downloader might experience. The measures are aggregated up to the album-week. For example, C_1 is the average time until download initiation among all observed requests for that album in a particular week.

Mean C_i values are presented in the last row of Table 6. The first three columns show that the typical file takes twenty minutes to download, starting from the initial search until the transfer is complete.[12] There are also long delays for top-selling albums, suggesting there is an ubiquitous scarcity of supply. While slow download speeds are the norm in our data, the estimates in Table 6 show that searching and downloading audio files in the U.S. is considerably easier when a larger number of German school children are on vacation. This reduction is even larger when the artist is on tour and when the album is highly ranked on the German charts.[13] The misspellings interaction significantly increases the time between a search and a download request as well as the number of unfulfilled downloads (C_2, C_4, C_5), but it has little effect on the time it

[12]Gummadi, Dunn, Saroiu, Gribble, Levy and Zahorjan (2003) independently document these long download times. This likely reflects the fact that only a third of the U.S. users in our data had a broadband connection.

[13]Note that the German tour and singles chart variable parameters are identified using only within album variation since fixed effects are included. This mitigates concerns that album popularity in the U.S. is driving the parameter estimates.

takes to transfer a file (C_1, C_3). This is consistent with the argument that misspellings create confusion, though they do not slow down the file transfer itself. The estimated effects on download times are economically significant. For example, a one standard deviation increase in the German vacation variable implies a 1.25 minute reduction in the time for a download to begin (C_1), which is an eighth of the typical delay..

These results are meaningful only if the cost of downloading influences the number of file transfers. This is not obviously true because P2P users can engage in other activities while files are being downloaded, which could mean they are insensitive to the time cost of file sharing. To check if the variation in download time that is due to our instruments has a significant impact on the number of transfers, we estimate the system

$$\begin{aligned} C_{it} &= Z_{it}\delta + v_i + \mu_{it} \\ D_{it} &= C_{it} + v_i + \varepsilon_{it} \end{aligned}, \qquad (2)$$

where Z_{it} is the full list of instruments and C_{it} denotes total download time (C_1+C_2+C_3). The last two columns of Table 6 shows that P2P users are fairly sensitive to the time cost of file sharing: a one standard deviation increase in download time reduces downloads by almost half of their mean. We find similar effects when we separately estimate equation (2) for each of the five C_i terms. These estimates confirm our initial claims. German vacations influence the cost of downloading, and this effect has an important impact on the number of downloads in the U.S.[14]

Specific Concerns with Individual Instruments[15]

[14] A different approach to show that German vacations influence downloading activity is to look at international data. We find that school holidays have an important effect only in countries whose time zones are complementary to Germany's. Appendix C of Oberholzer-Gee and Strumpf (2005) presents this point in detail.

[15] A general concern is that the instruments are based on high frequency variation in download costs. Unfavorable conditions might lead users to simply defer downloads to a later time, in which case our second stage estimates will be attenuated to zero. Oberholzer-Gee and Strumpf (2005) shows this concern is not warranted, since users are impatient and quickly lose interest in an album.

German-kids-on-vacation: A potential difficulty with the vacation variable is that it might be correlated with time-varying album popularity in the U.S. We perform a number of tests to see if this is the case. First, we check if German vacations happen to coincide with official U.S. holidays. We find that there is little overlap.[16] A second possibility is that German school vacations proxy for American vacations which are likely to have a direct impact on music sales. As there is no centralized data on holidays for all 14,000 U.S. school districts, we collect information on the number college students who are out of school during our study period. The sample includes all schools in the top two tiers of U.S. News and World Report's 2002 ranking. Information on school breaks is available for 157 schools, leaving us with data for 2.17 million students, almost a quarter of all U.S. college students. Figure 1 compares the vacation patterns in Germany and the U.S. There are marked differences. When some German kids are off in early fall, U.S. students are mostly in school. During the Thanksgiving break in the U.S., German kids are in school. Both populations are off during the Christmas break, although the break starts earlier for U.S. students. To test more formally if the number of German kids on vacation proxies for the number of U.S. kids, we include the latter in the first stage of equation (1). We find no evidence that the measured effect of German vacations on American music downloads is mediated by U.S. vacations.[17]

In a final test, we check more directly if the German vacation variable is in fact uncorrelated with U.S. demand for music albums. We do this by interacting the instrument with an album's rank on the U.S. MTV charts.[18] MTV rankings have the advantage that videos are often shown prior

[16] Estimates over our 17 week observation period yield: $\text{US Holidays}_t = 1.148\ (1.61) - 0.182\ (0.16) \times \text{German Kids}$, where US Holidays is the number of official American holidays (such as Columbus Day or Thanksgiving) in week t and German Kids is the German holiday instrument.

[17] Controlling for the entire set of instruments, the estimated effect of German vacations on downloads changes from 0.667 (0.054) without the U.S. students-on-break variable to 0.643 (0.057) with this variable.

[18] We thank one of our referees for this suggestion. We also used the Billboard Airplay ranking to explore these effects, with similar results.

to the release of a CD, at a time when songs from a forthcoming album first appear on file-sharing networks. This interaction is included in both stages of equation (1).

$$\begin{aligned} D_{it} &= X_{it}\beta + Z_{it}\delta + \varphi_1 Gkids_t \times MTV_{it} + \omega_{1s} t^s + \nu_i + \varepsilon_{it} \\ S_{it} &= X_{it}\beta + \gamma \hat{D}_{it} + \varphi_2 Gkids_t \times MTV_{it} + \omega_{2s} t^s + \nu_i + \mu_{it} \end{aligned}, \qquad (3)$$

where Z_{it} is our full set of instruments. As required under our assumptions, φ_1 is positive: German vacations have a larger effect for files that are more popular in the U.S. In the second stage, however, φ_2 is economically small and statistically insignificant. When an album becomes more popular in the U.S., this boost in popularity is not directly related to German vacations, supporting our claim that the holiday shocks are exogenous.

A second concern is that Germans supply only a narrow slice of music that is of interest to U.S. file sharers. If those who like the type of music that Germans make available substitute downloads for purchases in an atypical fashion, we measure a local average treatment effect, not a true population effect (Imbens and Angrist 1994). Fortunately, there is substantial overlap between American and German musical tastes. Of the albums that entered our sample via the Billboard 200, 62.65% are also on the top 100 German charts. More generally, we study Amazon rankings to compare sales ranks in the two countries (Goolsbee and Chevalier 2003). With the exception of Latin and Country music, Wilcoxon matched-pairs signed-ranks tests cannot reject the null of equal distributions for the eleven genres in our sample. In the robustness section of the paper, we test if the undersupply of Latin and Country music affects our estimates. We show that this is not the case, suggesting the measured effect of downloads on sales is likely to be a good estimate of the average population effect.

German-kids-on-vacation × indicator for misspellings in song titles: Because misspellings appear to be more likely in some genres than in others, one might argue that this indicator is

likely to proxy for album popularity. In our application, this concern is not valid for two reasons. First, as an empirical matter, we find that misspellings are not correlated with sales, even in models without album or genre fixed effects.[19] Second, all our specifications presented in the results section include album fixed effects which control for an album's time-invariant popularity.

A second difficulty with the misspelling instrument could be that misspellings cause our song matching algorithm to fail. This would result in a negative relationship between misspellings and measured downloads, even if misspellings had no effect on actual downloads. More importantly, the second-stage estimates would be attenuated towards zero, since the variation in fitted downloads would be largely due to noise. Several pieces of evidence suggest this is not true. First, the estimates in the last sub-section show that misspellings do in fact have real effects on transfer times and user behavior. Second, we can check for misspellings in unmatched downloads. If the criticism is correct, there should be more misspellings in the unmatched than in the matched sample. This is not the case.[20]

German-kids-on-vacation × rank of album on German charts: The idea underlying this instrument is that vacation periods in Germany will boost downloads in the U.S. more when many German users make a particular file available. Because the instrument is included along with album fixed effects, it is the timing of the chart rankings in Germany that identify downloads. However, if U.S. popularity shocks happen to coincide with high German chart positions, we would measure the effect of downloads on sales with a positive bias. We can test for this spurious correlation in two ways. First, assuming that the German vacation variable is a

[19] The effect of misspellings on sales is statistically insignificant and economically small. A one-standard-deviation increase in misspellings raises sales by a mere 11,000 copies (less than ten percent of the mean) during our entire study period.

