Queen's Economics Department Working Paper No. 1354

Purchase, Pirate, Publicize: Private-Network Music Sharing and Market Album Sales

Jonathan F. Lee
Queen's University

Department of Economics
Queen's University
94 University Avenue
Kingston, Ontario, Canada
K7L 3N6

1-2018

Purchase, Pirate, Publicize: Private–Network Music Sharing and Market Album Sales

Jonathan F. Lee[1]

Abstract

I quantify the effects of private–network music sharing on aggregate album sales in the BitTorrent era using a panel of US sales and private–network downloads for 2,109 albums during 2008. Exogenous shocks to the network's sharing constraints address the simultaneity problem. In theory, private–network activity could crowd out sales by building aggregate file sharing capacity or increase sales through word of mouth. I find evidence that private–network sharing results in decreased album sales for top–tier artists, though the economic impact is quite modest. However, private–network activity seems to help mid–tier artists. The results are consistent with claims that word of mouth is stronger for lesser–known artists and that digital sales are more vulnerable to increases in file sharing capacity. I discuss policy implications and alternatives to costly legal efforts to shut down private file sharing networks.

Keywords: intellectual property, copyright, file sharing, piracy, digital music

JEL–Classification: L82, L86, O34
Published Version: https://doi.org/10.1016/j.infoecopol.2018.01.001

[1]Queen's University, Dept. of Economics. Contact: leejf@econ.queensu.ca. This paper is a major revision of a previous project entitled "Private Filesharing Markets and Album Sales" (Lee, 2009). I am grateful to faculty and colleagues at Queen's University and Vanderbilt University, attendees of the American Law and Economics Association 2016 Annual Meeting, Stan Liebowitz, and anonymous referees for helpful comments and discussion. I am especially indebted to the John Deustch Institute, whose financial support enabled my use of the Nielsen SoundScan data. All errors are my own.

The relationship between media production and media piracy is not as straightforward as each side of the debate might claim. To copyright holders, every illicit transaction represents the loss of a legitimate purchase that might otherwise have happened. However, many pirates would never have purchased at the price the producer had set, and these new illicit consumers may increase exposure of the product. Such exposure may induce new transactions that might otherwise have never happened, and these transactions may accrue to the copyright holders themselves. How the tension resolves is thus an empirical question. Does the substitution of piracy for purchasing overwhelm the possibilities of a larger audience, or do new consumers outnumber the forgone sales to pirates?

This paper addresses that empirical question in the market for recorded music and its file sharing counterpart. Drawing from data on US album sales and on activity within a private file sharing network, I follow 2,109 albums over 27 weeks in 2008 to estimate the effect of an exogenous change in private–network file sharing activity on album sales. I find that the file sharing elasticity of sales is -0.02 for physical sales, -0.04 for digital sales, and -0.02 overall. I interpret these results as evidence that private–network piracy leads to a crowding out of legitimate sales and that this crowd–out is more significant in the market for digital music, but that the practical extent of these effects is quite small. The results are less clear–cut when controlling for artist popularity; effects are negative for top–tier artists but positive for mid–tier artists. I take these results as evidence that private–network activity leads to a crowding out of sales for artists with an established reputation but can act as a channel through which word–of–mouth increases exposure (and sales) of music by less–established artists. Again, the economic magnitudes of these effects are fairly small.

It is crucial to understand exactly what these results measure, especially when considering their implications for copyright policy. The data measure music piracy at a single private sharing network and measure sales for the entire US market; I cannot and do not attempt to quantify individual consumers' elasticities of demand in this paper. Instead, I measure the impact of a single file sharing network on the whole music market, which is the relevant measure when law enforcement organizations are deciding whether to take action against a single file sharing network. This would be a nonsensical effect to try and measure in conventional markets ("What is the effect of increased car theft in Honolulu on car sales for the US as a whole?"), but the social nature of sharing networks and the fact that the goods shared on them are infinitely replicable mean that activity in one corner of the market can spill over into and magnify activity in the rest of the market. The results in this paper, interpreted in the context of previous evidence on the relationship between sales and all piracy, are consistent with a private piracy elasticity of aggregate piracy of 0.15; that is, a 10% increase in private-network music sharing will spill over and manifest as a 1.5% increase in piracy overall.