[20] The rates are 0.041 (N=35614) and 0.038 (N=7163), in the unmatched and matched samples respectively. The Pearson $\chi 2$ statistic is 1.402.

valid instrument, we can perform overidentification tests for this and the other interactions that we use as instruments. These tests, reported in the results section of the paper, provide no indication that any of our instruments are invalid. A second and more direct test is to see whether shocks in U.S. demand are correlated with German popularity.[21] Under our hypotheses, U.S. demand shocks must not get magnified when albums become more popular in Germany. For example, we expect U.S. vacations to increase P2P activity, but this increase must not vary with German popularity. The model is,

$$D_{it} = Z_{it}\delta + \varphi_1 Ukids_t + \varphi_2 Ukids_t \times Gcharts_{it} + \varphi_3 Ukids_t \times MTV_{it} + \varphi_4 Gkids_t \times MTV_{it} + \omega_s t^s + \nu_i + \varepsilon_{it} \quad (4)$$

$Ukids_t$ denotes the number of U.S. college students on break (our measure of U.S. demand shocks), $Gcharts_{it}$ is a title's rank on the German charts, and MTV_{it} is the position on the MTV chart (our measure of U.S. popularity). The effect of interest in this specification, φ_2, shows whether a shock in demand in the U.S. is mediated by German popularity. This is not the case: φ_2 is -0.0008 with a standard error of 0.0134, and this effect is only one tenth of the size of the German kids × German chart interaction in our later specifications. The data show that relative popularity in Germany interacts with German but not with U.S. vacations.

VI. Results

Before turning to the estimates, it is instructive to graph some of the data.. Figure 2 shows the weekly time series of sales and purchases for one of the most popular albums in our sample. This "Superstar" album was largely ignored in file sharing networks until it became available for sale in week ten of our sample. This suggests it is the publicity associated with an official

[21] We thank one of our referees for this suggestion.

release which drives downloads as well as sales. Notice also the rapid but non-monotone decay in sales and downloads, which highlights the importance of using high-frequency data.

Panel Analysis

In Table 7 we report results for equation (1). The unit of observation is the album-week. The models include a control in both stages for time-varying U.S. popularity, the album's position on the American MTV charts, and a polynomial time trend of degree six. As expected, a simple OLS specification yields a large positive effect of 1.093 with a standard error of 0.023. A model which adds album fixed effects is given in column (1). While we continue to find a positive effect of downloads on sales, the relationship is now much weaker. The remaining estimates in Table 7 instrument for downloads. We begin by using the number of German kids on school vacation (column II). The first-stage estimates imply that a one standard deviation increase in the number of children on vacation boosts weekly album downloads by slightly more than one half of their mean,, an effect that is statistically significant and economically meaningful. Once we instrument for downloads, the estimated effect of file sharing on sales is small and statistically indistinguishable from zero.

We next consider specifications in which we add the band-on-tour-in-Germany interaction and the remaining time-varying instruments (columns III and IV). The tour and the German-chart interactions are of particular interest since they vary across albums as well as over time and provide an additional source of identification. The instruments have the expected first-stage signs. Tours and better chart positions magnify the effect of German students on vacation. The reverse is true for misspellings, which make it more difficult to search for files. Sargan

overidentification tests are reported at the bottom of the table. In these richer models downloads continue to have economically small and statistically insignificant effects on sales.

To help improve the precision of our second-stage estimates, in column (5), we allow the effect of the German vacation instrument to vary by album. The logic for including these interactions follows from the same arguments used for the other instruments. When German kids spend more time on P2P networks, the resulting supply shock will vary across albums because the students supply the files that happen to be popular in Germany at the time of the shock. As before, we face a potential problem with using this type of variation: If it so happens that the exogenous German shock is spuriously correlated with album-specific surges in popularity in the U.S., our estimates would be biased. The specification in column (5) addresses this issue in four ways. As before, we include album fixed effects to make sure it is the timing of the supply shocks that identify downloads. Second, we introduce album-specific U.S. popularity effects at both stages of the model by interacting the MTV variable with the album fixed effects. The model thus controls for changes in the U.S. popularity of a release. Third, relying on the assumption that the number of German kids on vacation is a valid instrument, we conduct overidentification tests in a specification that includes only two instruments: the vacation variable and one of the vacation × album-fixed-effect interactions. There are 680 such tests. To err on the side of caution, we exclude from the final specification all interactions whose overidentification tests cannot reject the null at a significance level of greater than 0.20. There are 21 such interactions. Fourth, we estimate a variant of equation (3), now with German kids × album fixed effect × U.S. MTV interactions. In the sales equation, these interactions are individually and collectively not different from zero.

Column (5) in Table 7 reports results with the album interactions. Our instruments retain their statistical significance.[22] The mean of the coefficients on the vacation-album-fixed-effect interactions is -1.143, leaving the average effect of vacations on downloads almost unchanged from the earlier specifications. Grouping the album interactions by genre, we find that vacations increase downloads the most for music types that are popular in Germany: the mean of the vacation-album-fixed-effect coefficients is -0.71 for International albums and -0.91 for Rock. In contrast, the effect of vacations is much smaller, but still positive, for genres that are less popular in Germany (the mean interactions are -1.52 for Latin music, -1.54 for Country, and -1.57 for Holiday music.) At the second stage, the estimated effect of downloads on sales is virtually unchanged in this specification, but the standard error drops considerably.

To see if our results are driven by our modeling choice for the time trend in downloads and sales, we replace the polynomial time trend with week fixed effects in columns (6) and (7) of Table 7. In these specifications, we lose the German-kids-on-vacation instrument because it does not vary across releases. The results remain similar, with more precise second-stage estimates when we allow the effect of vacations to vary by release (column VII).

Table 7 suggests file sharing had a surprisingly small effect on sales that is statistically indistinguishable from zero. The instrumented point estimates fall within a very narrow range and suggest that file sharing did not heavily impact the music industry as a whole. If file sharing were to be eliminated, the most negative estimate (column VI) implies industry sales for all of 2002 would increase by 6.5 million albums. Using the most positive estimate (column VII), industry sales would fall by 8.9 million copies.[23] In 2002, the industry sold 803 million CDs.

[22] The vacations × misspellings interaction is collinear with the vacations × album fixed effects and cannot be included in this specification.

[23] The impact is the difference between predicted sales and the fitted value when downloads are set at zero. Using equation (1), the summed impact for our album sample and for our 17 week observation period is $\sum_t \sum_i S_{it}(D_{it}) - S_{it}(0)$

The robustness of these results extends to specifications not reported in Table 7. For example, we arrive at the same conclusions if we omit the misspelling or the German rank instrument.

Dynamic Analysis

The models in Table 7 only allow for a contemporaneous effect of downloads on sales, but it is quite possible that downloads influence sales at a later point in time. For example, users might sample music which they consider buying in the future. In Table 8, we address this issue by studying the effect of several weeks of downloads on sales and by estimating Generalized Methods of Moments (GMM) models.

A difficulty with the first approach is that downloads are highly correlated across time, which prevents us from including downloads in past weeks as individual covariates. Instead, we study the effect of a weighted sum of current and past downloads on current sales. Downloads are instrumented using the core set of instruments (specification IV in Table 7) or the extended set (specification V). Our formal measure is the weighted stock of current and previous weekly downloads, $D_t^{Stock} = \sum_{s\geq 0}\delta_s \times D_{t-s}$.[24] In these models, we continue to find small and statistically insignificant effects for the weighted sum of three weeks of downloads, both in specifications with a polynomial time trend (Table 8, I&II) and with week fixed effects (III&IV). As in the panel results, standard errors drop significantly with the extended set of instruments (II&IV). We also constructed stock variables for the sum of downloads during the past four and six weeks and found no evidence of a sales crowd-out in these models.