In equilibrium, sales and piracy are simultaneously determined: the unobserved effects of album popularity, media exposure, and other variables that impact music consumption will influence sales and downloads alike. Thus identification of the effect of piracy on sales requires an exogenous covariate. Fortunately, the file sharing data that I use include such covariates. The file sharing network under study requires that a user's ratio of lifetime uploading to downloading must exceed a certain threshold, or the user will be banned from the network. In other words, users must give back in some proportion to what they receive. It follows that the more slack this constraint is for a user, the more that user can download. There are events during the sample period where users are credited for uploading, but not for downloading, known as freeleeches. These freeleech periods alter the slackness of the user's ratio constraint, which elicits exogenous variation in file sharing on the entire network. I use an assortment of freeleech measures and ratio slackness measures as instrumental variables, and I provide robust support for instrument suitability in first–stage results as well as in post–estimation testing.

The results are of both academic and practical interest. The relationships between physical, digital, and illicit markets is illuminating in its own right, and the interaction of conventional markets with diffuse digital markets is of broad interest to economists. But the results can also inform business and policy decisions in the market for music and for other media as well. Trade groups such as the Recording Industry Association of America (RIAA) and the International Federation of the Phonographic Industry (IFPI), alongside national and international law enforcement agencies, spend considerable effort and resources to deter piracy and shut down file sharing networks like the one studied in this paper. If these networks have only a small effect on sales, these efforts may be better allocated elsewhere.[2] I discuss alternative policies that include a more focused anti–piracy effort that concentrates only on top–tier artists' music, facilitating legitimate digital consumption,

[2] See BBC (2007) and Fisher (2007) for an example of this point.

and varying licensing royalties to amplify file sharing's benefits and mitigating its costs. The paper's results should help to inform policy and business strategies by trade groups, law enforcement agencies, and policymakers.

Review of Existing Literature

Researchers have spent considerable time studying the effect of file sharing on the music market. A clear picture has not emerged, but research does focus on two main arguments. The "traditional" view argues that piracy simply substitutes away from legitimate sales, which is tantamount to theft in the short run and degrades the incentives to create music in the long run. Therefore strong protection of intellectual property is needed to inhibit piracy and provide adequate incentives to create new music. The other view argues that even if substitution does occur, it is certainly not at a one–to–one rate, and that file sharing is a highly effective distribution method which allows sampling, spreads information about music quality, and gives smaller artists easy and direct access to listeners. These channels can create new consumers who would never have purchased the music otherwise. Theoretical and empirical work has investigated both arguments, and consensus is elusive.

Numerous surveys and meta–analyses of existing research have been carried out to determine which of the two arguments is more relevant. Depending on the study, authors conclude that consensus has not been reached (Connolly and Krueger, 2006), that the effect is negligible (Oberholzer-Gee and Strumpf, 2010), or that the effect is positive (Dejean, 2009). Other studies examine the evidence and conclude that the effect is decidedly negative (Liebowitz, 2005a,b, 2006a,b). I provide a short overview of the literature below, but the interested reader should consult these reviews for a more thorough consideration.

Theorists have argued for the possibility of a "sampling" effect, wherein file sharing allows users to try before they buy, and concluded that empirical testing is needed to determine whether the sampling effect actually outweighs the conventional substitution effect (Peitz and Waelbroeck, 2006a,b; Gopal et al., 2006). I interpret the current paper's findings in the context of a word–of–mouth effect which is similar to the sampling effect, but incorporates social network structure.

Since the effect of file sharing is fundamentally an empirical question, many studies have been carried out to determine the effect's direction and importance. The majority of these studies find a negative effect, whether using survey data (Waldfogel, 2010; Zentner, 2006; Rob and Waldfogel, 2006;

Leung, 2008), macro–level data with proxies for file sharing such as broadband access (Peitz and Waelbroeck, 2004; Danaher et al., 2014; Hui and Png, 2003; Liebowitz, 2008), or the emergence of file sharing as a natural experiment (Mortimer et al., 2012; Hong, 2013). Other studies find no statistically significant effect in survey data (Andersen and Frenz, 2010) or on long–run trends in music quantity (Waldfogel, 2011) and music quality (Waldfogel, 2012). However, none of these studies observes both sales *and* piracy at the album level; they instead rely on survey–based, proxied, or aggregated measures of file sharing activity.

Only a few studies exist that observe sales and file sharing at the album level. Oberholzer-Gee and Strumpf (2007) find no evidence of a statistically significant effect of file sharing on album sales, using German school vacations as a source of exogenous variation in available files. Blackburn (2006) estimates the effect of album–level file sharing supply on sales, using RIAA legal action as an exogenous file sharing risk shock. The author concludes that sales for less popular artists benefit from file sharing, sales for more popular artists suffer, and that these effects zero out on net.

This paper uses a similar data structure to the above album–level studies but nevertheless makes novel contributions. I collect a unique dataset of album–level file sharing transactions from a technologically modern environment with a longer and wider panel of albums than other similar datasets. The size of the dataset facilitates the distinction between physical and digital sales, as well as a finer gradation of artist popularity. The exogenous variation used is a product of the file sharing network itself, not of user behavior, inherent characteristics of an album, or macro–level trends, and is unique in that quality. The findings of the paper thus shed new light on aspects of the sales–piracy relationship, whether these aspects have been studied extensively (*e.g.*, elasticities) or have received less attention (*e.g.*, physical–digital and popularity distinctions).

Further, this paper makes entirely new contributions to the policy debate surrounding piracy. The research noted above aims either does not disaggregate (focusing on piracy as a whole) or disaggregates by characteristics of the good in question (e.g., by album, genre, or popularity). This paper disaggregates by characteristics of the good, but it also disaggregates by the method of piracy employed. Previous work has been forced to treat all piracy as identical; I focus on private file sharing networks as distinct from public ones. The policy measures and law enforcement strategies used to specifically address private–network piracy are and must be different from those used to address piracy as a whole, but the current body of research cannot inform them appropriately. Thus, this paper focuses on the relationship between private–network piracy and album sales, paying special attention to the features that distinguish this channel of piracy from others.

The paper proceeds as follows. Section 1 describes the environment in which modern file sharing takes place, and Section 1.1 discusses the channels through which it interacts with legitimate markets. Section 1.1 also discusses the channels through which private–network piracy spills over to the larger file sharing environment. Section 2 describes the data, which are used in Section 3 to better understand how the file sharing network functions and in Section 4 to estimate the sales–piracy relationship. Section 5 offers policy suggestions in light of the paper's findings, and Section 6 offers concluding remarks.

1. The File Sharing Environment

The extralegal sharing of digital music began in earnest in 1999, when the peer–to–peer (P2P) service Napster came online. Napster, as well as other similar P2P contemporaries, enabled users to search for music in other users' libraries. The user could then download the music directly from the other users on the network. The popularity of file sharing exploded under P2P technology on networks like Napster, Gnutella, and KaZaA through the early 2000s. However, the technology was not without problems. Multiple versions of a song, of varying authenticity and quality, were available, and the user could only tell which was best by completing a download. Further, the actual download would only complete if the sharing user remained online for the duration of the transfer, and the speed of the download depended heavily on the quality of the sharing user's connection.

In the mid–2000s, the BitTorrent file sharing protocol gained popularity. In contrast with earlier P2P networks, BitTorrent uses a more diffuse method for file distribution. A user ("peer") downloads a small file that contains information about the "tracker", which is a server that facilitates peer connections. The user downloads tiny portions of the desired file (*e.g.*, music) from many different peers simultaneously, then combines these pieces together to build a complete copy of the file. The main benefits of BitTorrent come from the redundancy and decentralization of shared files inherent in the process. Everyone in the peer group obtains the same version of the file, transfers continue even if some peers leave the group, and users can preferentially connect to high–speed peers when available to increase overall transfer speeds. These benefits have been amplified by increased broadband penetration over time, which has increased connection speeds and the feasibility of "always–on" internet connections. Further, BitTorrent trackers only host the torrent files, not the actual copyrighted content, so legal challenges were made more difficult.