$= \gamma \times \sum_t \sum_i D_{it}$. We multiply this number by a scaling factor to get the annual impact for the entire music industry, $\gamma \times 240$m (this calculation is described in more detail below Table 11).

[24] The weights δ_s are chosen in a grid search that minimizes the unexplained fraction of the variance in our sales equation subject to $\delta_s \geq \delta_{s+1}$. The optimal weights $(\delta_0, \ldots, \delta_T)$ are (1,0.1,0.1). It is interesting that the weights which best fit our data give much importance to downloads in the current week, while downloads further back in the past do not heavily influence sales. Oberholzer-Gee and Strumpf (2005) presents additional results showing that file sharers are impatient. These findings are consistent with those of Einav (2004) for movie consumption.

Models (5) and (6) in Table 8 use the GMM estimator developed by Arellano and Bond (1991). The GMM models are more general than the previous specifications in the sense that we do not need to make any assumptions about the appropriate lag structure. The lag of sales that is included on the right-hand side accounts for any effect that past downloads might have had on current sales. The model is estimated in first differences. We instrument for past sales using suitable lags of their own levels and our core set of first-differenced instruments.[25] Arellano-Bond tests for autocorrelation are applied to the first-difference equation residuals. Second-order autocorrelation would indicate that some lags of the dependent variable which are used as instruments are endogenous, but the tests reveal no such problem. The results of these models, with a polynomial time trend as in (5) or with week fixed effects as in (6), are similar to our previous findings. The estimates are fairly precise, making these GMM models an alternative to using our extended set of instruments.

"Drop-out" Hypothesis

A possible explanation for our inability to find a statistically significant relationship between file sharing and sales is that file sharers and consumers who purchase music are in fact two separate groups. According to this hypothesis, growth in file sharing does displace sales but we cannot identify this effect because our data do not reflect the increasing number of file sharers.

There are three responses to this conjecture. First, it is inconsistent with what we know about consumer behavior. The premise underlying the "drop-out" hypothesis is that file sharers no

[25]The formal model is,

$$S_{it} = \alpha S_{i,t-1} + X_{it}\beta + \gamma D_{it} + \omega_s t_s + \nu_i + \mu_{it}.$$

The lagged sales term soaks up any delayed effect of downloads, regardless of how far in the past they occurred (taking a Koyck transformation yields a specification with infinite lags of downloads on the right hand side). Estimating in first differences purges the album fixed effects. We instrument for the first-differenced $S_{i,t-1}$ which are now endogenous.

longer buy CDs. However, every survey we are aware of, including the industry studies listed in the literature section, indicates that downloaders, even heavy ones, continue to purchase legal CDs. We corroborated these findings with our own survey of individuals who were engaged in file sharing (Oberholzer-Gee and Strumpf 2005). Ninety percent reported that they recently purchased a CD, a value reaching one hundred percent among the most active downloaders.

Secondly, we can test the "drop-out" hypothesis directly by controlling for the increasing number of users. An implication of the hypothesis is that our download sampling rate declines over time because the servers for which we have data handle a limited number of users. Growth in file sharing, however, is managed by additional server capacity which we do not observe. If we accounted for this growth, the hypothesis suggests, we would find a displacement effect because the "drop-outs" are replacing purchases with transfers. We address this issue by scaling up the number of downloads in our sample to reflect the growth in file sharing. We use the number of FastTrack/KaZaA users as a proxy for the rate of growth.[26] Because the number of users increased by over a third over our observation period, we should be able to detect a drop-out effect if it exists. Table 9 reports these estimates for three panel models, three models using a stock of previous downloads, and for two GMM models. In all these specifications, downloads still do not have a significant effect on sales. A third approach to testing the drop-out hypothesis is to compare the long-run sales growth of individual genres of music. We return to this point in Section VII.

[26] We use 22 data points on the number of KaZaA users in the period from 9/9/2002 to 2/4/2003 to fit a fractional polynomial trend in the number of users. The model explains 85% of the variation.

Robustness Tests

To further corroborate our results, we perform a large number of robustness checks, some of which we report in Table 10.[27] The tests fall in three broad categories: models for subsets of our sample, alternative econometric specifications, and models that allow the effect of file sharing on sales to vary by popularity. We first investigate the importance of the holiday season when many consumers purchase CDs as gifts. It is possible that downloads are less substitutable for sales during this period due to the reluctance to give downloaded music as a present. Note that this is also an argument against the idea that file sharing is the main cause of the sales decline, since purchases are heavily concentrated in the holiday season. Still, it is straightforward to test for this effect. In Table 10, we exclude the December data from our sample. We report these results for specifications IV, VI and VII of Table 7. Even without the December data, there is no statistically significant effect of file sharing on sales. In a second test, we omit albums that are not downloaded during our study period. These less popular releases might have little sales even in the absence of file sharing, making the effect of P2P on sales miniscule by definition. Omitting these albums, however, does not change our conclusions. The same holds if we restrict our sample to better-selling albums.

We next test if the undersupply of Latin and Country music influences our estimates. Recall from Section V.D. that this would cause a problem only if the substitutability of downloads and album purchases varies across music genres. The last specification in the first panel of Table 10 re-estimates our models without Latin or Country releases. As expected, this increases the effect of vacations on downloads, from a coefficient estimate of 0.667 in model IV of Table 7 to 0.744 in this model. However, the measured effect of downloads on sales remains similar, a finding

[27]We thank our referees for suggesting several of these points. Many additional robustness tests can be found in Oberholzer-Gee and Strumpf (2005). This working paper also presents pooled specifications utilizing only cross-album variation, and these estimates also show file sharing has little impact on sales.

that is consistent with the idea that the substitutability of downloads and purchases is roughly similar across genres.

In the second panel in Table 10, we explore two alternative specifications. To reduce the importance of outlier albums with a large number of sales, we use log(sales) as the dependent variable. The impact on sales continues to be insignificant in all three specifications. In the next model, we first-difference both sales and downloads and express them as percentage changes. An advantage of this model is that it nicely captures album-specific trends in popularity. Unfortunately, this advantage comes at the cost of a reduced number of observations due to the first-differencing and the weeks with zero downloads or sales. Using our core set of instruments, we now find a positive and statistically significant but economically small effect of downloads on sales. However, the estimated coefficient drops considerably and is insignificant when we introduce week fixed effects.

The previous models constrained the effect of downloads on sales to be identical for all releases. In the bottom panel of Table 10, we relax this assumption. We first explore the idea that the effect varies by artist popularity. We do this by interacting the download variable with two measures of popularity: an artist's last and his best-ever Billboard ranking. The rankings themselves are subsumed in the album fixed effects, but the interaction term varies by week. To make it easier to interpret the results, Billboard ranks are coded as [201 − actual rank] so that larger numbers indicate greater popularity.[28] We estimate these models using specification IV in Table 7. There is no indication that more popular artists are affected differentially. Neither the interaction terms nor the joint effect of the main and interaction terms are statistically significant.

[28]More precisely, the term is a three-way interaction: [downloads × indicator that the artist had a Billboard ranking × (201−Billboard rank)].

From a welfare point of view, it is particularly interesting to study variations in the effect of file sharing across younger and older artists because such differences might influence their decision to start and continue a career in music. Interacting downloads with the number of albums an artist produced, we find no significant differences across more or less experienced performers. Finally, we investigate whether the effect of downloads on sales varies with the number of popular songs on an album. As documented earlier, most file sharers obtain just a few songs from an album. One might suspect that P2P is a fairly good substitute for albums with only one or two popular songs. We calculate a Herfindahl index for each album-week as a measure of concentration of downloads. The index is included in both the first and the second stage. There is no evidence that albums with more concentrated downloads suffer disproportionately from file sharing.

VII. Quasi-experimental Evidence

Our data also allow us to study the impact of P2P on sales in a quasi-experimental context. In particular we can examine how album sales respond to exogenous variation in file sharing intensity due to seasonality, geography, music genre, or secular growth. One of the advantages of this approach is that we can utilize several years of data, which allows us to investigate the long-term impact of file sharing. In all cases we continue to use sales data from Nielsen SoundScan (2005).