As with any sharing arrangement, free–riding is a chief concern. Fortunately, the programs ("clients") that use BitTorrent implement strategies to punish free–riding from peers. The most commonly seen strategy is a variation of tit–for–tat called a "choking" algorithm. Suppose our client has obtained 50% of the desired file from other peers, and it must find other clients from which to download the rest. Our client requests pieces from the other peers, which must determine whether or not to share; simultaneously, our client must determine whether or not to share the already–obtained pieces with peers who request them. A choking algorithm sends pieces to peers from which it is currently receiving other pieces (preferring fast connections over slower ones) and refuses to share to (chokes) the other connections. Thus our client will only share its pieces to peers who are sharing back, and peers who share slowly or not at all are choked off from our client. Determining which peers to choke and which to unchoke occurs once every period, typically ten seconds. Of course, such a strategy could easily deteriorate into a zero–sharing outcome, so clients also maintain an "optimistic unchoke" to which they share unconditionally. The client randomly chooses a new optimistic unchoke every three periods, which limits no–sharing punishments and also allows our client to seek out high–speed peers it had not previously connected with.[3] Different clients may employ variants and refinements on this choking algorithm, but the basic structure defines how clients share with each other.[4]

The BitTorrent protocol attempts to discourage contemporaneous free–riding through choking algorithms, but these can realistically only be applied to the peer group for one file. Dynamic free–riding can persist, in which

[3] This description is a simplified version of the one by BitTorrent's creator; see Cohen (2003).

[4] Though choking presents a workable strategy that has led to the widespread use and success of BitTorrent, it is by no means the optimal strategy. One strategically–designed client provides 70% performance against real–world clients using conventional choking by exploiting excess sharing of others (Piatek et al., 2007).

users participate in sharing only long enough to obtain a copy of the file, not bothering to reciprocate by actively sharing other files it has finished downloading. A healthy network requires at least some minimum level of active sharing without downloading ("seeding"), but seeding is hard to incentiveize. Two different kinds of file sharing networks have emerged, public and private, each with its own way of contending with free–riding. Public networks have no formal enforcement mechanism to discourage free–riding other than altruism, social norms of good behavior, and the hope of reciprocation in the future, and as such the free–riding problem persists in public networks. Private networks, on the other hand, require users to maintain certain standards of sharing to be allowed access to the network at all. Various methods of enforcement exist, all of which leverage the threat of expulsion from the network. Depending on the network, users must give in some required proportion to what they receive or remain available for sharing for some amount of time after download completion. If users do not fulfill these requirements, they are eventually banned from accessing the network. As a result, private networks tend to be smaller, have files available for longer, and maintain higher quality standards than public networks. The file sharing network studied in this paper is a private one with a minimum ratio requirement for continued network access; this requirement is detailed in Section 4.2.

BitTorrent remains the most popular file sharing protocol, accounting for more than half of total file sharing bandwidth and 2.26% of global internet traffic during 2013 (Palo Alto Networks, 2014). Dozens of active trackers exist, ranging from general interest networks to those that specialize in particular genres of music, TV, or movies. Of central interest is how these networks interact with and affect activity in legitimate markets for the good that is being shared, and the answer to this question is not clear. File sharing distributes copies of a good that might otherwise be purchased in legal markets, so these networks could displace legitimate market activity and replace it with relatively costless file sharing. However, file sharing also facilitates discovery of new goods: in legal markets, it is more costly to sample the product on offer. Thus file sharing allows consumers to verify the quality of the product before purchase, which could elicit additional sales from marginal consumers. Similarly, users who would never have purchased the product at any realistic market price may still participate in file sharing. These users can relay their newfound knowledge of the product's quality to their social connections through word–of–mouth, who may themselves purchase the product. Of course, these negative and positive effects might simply be too small to matter: consumers could have such a strong preference for shared digital goods or for purchased physical media that crowding out does not occur, while the social network of a file sharer might be composed only of other file sharers and word–of–mouth would not spread to potential customers.

Below, I describe the channels through which these possibilities could manifest in further detail. The discussion motivates the idea that the net effect of file sharing activity on music sales comprises capacity and word–of–mouth effects, and that the word–of–mouth effect should be stronger for artists whose reputations are less established.

1.1. The Sharing–Sales Relationship

Consider the market for a music album, the true quality of which is unknown to consumers. Given their beliefs about the album's quality, as well as their preferences over consuming music physically or digitally, legitimately or illicitly, consumers will decide whether and how to consume the album when it becomes available. Those that do consume learn the true quality of the album and may relay that information to their social connections: friends, colleagues, followers on social media, etc. If these peers are consumption–marginal, this information may elicit new consumption that would not otherwise have occurred. These new consumers then inform some of their own peers of the album's true quality. As the quality signal propagates through the whole social network, consumption–marginal agents may be induced to consume even though they decided not to when the album first became available. This is the essence of what I will refer to as the "word–of–mouth effect", through which the social network creates additional consumption.

The goal of the paper is to determine how additional file sharing activity will impact music consumption in this market, which could happen in two ways. The first uses the same channels as described above: if changes occur that make file sharing more attractive to marginal consumers, additional consumption occurs and the word–of–mouth effect is activated. The second way is unique to file sharing consumption: the nature of sharing is that consumers become the subsequent distributors of the music, motivated by altruism or the rules of the file sharing network. Thus early file sharers increase the capacity of file sharing networks, reducing costs and making sharing more attractive for later potential consumers. This "capacity effect" could simply elicit additional file sharing from consumers that would never have purchased otherwise, but it could also crowd out purchases from those who would otherwise have been motivated to consume legally.

The net effect of file sharing activity on music sales will thus depend on the relative magnitudes of the capacity and word–of–mouth effects. Both of these depend critically on consumer preferences and the structure of social connections. If a consumer strongly prefers one method of consumption (*i.e.* physical media, digital purchases, or illicit file sharing) over the others, then she will either consume using her preferred method or she will not consume at all.[5] This implies that the word–of–mouth effect will only increase sales, and only for those who prefer purchasing to pirating. Any sales decrease must therefore be a result of the capacity effect, which affects consumers' preferred consumption method. So the net effect will capture new sales from word–of–mouth and lost sales due to changing file sharing capacity.

Consumer uncertainty over the inherent quality of an album can enhance or mitigate these effects. If an album's true quality is much higher than expected, then the word–of–mouth effect will be quite large: expectations will change drastically and change many consumers' decisions. However, if true quality is below expectations, then word–of–mouth will have little effect: a negative signal will not induce new purchases. Thus the word–of–mouth effect should more strongly mitigate the capacity effect for unknown artists of high inherent quality than for unknown artists of low inherent quality. Put differently, file sharing should reduce uncertainty and help consumers to better recognize high–quality artists, leading to higher sales relative to well–known or low–quality artists.

Because these effects propagate across the entire social network, an exogenous change that impacts one set of consumers (one private network, say) can nevertheless affect a much larger group (*i.e.*, all file sharing networks); the word–of–mouth and capacity effects are distinct from classic wealth and substitution effects, even aggregated. Methodologically, this implies that one need not observe exogenous variation in file sharing for all consumers to draw inference. Instead, one only needs to observe exogenous variation in some set of consumers, as long as sufficient connections exist between that set and the network at large. To be clear, these two approaches are not

[5] It is reasonable to think that these preferences will depend on more than just the good's price. For example, the consumer may or may not have an MP3–capable stereo in their car, or they may have strong opinions on digital rights management (DRM) used to restrict playback of purchased digital music.

equivalent: while both are valid, each informs a different policy question. The aggregate variation approach can address the potential effects of broad–based measures, such as harsher punishment of copyright infringement, but it is ill–suited to determine the effects of shutting down an individual file sharing network. This paper addresses the latter question: what is the effect of private–network piracy on the aggregated market?

The propagation of these effects depends on an initial seed of consumption. To the econometrician, consumption due to file sharing capacity and due to high quality assessment are indistinguishable without a source of exogenous variation in file sharing capacity. Fortunately, I have access to a unique dataset comprising album sales and file sharing on a private network, which includes sources of exogenous variation in file sharing. I describe these data in Section 2.

2. Data

While most previous studies have relied on aggregate measures of file sharing, this paper analyzes the sales–piracy relationship using an album–level panel dataset of downloads and sales.[6] For the 27 weeks from July 10th to December 16th, 2008, I observe the number of albums sold in the US (both physical and digital) and illegal downloads on a private file sharing network for a variety of albums. I merge these two datasets to investigate the relationship between file sharing and legitimate sales.[7]

2.1. Album Sales Data

Data on album sales are provided by Nielsen SoundScan, which compiles US retail sales figures for music each week. Nielsen tracks sales both in physical retail locations as well as on digital platforms such as Amazon or iTunes, and distinguishes between the two in their data. For each album in the sample period, I observe the album and artist names, the number of physical and digital copies sold, the publisher, and the number of weeks

[6]The ideal dataset would be a panel of individuals in the midst of a purchase/pirate decision, but it is unlikely such data could be obtained without introducing self–reporting problems.