The first experiment involves variation over time. The number of file sharing users in the U.S. drops twelve percent over the summer (estimated from BigChampagne 2006) because college students are away from their high-speed campus Internet connections. If downloads crowd out sales, we should observe that the share of albums sold in the summer increases following the

advent of file-sharing. We consider a differences-in-differences approach and compare the share of summer sales in the period prior to file sharing (the control group) with sales following the introduction of file sharing (the treatment group). We calculate the share of album sales occurring in the May to September period using weekly SoundScan data. We find that the introduction of widespread file-sharing has had virtually no impact on summer sales. In the four years (1995-1998) preceding the introduction of Napster, the average share of summer sales was 37.0% with a range of 36.4-37.8%. During the more recent period of extensive file-sharing (1999-2005), the average share of summer sales was 37.2% with a range of 35.9-37.8%.

A second experiment considers spatial variation. Recall that U.S. users download over a third of their music files from Western European countries such as Germany and Italy. Due to time zone differences, such transfers are easier for East rather than West Coast users. This is because the peak file-sharing period (7pm to 3am) overlaps between Western Europe and the East Coast, which have a six hour time difference, but not between Europe and the West Coast, which have a nine hour difference. So East Coast users can draw on a larger base of files from international users than West Coast users. Consistent with these differences, we find that there is more file sharing on the East Coast than on the West Coast.[29] If file sharing had a large negative effect on record sales, then sales during the file sharing era should decrease more on the East Coast than on the West Coast. For the period 1998-2002, we obtained total album sales for the one hundred one largest "Designated Market Areas" from SoundScan. Despite the differences in the availability of files, sales have not noticeably varied across the country. In 1998, the last year in the pre-P2P period, the share of album sales in the Eastern Time Zone was 43.9%. This share has hardly moved since then. In 1999-2002, the mean was 43.5% and the range was 42.7-44.0%.

[29] Unfortunately, IP addresses can only be matched imperfectly to locations, so this finding is merely suggestive.

This is consistent with some common national factors, rather than file-sharing, driving sales trends.

A third experiment, which also provides a test of the "drop-out" hypothesis, is to see whether download intensity influences long-run sales growth after explicitly controlling for trends in music format popularity. The model for the period 1999-2005 is,

$$\text{Sales Growth}_g = \alpha + \gamma \times \text{Downloads}_g + \lambda \times \text{Listenership}_g + e_g \qquad (5)$$

where g indicates genre, Sales Growth_g is the percentage growth in sales over 1999-2005, Downloads_g are measures of genre-specific download intensity from our data, and Listenership_g is the genre-specific radio listenership growth rate (Arbitron 2006) which controls for trends in popularity. Since downloading is relatively concentrated across genres (Table 3), the "drop-out" hypothesis predicts a greater sales reduction for genres which are popular on file sharing networks. The estimated γ is not statistically significant using either download levels or downloads relative to purchases. For example, using mean downloads per album and controlling for genre sales levels, the estimated γ is 0.05 with a standard error of 0.52 (the mean for downloads is 61.2, and for sales growth it is -5.8).

Finally, we consider whether growth in file sharing can be linked to changes in total album sales. The key question is whether periods of particularly rapid growth in the user-base are linked to sharper sales reductions. A simple test is to consider annual sales since the advent of widespread file sharing in 1999. According to SoundScan, album sales increased in three of the seven years over this period, in contrast to movie ticket sales which rose in only two years. It is worth stressing that extended sales slumps are common in the music business, even prior to file sharing. While real revenues have fallen 28% over 1999-2005, real revenue fell 35% during the

collapse of disco music in 1978-1983. Real sales also dropped 6% over 1994-1997.[30] More direct evidence comes from regressing total album sales, including paid digital downloads, on the average number of simultaneous file sharing users in the U.S. (BigChampagne 2006),

$$Sales_t = \gamma \times Users_t + \nu_m + \mu_t \qquad (6)$$

where t indicates a month, and ν_m are monthly fixed effects which account for seasonality. Using monthly data from August 2002-May 2006 (N=46) and defining Sales and Users in millions (with respective sample means of 56.0m and 5.0m), the estimated γ=-0.427 with a robust standard error of 0.33. There is little evidence that growth in the number of users has had a statistically or economically significant effect on sales.[31] The estimates remain insignificant if equation (6) is estimated in first differences.

The results of these quasi experiments are consistent with our earlier findings. Looking at variation in downloading intensity that is due to geography, seasonality, the genre of music, or secular growth, we find no evidence that the advent of P2P technology is the primary cause of the recent slump in music sales.

VIII. Conclusions

Using detailed records of transfers of digital music files, we find that file sharing has had no statistically significant effect on purchases of the average album in our sample. Even our most negative point estimate (Table 7, model VI), implies that a one standard deviation increase in file-sharing reduces an album's weekly sales by a mere 368 copies, an effect that is too small to be statistically distinguishable from zero. Because our sample was constructed to be representative of the population of commercially relevant albums, we can use our estimates to

[30]These are calculated from nominal RIAA revenues listed in Lesk (2003) and RIAA (1998; 2006).
[31]If file sharing were eliminated, the point estimates imply monthly sales would only increase by 2.1m.

test hypotheses about the impact of P2P on the entire industry. Using ninety-five percent confidence bands, these tests are presented in Table 11. Taking into account all our (instrumented) estimates including the least precise results in Tables 7-9, we can reject a null that P2P caused a sales decline greater than 24.1 million albums. For reference, the music industry sold 803m CDs in 2002, which was a loss of 80m from the previous year (RIAA 2004). Our estimates become more precise if we relax the assumption that file sharing only impacts contemporaneous sales and if we allow for growth in the number of file sharers. For example, the scaled GMM models in Table 9 reject a null of losses greater than 6.6 million. Relying on our five most precise estimates, we conclude that the impact could not have been larger than 6.0 million albums. While file sharers downloaded billions of files in 2002, the consequences for the industry amounted to no more than 0.7% of sales

If file sharing is not the culprit, what other factors can explain the decline in music sales? Several plausible candidates exist. A first reason is the change in how music is distributed. Between 1999 and 2003, more than 14% of music sales shifted from record stores to more efficient discount retailers such as Wal-Mart, possibly reducing inventories. As a result, album shipments, which are often cited to document the decline in the legal demand of music, fell much more than actual sales.[32] A second factor is the ending of a period of atypically high sales, when consumers replaced older music formats with CDs. Perhaps more important than these developments is the growing competition from other forms of entertainment. A shift in entertainment spending towards recorded movies alone can largely explain the reduction in sales. The sales of DVDs and VHS tapes increased by over $5 billion between 1999 and 2003. This figure more than offsets the $2.6 billion reduction in album sales since 1999. Consumers also

[32] In the 1999 to 2003 period, the number of shipped albums fell by 301 million but the number of albums that were sold declined by only 99 million.

spent more on video games, where spending increased by 40%, or $3 billion, between 1999 and 2003, and on cell phones. Teen cell phone use alone tripled between 1999 and 2003.

An interesting question is whether our results continue to hold in more recent years. Since the time of our study, P2P technology has become more efficient, broadband access is much more widespread, and the number of file sharers has doubled. While a full analysis is outside the scope of this paper, there are several trends that are inconsistent with the view that P2P now displaces sales on a large scale. First, our natural experiments, for which we have data up to 2005, give no indication that file sharing has caused a sales decline in more recent years. Second, music sales have been flat or even rising in major markets with a quickly growing file-sharing population. For example, in 2005 retail music sales rose in four of the five largest national markets. Third, in the United States the entire drop in 2005 album sales is due to losses at a single firm, the recently merged Sony-BMG, which has experienced severe post-merger integration difficulties. If file sharing were responsible for the observed sales decline in the U.S., we would not expect this activity to only affect the products of a single firm.