[7]I do not have data on pricing for albums, nor a coherent way to "price" downloads, so I work in quantities. As the market is populated by consumers with unit demand for a given album, estimating a quantity relationship should be appropriate.

since the album's release for the top 1000 selling albums each week. I report summary statistics in Table 1.

Album sales are considerably right–skewed: a small number of "superstar" albums account for the majority of sales. Overall, digital sales are a small fraction of total sales, but certain albums have a much higher digital share of sales than others. Digital sales exhibit different life–cycle patterns from their physical counterparts; I discuss these differences below.

Figure 1 shows the average album's sales as a percentage of its best week's sales over its life. A typical album's sales reach their peak in the first week of release, decay exponentially through week 10, and tend to stabilize afterward. This apparent stability may be excessive; albums whose sales drop considerably will fall off the charts and out of sample.

Figure 2 shows average album sales each week across the sample period. Sales tend to hold steady in most weeks, but tilt sharply upward in November and December due to the holiday retail season. Physical sales track this trend closely, but digital sales exhibit more variation week–to–week and are also more prominent during summer months, as can be seen in Figure 2c.

Since the data capture sales across almost 1,800 artists, it will be helpful to divide albums into tiers by some measure of artist popularity. For each week, I observe albums' ordinal ranking by copies sold (*e.g.*, the best–selling album has rank 1). Then for each artist, I determine their best album's peak rank and average album's peak rank, and use both as measures of an artist's popularity. I divide artists into quartiles according to these measures.[8] Table 2 presents these tiers and their composition. The number of albums in each tier varies, even though the number of artists in each tier is equal. Since album success can vary for a given artist, the best–rank tier system sorts considerably more albums into higher tiers than the mean–rank tier system, which is more balanced.

2.2. File Sharing Data

File sharing data are gathered from a private file sharing website.[9] During the observation period, this network acted as a tracker for over 250,000 different albums, and more than five million downloads occurred. As a private

[8] All results in the paper are qualitatively similar if two, three, four, or five tiers are used. I present results for four tiers.

[9] As a condition of data access, the name of the website has been withheld.

tracker, its users must satisfy a minimum upload/download ratio requirement or face expulsion from the network.

I report summary statistics for file sharing in Table 3. The data are significantly right–skewed: a small share of the albums account for the large majority of file sharing activity, and over 25% of albums in the file sharing data are never downloaded during the sample period. Indeed, many albums remain available long after their popularity has waned. Figure 3 depicts the average life–cycle of an album in the downloading data. The graph is constructed by calculating new downloads each week as a percentage of the highest week's downloads.[10] An album is downloaded most in the first three weeks of its being posted on the site, and activity declines steadily afterward.

Figure 4 shows downloads for an average album across the sample period. File sharing activity is fairly constant (and low), with a few exceptions: large spikes occur around weeks 15 and 20. These exceptional weeks coincide with "freeleech" periods, during which uploading improves a user's ratio but downloading does not harm it. In essence, users can download freeleech albums without penalty, but will receive credit for sharing them with others. This acts as a large positive capacity shock, incentivizing contemporaneous downloads. These downloads increase the ratio of the sharing users, who can then download more in the future without violating the ratio requirement. There are three major freeleech phases during the sample period.

Observed Freeleeches

The first freeleech phase is the New Album Contest, which occurred between September 12th and 19th (during weeks 13 and 14 of the sample period). During this seven–day period, any new file that was added to the network was granted freeleech status for 6 hours from the time it was uploaded. Small rewards were given out to those users who uploaded the most new files, including elite user status and the ability to invite others to join the network. During this period, more than 22,000 new albums were uploaded. This contest was not anticipated, and began as soon as it was announced.[11]

The second freeleech phase occurred directly after the first. The network's goal was to reach 150,000 available albums. If the users reached this goal,

[10] For Figure 3, I only include albums where I observe their first upload.

[11] Anticipation of a freeleech would make the estimation strategy problematic, since users could delay downloading until the freeleech. To allay these concerns, I explicitly test for anticipation in Section 4.2.2 and find no evidence to suggest it occurred here.

the reward would be a 24–hour freeleech on all files. Since the number of new albums far exceeded the necessary amount (about 3,000), this freeleech period was stretched into about two–and–a–half days, from September 19th through 22nd (weeks 14 and 15 of the sample period). This freeleech period was anticipated, since it was announced along with the contest. However, the original duration was set to 24 hours, so its extended length was unanticipated.

The final phase was a celebration of the network's birthday. During this period, all newly uploaded files were freeleech for the first six hours after being added to the network. This policy was in place for the three days from October 31st to November 2nd (weeks 20 and 21 of the sample period). This freeleech was not announced before it began, and so was likely unanticipated as well.[12]

Table 4 describes freeleech patterns during these periods. Relatively few albums were on freeleech during any given period, except for the site–wide phase in weeks 14 and 15. Albums in the matched sample exhibit more freeleech activity. These freeleeches impacted users' ability to download and share files contemporaneously, but they also increased future capacity by slackening users' sharing constraints. I describe a measure of this the constraint presently.

Wealth Measures: User Ratio and Buffer

Even though freeleech periods allow downloading without penalty, uploading is still credited. Thus freeleeching generates more contemporaneous downloads and also increases the future downloading potential of users who upload during the freeleech: if a user's ratio is well above her required minimum, she can download more than if her ratio were lower. I derive two measures of the slackness in the ratio requirement: a user's ratio and a user's buffer (the amount of data she could download before she hits her minimum allowed ratio). I interpret these as wealth measures and define them more precisely in Section 4.1.

The data include the mean user's ratio each week, as well as the median user's buffer each week.[13] Figure 5 plots these measures as they change

[12]The event was hinted at four days prior in a forum post by the site administrators, saying only "Shhh! Don't tell anyone, BUT, Stay Tuned for Friday...!!!" Some users suspected a freeleech, while others suspected a newly–redesigned user interface.

[13]User data were anonymized as a condition of data access, so I cannot measure indi-

during the sample period. The plot shows clear trends. During and after the freeleech periods, user wealth increases significantly, declining steadily afterward.[14] I analyze the effects of these measures on file sharing activity, alongside freeleeches, in Section 4.

2.3. Merged Panel

To investigate the interaction between file sharing and album sales, I match albums from the file sharing dataset with albums from the sales dataset. The match is not exhaustive.[15] Just over one percent of albums from the file sharing data are found in the sales data. However, this is to be expected: I only observe sales of the top 1000 albums each week, so less popular or older albums from the file sharing data will not be matched. Almost three–quarters of albums from the sales data were matched to albums in the file sharing data, however. I am confident that the unmatched albums are truly unmatchable due to the merging procedure used, which I describe below.

First, I converted album names to uppercase in the file sharing data, since album names are stored in uppercase letters in the sales data but are mixed–case in the file sharing data. This went smoothly except for a few artists or albums with nonstandard characters (*e.g.* Sigur Rós or Beyoncé). For these exceptions, downloads may be slightly under–counted due to different typesetting for different versions on the file sharing network. I then associated sales and downloads between identical artist–album–week observations in the two datasets.

In the sales data, album names are truncated after 30 characters. Artist names are also inverted (*e.g.*, "Twain, Shania" instead of "Shania Twain"), and punctuation differences may also exist between the two datasets (*e.g.*, "&" instead of "and" or "Jay–Z" instead of "Jay Z"). This sometimes results in a failure to match.

For each album in the sales data with no exact match in the file sharing data, I manually searched for an entry in the file sharing data with an alternate spelling or other discrepancy. Sometimes I was able to find a match,

vidual wealth effects.

[14]This reveals one of the main reasons why the network would implement a freeleech. If the ratio requirement is not slack enough, users will stop downloading and the network will cease to function. Freeleeches inject liquidity and sharing continues.

[15]It was, however, exhausting.

but other times I could not locate one; *e.g.*, in cases of holiday compilations, religious music, or anthologies that were either never listed on the network or were only added after the sample period. These albums make up the quarter of unmatched albums in the sales data. Given this thorough matching procedure, I believe the merged panel comprises all albums that appeared in both datasets during the sample period, with possible under–counting of downloads for a few albums.