The advent of the new P2P technologies can be considered in a broader context. A key question is how social welfare changes with weaker property rights for information goods. To make such a calculation, we would need to know how the production of music responds to the presence of file sharing. Based on our results, we do not believe file sharing had a significant effect on the supply of recorded music. For artists who produce commercially relevant products, the effects documented in this study are simply too small to change the number or quality of recordings that they release. And for new bands that are about to launch their career, the probability of success is so low as to make the expected income from producing music virtually zero, so file sharing will not change the relevant incentives. If we are correct in arguing that downloading has had

little effect on the incentives to produce music, we agree with Rob and Waldfogel (2006) who find that file sharing likely increased aggregate welfare. The limited shifts from sales to downloads are simply transfers between firms and consumers. But the sheer magnitude of P2P activity, the billions of songs downloaded each year, suggests the added social welfare from file sharing is likely to be high.

References

Agentur Lindner 2004. http://www.agentur-lindner.de/special/schulferien/index.html.

Arbitron 2006. *Format Trends Report.* http://wargod.arbitron.com/scripts/ndb/fmttrends2.asp.

Arellano, Manuel and Stephen Bond 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *The Review of Economic Studies* 58 (2): 277-97.

Bakos, Yannis, Erik Brynjolfsson and Douglas Lichtman 1999. "Shared Information Goods." *Journal of Law and Economics* 42: 117-156.

Berry, Steven 1994. "Estimating Discrete-Choice Models of Product Differentiation." *Rand Journal of Economics* 25: 242-262.

BigChampagne 2006. "Average Simultaneous U.S. Users: August 2002-May 2006." personal correspondence.

Billboard 2006. Fuzzy Math. 1 July: 24-26.

Blundell, Richard and Stephen Bond 1998. "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87 (1): 115-43.

Boldrin, Michele and David Levine 2002. "The Case Against Intellectual Property." *American Economic Review: Papers and Proceedings* 92: 209-212.

Bresnahan, Timothy, Scott Stern, and Manuel Trajtenberg 1997. "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the late 1980s." *Rand Journal of Economics* 28: S17-S44.

Business Software Alliance 2003. *Eighth Annual BSA Global Software Piracy Study.* http://www.bsa.org/globalstudy2003/index.cfm.

Central Intelligence Agency 2002. The World Factbook. https://www.cia.gov/cia/publications/factbook/index.html.

Central Intelligence Agency 2003. The World Factbook. https://www.cia.gov/cia/publications/factbook/index.html

Consumer Expenditure Survey 2004. http://www.bls.gov/cex/.

DEG 2005. "Industry Boosted by $21.2 Billion in Annual DVD Sales and Rentals." The Digital Entertainment Group. http://www.dvdinformation.com.

Edison Media Research 2003. *The National Record Buyers Study III.* Sponsored by Radio & Records. http://www.edisonresearch.com.

Einav, Liran (forthcoming). "Seasonality in the U.S. Motion Picture Industry". *Rand Journal of Economics*.

Forrester 2002. "Downloads Save the Music Business." http://www.forrester.com.

Forrester 2004. "US Antipiracy Bill Won't Stop File Sharing." http://www.forrester.com.

Gentzkow, Matthew forthcoming. "Valuing New Goods in a Model with Complementarity: Online Newspapers." *American Economic Review.*

Goolsbee, Austan 2000. "In a World Without Borders: The Impact of Taxes on Internet Commerce." *Quarterly Journal of Economics* 115: 561-576.

Goolsbee, Austan and Judy Chevalier 2003. "Price Competition Online: Amazon Versus Barnes And Noble," *Quantitative Marketing and Economics* 1 (2), June 2003, 203-222.

Gummadi, Krishna, Richard Dunn, Stefan Saroiu, Steven Gribble, Henry Levy, and John Zahorjan 2003. "Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload." Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP-19).

International Federation of the Phonographic Industry 2002. *Recording Industry in Numbers 2001*. International Federation of Phonographic Industry.

Imbens, Guido and Joshua Angrist 1994. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62: 467-475.

Internet2 Netflow Statistics (2004). *Internet2 NetFlow: Weekly Reports.* `http://netflow.Internet2.edu/weekly/.`

Jupiter Media Metrix (2002). "File Sharing: To Preserve Market Value Look Beyond Easy Scapegoats." `http://www.jupiterresearch.com.`

Karagiannis, Thomas, Andre Broido, Nevil Brownlee, kc claffy, and Michalis Faloutsos (2004). "Is P2P dying or just hiding?" Presented at Globecom 2004 in November-December 2004. `http://www.caida.org/outreach/papers/2004/p2p-dying/.`

Klein, Benjamin, Andres Lerner, and Kevin Murphy (2002). "The Economics of Copyright 'Fair Use' in a Networked World." *American Economic Review: Papers and Proceedings.* 92: 205-208.

Kultusministerkonferenz, Statistische Veröffentlichungen (2002). Nummer 162 vom August.

Lesk, Michael 2003. "Chicken Little and the Recorded Music Crisis." IEEE Security & Privacy (September): 73-75.

Liang, Jian, Rakesh Kumar, and Keith Ross 2004. "Understanding KaZaA." Manuscript, Polytechnic University.

MPAA 2005. "U.S. Entertainment Industry: 2004 MPA Market Statistics." Motion Picture Association: Worldwide Market Research.

Musikmarkt 2002. "Deutschland Single-Charts." `http://musikmarkt.lw-t1.thuecom-medien.de/content/charts/history.php3?jahr=2002.`

Nevo, Aviv 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69: 307-342.

Nielsen SoundScan (2005). http://home.soundscan.com/about.html.

Niesyto, Horst 2002. Digitale Spaltung - digitale Chancen: Medienbildung mit Jugendlichen aus benachteiligenden Verhältnissen. Mimeo, Pädagogische Hochschule Ludwigsburg.

Oberholzer-Gee, Felix and Koleman Strumpf 2005. "The Effect of File Sharing on Sales: An Empirical Analysis." Manuscript, Harvard Business School and the University of North Carolina at Chapel Hill.

OECD 2004. *OECD Information Technology Outlook 2004.* Paris: Organisation for Economic Co-operation and Development.

Plant, Arnold 1934. "The Economic Aspects of Copyright in Books." *Economica* 1: 167-195.

Posner, Richard 2005. "Intellectual Property: The Law and Economics Approach." *Journal of Economic Perspectives* 19 (2): 57-73.

RIAA 1998. *RIAA 1996 Statistical Overview.* Archived copy from the Internet archive, http://web.archive.org/web/19980124173720/www.riaa.com/market/releases/statover.htm.

RIAA 2004. *RIAA Market Data: The Cost of a CD.* Archived copy from the Internet archive, http://web.archive.org/web/20030416004543/http://www.riaa.com/MD-U.S.-7.cfm.

RIAA 2006. *The Recording Industry Association of America's 2005 Yearend Statistics.* http://www.riaa.com.

Rob, Rafael and Joel Waldfogel 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(1): 29-62.

Shapiro, Carl and Hal Varian 1999. *Information Rules: A Strategic Guide to the Network Economy*. Boston: Harvard Business School Press.

Takeyama, Lisa 1994. "The Welfare Implications of Unauthorized Reproduction of Intellectual Property in the Presence of Demand Network Externalities." *The Journal of Industrial Economics* 42: 155-166.

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

Varian, Hal 2000. "Buying, Sharing and Renting Information Goods." *The Journal of Industrial Economics* 48: 473-488.

Windmeijer, Frank 2000. "A finite sample correction for the variance of linear two-step GMM estimators." Institute for Fiscal Studies, IFS Working Papers: W00/19.

Zentner, Alejandro 2006. "Measuring the Effect of Music Downloads on Music Purchases." *Journal of Law & Economics* 49(1): 63-90.

TABLE 1

SAMPLE SALES BY CATEGORY

	Observations	Mean sales	Std dev	Min	Max
Full sample	680	143,096	344,476	74	3,430,264
Catalogue	50	46,833	40,031	219	223,085
Current Alternative	117	118,599	130,257	9,210	785,747
Hard Music Top Overall	19	28,304	22,103	2,945	86,416
Jazz Current	21	21,940	62,522	86	290,026
Latin	21	27,590	35,840	3,143	153,209
New artists	50	15,816	13,635	319	61,673
R&B	144	46,512	67,050	2,151	457,338
Rap	76	39,307	61,278	1,069	324,426
Top Current ("Billboard 200")	83	744,022	710,054	4,092	3,430,264
Top Current Country	66	87,839	130,096	74	669,575
Top Soundtrack	33	44,920	79,264	1,788	318,538

NOTE.— These figures only include sales over our seventeen week observation period. Most of the top-selling albums are classified as "Current" for the purposes of this table

TABLE 2
THE GEOGRAPHY OF FILE SHARING
(numbers in %)

Country	Share of users	Share of downloads	Users in U.S. download from (%)	Users in U.S. upload to (%)	Share World Population	Share World GDP	Share World Internet Users	Software Piracy Rate
United States	30.9	35.7	45.1	49.0	4.6	21.2	27.4	23
Germany	13.5	14.1	16.5	8.9	1.3	4.5	5.3	32
Italy	11.1	9.9	6.1	5.7	0.9	2.9	3.2	47
Japan	8.4	2.8	2.5	1.8	2.0	7.2	9.3	35
France	6.9	6.9	3.8	4.7	1.0	3.1	2.8	43
Canada	5.4	6.1	6.9	7.9	0.5	1.9	2.8	39
United Kingdom	4.1	4.0	4.2	4.2	1.0	3.1	5.7	26
Spain	2.5	2.6	1.8	2.0	0.6	1.7	1.3	47
Netherlands	2.1	2.1	1.9	1.6	0.3	0.9	1.6	36
Australia	1.6	1.9	0.8	2.2	0.3	1.1	1.8	32
Sweden	1.5	1.7	1.8	1.5	0.1	0.5	1.0	29
Switzerland	1.4	1.5	0.9	1.0	0.1	0.5	0.6	32
Brazil	1.3	1.4	1.2	1.3	2.9	2.7	2.3	55
Belgium	0.9	1.2	0.5	1.0	0.2	0.6	0.6	31
Austria	0.8	0.6	0.6	0.4	0.1	0.5	0.6	30
Poland	0.5	0.7	0.7	0.5	0.6	0.8	1.1	54

NOTE.— Shares of users and downloads is from the file sharing dataset described in the text. All other statistics are from the Central Intelligence Agency (2002, 2003), except the software piracy rates which are from the Business Software Alliance (2003). All values are world shares, except the piracy rates are the fractions of business application software installed without a license in the country. All non-file sharing data are for 2002 except population which is for 2003.

TABLE 3
DOWNLOADS BY GENRE

	# songs (# albums) in sample	Mean # of downloads	Std dev	Min	Max
			Song level		
All genres	10271	4.645	21.462	0	1258
Catalogue	714	4.361	10.370	0	152
Alternative	1707	7.021	18.153	0	312
Hard	270	4.830	8.684	0	52
Jazz	261	0.333	0.920	0	7
Latin	309	0.550	2.927	0	28
New artists	711	0.609	7.039	0	184
R&B	2249	1.635	7.680	0	159
Rap	1227	0.920	4.887	0	82
Current	1342	17.182	51.286	0	1258
Country	913	1.974	6.382	0	128
Soundtrack	568	1.673	5.301	0	61
			Album level		
All genres	680	70.162	158.628	0	1799
Catalogue	50	62.280	103.114	0	680
Alternative	117	102.436	122.794	0	674
Hard	19	68.632	82.899	0	264
Jazz	21	4.143	4.542	0	13
Latin	21	8.095	26.344	0	121
New artists	50	8.660	33.097	0	229
R&B	144	25.542	56.494	0	433
Rap	76	14.855	24.487	0	119
Current	83	277.807	333.935	2	1799
Country	66	27.303	51.649	0	344
Soundtrack	33	28.788	36.611	0	185

TABLE 4

DOWNLOADS BY SALES – ALBUM LEVEL

	Obs	Mean # of downloads	Std dev	Min	Max	Mann-Whitney
1st quartile: mean 7,235 copies [up to 12,493 copies]	170	11.358	38.472	0	402	- 14.067**
2nd quartile: mean 21,022 copies [up to 31,115 copies]	170	20.929	52.082	0	433	-12.431**
3rd quartile: mean 57.940 copies [up to 100,962 copies]	170	48.088	55.223	0	264	-8.187**
4th quartile: mean 486,184 copies [max 3,430,264 copies]	170	200.270	265.369	0	1799	

NOTE.— Mann Whitney test statistics are for the null that the 4th quartile with the highest sales comes from the same population as the other sales quartiles.

** significant at the 1% level

TABLE 5

SUMMARY STATISTICS

	Observations	mean (std dev)	min	max
Sales (1,000s)	10093	9.580 (34.361)	0	874.137
Downloads	10093	4.360 (13.644)	0	368
German kids on Vacation (million)	10093	9.855 (3.576)	0	12.491
Band on tour in Germany	10093	0.003 (0.053)	0	1
Misspelling indicator	10093	0.062 (0.187)	0	1
Rank of single on German charts (calculated as 101 minus rank)	10093	1.576 (10.268)	0	100
Rank of single on MTV charts (calculated as 101 minus rank)	10093	2.158 (13.568)	0	100
Billboard rank previous album (calculated as 201 minus rank)	10093	61.136 (82.314)	0	200
Best Billboard rank ever (calculated as 201 minus rank)	10093	83.548 (89.994)	3	200
# previous releases	10093	6.718 (15.574)	0	194
HHI downloads	10093	2.460 (3.672)	0	10000

TABLE 6

DOWNLOAD TIMES: RELATION TO INSTRUMENTS AND IMPACT ON NUMBER OF TRANSFERS

	(1)	(2)	(3)	(4)	(5)	(6)	
						Impact of download time on download quantity	
	Time: Download Request to Initiation (sec) C_1	Time: Search Request to Download Request (sec) C_2	Time: Initiation Download to Completion (sec) C_3	Ratio: # Search Requests to # Downloads C_4	Percentage: Download Requests which are not completed C_5	Download Time (1st stage) $C_1+C_2+C_3$	Downloads (2nd stage) D_{it}
German kids on Vacation (million)	-32.005	-4.336	-26.031	-0.453	-2.351	-62.420	
	(5.51)**	(0.29)**	(2.69)**	(0.05)**	(0.10)**	(5.24)**	
German kids × Band on tour	-49.914	-3.966	-35.015	-0.480	-2.927	-89.010	
	(20.31)*	(1.73)*	(13.35)**	(0.22)*	(0.51)**	(17.83)**	
German kids × Misspellings	22.494	6.157	8.609	0.672	1.963	7.302	
	(33.66)	(2.182)**	(17.76)	(0.25)**	(0.58)**	(40.59)	
German kids × rank German charts	-0.347	-0.034	-0.471	-0.005	-0.024	-0.849	
	(0.18)*	(0.02)	(0.16)*	(0.00)*	(0.01)*	(0.22)**	
Download time							-0.006
							(0.00)**
Album Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1662	1952	1332	2164	1952	1332	1332
Mean for Dependent Variable	609.08	91.02	796.20	12.21	62.96	1491.18	7.25

NOTE.— Albums or album-weeks are omitted when the dependent variable is undefined (e.g. for C_1 when there are no successful album download initiations). Robust standard errors are in parentheses. These estimates are based on data from weeks 3-6 of our observation period (the data come from more detailed log files which are only available during these weeks).

* significant at the 5% level

** significant at the 1% level

TABLE 7

PANEL ANALYSIS - DOWNLOADS AND ALBUM SALES

	(1)	(2)		(3)		(4)		(5)		(6)		(7)	
	Sales	1st stage down-loads	2nd stage Sales	1st stage down-loads	2nd stage sales	1st stage down-loads	2nd stage Sales	1st stage down-loads	2nd stage sales	1st stage down-loads	2nd stage Sales	1st stage down-loads	2nd stage sales
# downloads	0.277		0.003		0.024		-0.010		0.005		-0.027		0.037
	(0.025)**		(0.194)		(0.189)		(0.158)		(0.062)		(0.270)		(0.065)
German kids on vacation		0.671		0.670		0.667		1.818					
		(0.054)**		(0.054)**		(0.054)**		(0.125)**					
German kids × band on tour				0.469		0.474		0.470		0.464		0.451	
				(0.168)**		(0.167)**		(0.161)**		(0.167)**		(0.161)**	
German kids × Misspellings						-0.288				-0.290			
						(0.124)*				(0.124)*			
German kids × Germ charts						0.012		0.007		0.012		0.007	
						(0.001)**		(0.002)**		(0.001)**		(0.002)**	
U.S. MTV rank	0.079	0.036	0.089	0.037	0.088	0.035	0.089	0.058	-0.194	0.036	0.092	-0.042	-0.183
	(0.020)**	(0.008)**	(0.021)**	(0.008)**	(0.021)**	(0.008)**	(0.021)**	(0.103)	(0.256)	(0.008)**	(0.022)**	(0.102)	(0.255)
German kids × album FE	No	No	No	No	No	No	No	Yes	No	No	No	Yes	No
MTV × album FE	No	No	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Polynomial time trend	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No
Week FE	No	No	No	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Album FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	10093	10093	10093	10093	10093	10093	10093	10093	10093	10093	10093	10093	10093
Prob χ^2>0 on excluded instruments		0.0000		0.0000		0.0000		0.0000		0.0000		0.0000	
Sargan test (p-value)					0.73		0.70		0.98		0.50		0.97
R-squared	0.75	0.74	0.76	0.74	0.76	0.73	0.76	0.74	0.79	0.82	0.77	0.85	0.79

NOTE.— The unit of analysis is the album-week. Dependent variables are the number downloads at the 1st stage (summing all songs on an album) and album sales (1,000s).

Robust standard errors are in parentheses. Since all models include album fixed effects, the reported *R*-squared is the sum of the explained within-variance and the fraction of the variance that is due to the fixed effects. Album-weeks prior to the release date are excluded from the sample.

* significant at the 5% level

** significant at the 1% level

TABLE 8

DYNAMIC PANEL ANALYSIS - DOWNLOADS AND LAGGED ALBUM SALES

	(1) 2nd stage Sales	(2) 2nd stage sales	(3) 2nd stage Sales	(4) 2nd stage sales	(5) GMM Δ sales	(6) GMM Δ sales
Weighted ∑ of three weeks of downloads (instrumented)	0.097 (0.115)	0.048 (0.039)	0.022 (0.170)	0.045 (0.041)		
Δ downloads					0.029 (0.074)	0.047 (0.078)
U.S. MTV rank	0.092 (0.015)**	-0.016 (0.169)	0.097 (0.016)**	-0.022 (0.168)	0.085 (0.091)	0.041 (0.080)
lagged sales					0.166 (0.100))	0.261 (0.117)*
German kids × album FE in 1st stage	No	Yes	No	Yes	No	No
MTV × album FE	No	Yes	No	Yes	No	No
Polynomial time trend?	Yes	Yes	No	No	Yes	No
Week Fixed Effects?	No	No	Yes	Yes	No	Yes
Album Fixed Effects?	Yes	Yes	Yes	Yes	No	No
1st-stage specification is as in Table 7, model	4	5	6	7		
Observations	8739	8739	8739	8739	8739	8739
Arellano-Bond test for AR(1) in first differences: Pr > z					0.302	0.204
Arellano-Bond test for AR(2) in first differences: Pr > z					0.638	0.522
R-squared	0.92	0.96	0.92	0.97		

NOTE.— The dependent variable is album sales (1,000s). The number of downloads is instrumented using the Table 7 specification listed in the fifth row from the bottom. The weighted sum of three weeks of downloads includes the current week. The weights are chosen in a grid search which minimizes the unexplained fraction of the variance in our models. Models (5) and (6) use the Generalized Method of Moments estimator developed by Arellano and Bond (1991). In this model, the typical standard error estimator tends to be downwards biased (Blundell and Bond 1998). Standard errors are corrected using the two-step covariance matrix derived by Windmeijer (2000). Arellano-Bond tests for autocorrelation are applied to the first-difference equation residuals. Second-order autocorrelation would indicate that some lags of the dependent variable which are used as instruments are endogenous. The tests reveal no such problem. Album-weeks prior to the release date are excluded from the sample.

* significant at the 5% level

** significant at the 1% level

TABLE 9

ROBUSTNESS CHECK WITH SCALED DOWNLOADS – TESTING THE "DROP-OUT" HYPOTHESIS

	(1)		(2)		(3)		(4)	(5)	(6)	(7)	(8)
	1st stage downloads	2nd stage Sales	1st stage downloads	2nd stage sales	1st stage downloads	2nd stage sales	2nd stage Sales	2nd stage Sales	2nd stage Sales	GMM Δ sales	GMM Δ sales
Scaled downloads		-0.009		0.022		0.029					
		(0.126)		(0.046)		(0.049)					
Weighted ∑ of three Weeks downloads							0.078	0.038	0.037		
							(0.093)	(0.030)	(0.031)		
Δ downloads										0.072	0.123
										(0.053)	(0.072)
German kids on Vacation (million)	0.856		2.608								
	(0.073)**		(0.171)**								
German kids × Band on tour	0.602		0.600		0.585						
	(0.225)**		(0.216)**		(0.216)**						
German kids × Misspellings	-0.377										
	(0.167)*										
German kids × rank German charts	0.014		0.008		0.008						
	(0.002)**		(0.002)**		(0.002)**						
U.S. MTV rank	0.036	0.089	-0.084	-0.198	-0.059	-0.182	0.093	0.139	-0.023	0.085	0.044
	(0.011)**	(0.020)**	(0.137)	(0.255)	(0.137)	(0.255)	(0.015)**	(0.158)	(0.168)	(0.097)	(0.077)
Lagged sales										0.166	0.261
										(0.101)	(0.118)*
German kids × album FE in 1st stage	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No
MTV × album FE	No	No	Yes	Yes	Yes	Yes	No	Yes	Yes	No	No
Polynomial time trend	Yes	Yes	Yes	Yes	No	No	Yes	Yes	No	Yes	No
Week Fixed Effects?	No	No	No	No	Yes	Yes	No	No	Yes	No	Yes
Album Fixed Effects?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No
Specification as in Table (model)	7 (4)	7 (4)	7 (5)	7 (5)	7 (7)	7 (7)	8 (1)	8 (2)	8 (4)	8 (5)	8 (6)
Observations	10093	10093	10093	10093	10093	10093	8739	8739	8739	8739	
R-squared	0.74	0.76	0.85	0.79	0.87	0.79	0.82	0.86	0.87		
AB test for AR(1)										0.305	0.201
AB test for AR(2)										0.643	0.531

NOTE.— Dependent variables are album sales (1,000s) and scaled downloads at the 1st stage. Downloads are scaled to reflect the growth of KaZaA users over the sample period. For the fixed-effects models, the reported *R*-squared is the sum of the explained within-variance and the fraction of the variance that is due to the fixed effects. Album-weeks prior to the release date are excluded from the sample.

* significant at the 5% level

** significant at the 1% level

TABLE 10

ROBUSTNESS CHECKS

Table 7 (4) Coefficient downloads (std. error)	Table 7 (6) Coefficient downloads (std. error)	Table 7 (7) Coefficient downloads (std. error)	*N*	Specification
-0.010	0.005	0.037	10093	Benchmark specifications, models (4), (6) and (7)
(0.158)	(0.062)	(0.065)		in Table 7
	Changes in Sample			
0.064	-0.001	-0.013	7399	Without holiday sales
(0.376)	(0.108)	(0.112)		
0.018	0.034	0.079	7890	Without albums that are not downloaded
(0.166)	(0.071)	(0.075)		
0.051	0.083	0.161	5033	Albums that sell more than 151,284 copies (50^{th}
(0.184)	(0.090)	(0.097)		percentile) during the sample period
0.037	0.062	0.092	8567	Without Latin and Country albums
(0.135)	(0.055)	(0.058)		
	Changes in Model Specification			
-0.006	0.001	0.004	10093	Dependent variable is log of sales
(0.007)	(0.003)	(0.003)		
0.083	0.019	0.005	3232	Sales and downloads are expressed as percentage
(0.029)**	(0.026)	(0.022)		changes
	Does the estimated effect vary by popularity?			
Main effect downloads	Interaction	H_0 sum = 0 (Prob > F)		Downloads (instrumented) are interacted with…
-0.095	0.001	0.6119	10093	Billboard rank of artist's prior album
(0.185)	(0.001)			
-0.130	0.001	0.5015	10093	Best Billboard rank for artist during career
(0.192)	(0.001)			
0.002	0.002	0.9822	10093	Number of previous albums
(0.181)	(0.007)			
-0.128	0.039	0.5917	10093	Herfindahl index measuring concentration of
(0.175)	(0.026)			downloads

NOTE.— Dependent variables are album sales (1,000s) and # downloads at the 1st stage. Robust standard errors are in parentheses. For the popularity results in the lower panel, the specification is model (5) in Table 7. Album-weeks prior to the release date are excluded from the sample.

* significant at the 5% level

** significant at the 1% level

TABLE 11
HYPOTHESES TESTS

Class of Models	Lower bound of 95% confidence interval Can reject hypothesis that the impact of file sharing is larger than (in million albums)
All models (Tables 7 through 9)	-24.1
Models with German vacation × Album FE interactions	-12.7
Models with scaled downloads (Table 9)	-12.4
GMM models with scaled downloads (Table 9)	-6.6
5 models with smallest standard errors	-6.0

NOTE.— These values represent the overall, industry-wide impact of file sharing for 2002 as implied by the various specifications. The lower bound is the minimum of the 95% confidence interval around the mean impact. Details of this calculation are listed below. The second column of each row reports the median lower bound for that class of models.

The lower bound is calculated as $\sum_t\sum_i (D_{it}\times 5.04\times 1000)\times(\gamma - 2\times se(\gamma)) = 240m\times(\gamma - 2\times se(\gamma))$, where γ is the point estimate from equation (1). The factor 5.04 scales the results from our sample to all releases and the entire year 2002. It is calculated as: Aggregate impact = (Effect of file sharing on sample sales over observation period) × (population sales/sample sales) × (file sharing activity over year/file sharing activity in observation period). From our sales data, the ratio (population sales/sample sales) is 2.27. The second ratio is (File sharing activity over year/file sharing activity in observation period) = 2.22, which is calculated from weekly file sharing traffic rates over the 2002 calendar year on the Internet2 backbone (Internet2 Netflow Statistics 2004) and the monthly average number of U.S. file sharing users (BigChampagne 2006). Note that the second conversion factor is close to a naïve correction based simply on time, (52 weeks in year/17 weeks in observation period) = 3.06.

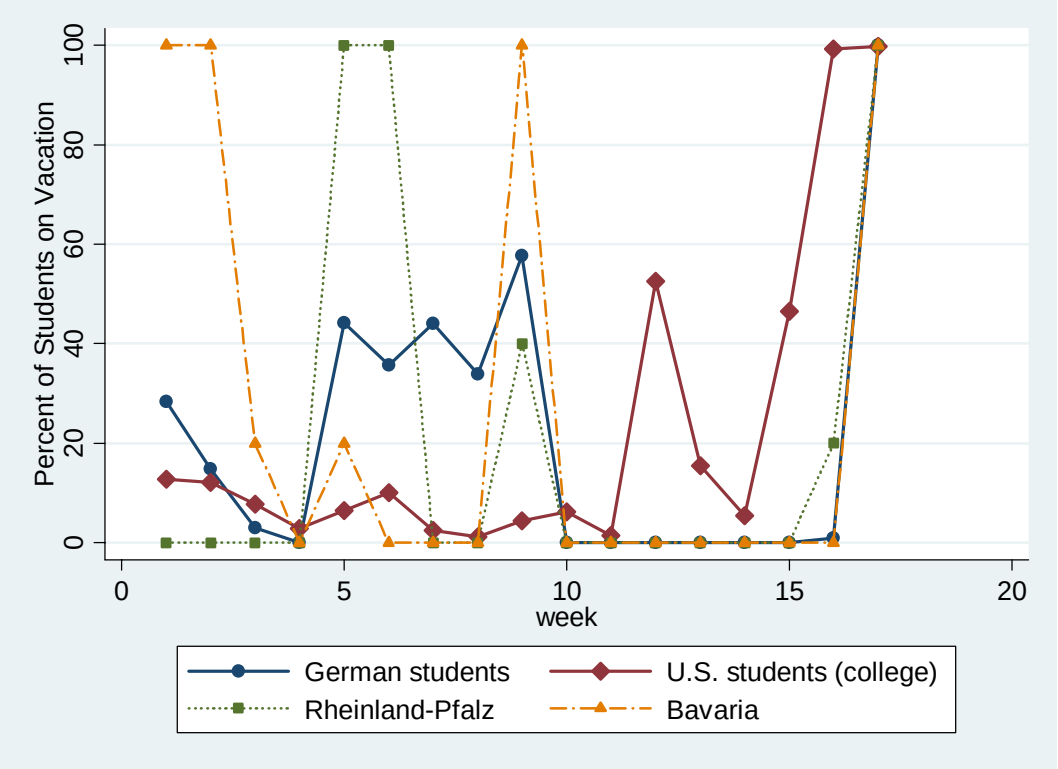


Fig. 1.— Timing of German and U.S. School Vacations

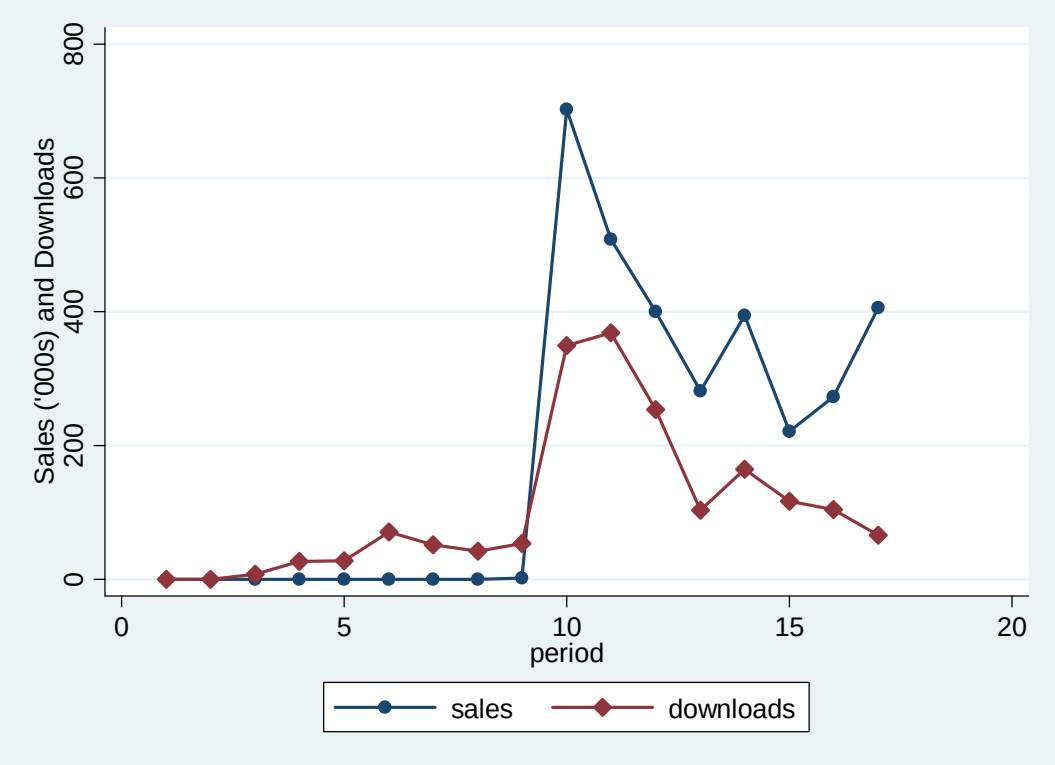


Fig. 2.– Dynamics of Downloads and Albums Purchases for a Popular Album
(by week, sales in thousands)