Summary statistics of the matched albums are reported in Table 5. Albums in the merged panel still exhibit right–skewness, though less than in the individual datasets, and both sales and downloads are considerably higher than that of unmatched albums. Figure 6 shows that an album's downloads and sales track similar patterns, starting high and decaying gradually as they age.[16] Figure 7 depicts average downloads and sales across the sample period, demonstrating sales' seasonality and downloads' relative lack thereof. Large deviations in sales, similar in magnitude with contemporaneous spikes in downloading from freeleeches, are not present.

The matched albums comprise a unique, album–level dataset that can shed light on the sales–piracy relationship. In Section 3, I use these data to better understand and describe the file sharing activity that occurs on this network, how it varies over time and across albums, and how freeleech periods influence users' behavior on the network.

3. Descriptive Analysis

The rich file sharing and sales data, described above, shed light on what kinds of music users share, how their behavior changes during freeleech periods, and how sharing changes over an album's life cycle. In particular, I estimate the expected number of downloads of an album in a given week, conditional on various parameters using a fixed effects Poisson regression (Hausman et al., 1984). The results presented in this section should be read as descriptive correlations.

The characteristics of the music in question play a significant role in determining how often it is shared. Table 7 presents relative likelihoods of downloading an album by genre. The first three columns consider all albums

[16] My data include some pre–release file sharing activity, piracy that occurs before the album is commercially available. See Hammond (2014) for a study of this particular phenomenon.

in the piracy dataset, and the last three columns restrict the sample to albums matched in the sales dataset. Within each, the first column considers the effects of genre and freeleech status separately, while the second and third columns of each consider genre–freeleech interaction effects. I include a six–degree time polynomial, album age dummies, median user buffer, average filesize, and album–level fixed effects as additional controls.

We see that users on this network tend to share certain albums more than others. Active genres in the full sample include metal, decade–specific, and culture–specific (*e.g.*, Latin music or Korean pop), but changes occur during a freeleech. Jazz, spoken word, and alternative become relatively more popular on freeleech, for instance. Somewhat different patterns exist for albums in the merged sample. For instance, culture–specific and decade–specific are relatively less popular, while jazz, classical, and spoken word are relatively more popular. Of note are the results for holiday, religious, and soundtrack music. Contrary to other genres, activity markedly decreases during a freeleech. In fact, this is an artifact of observation and the timing of freeleeches: the data cover weeks through mid–December, when holiday and religious music are most popular, but freeleeches predominately occur before the holiday season. For all other genres, freeleeches boost sharing activity.

The effects of a freeleech also differ by an album's age. Table 8 compares sharing patterns across albums by age and freeleech status. The table again presents results for the full and matched samples separately. Coefficients are calculated relative to a newly–released, non–freeleech album. I include a six–degree time polynomial, genre dummies, and album–level fixed effects as additional controls.[17]

The "Baseline" specification shows that albums are shared most at release, with activity declining thereafter (mirroring the pattern in Figure 1c), and that activity significantly increases during a freeleech. The decline in activity is slower for matched albums, and the effect of a freeleech is somewhat smaller. Examining the "Age–FL Interactions" specification shows that at every age, a freeleech increases sharing activity, though the freeleech's effect reaches its peak at about two weeks after release in the matched sample. I speculate that for many new releases, activity is already high and there are relatively few potential downloaders left to be attracted by a freeleech.

[17] The "FL Dummy" regressions from Table 7 and the "Baseline" regressions from Table 8 are all from the same regression.

Freeleeches may have dynamic effects as well, through intertemporal substitution or by increasing the supply of available music after a freeleech has occurred. If users believe a freeleech will occur in the near future, they may decrease downloading now in anticipation, and an active freeleech may cause users to download today when they would have eventually done so at a later date. On the other hand, if a freeleech increased downloading and those downloaders then shared the album in the following weeks, this positive supply shock should increase activity in later weeks as album availability increases. I test for each of these effects in Table 9 using the following specification:

$$\ln(d_{it}) = \sum_{k=-3}^{3} \beta_k fl\_avgh_{it+k} + w_t + a_i + \epsilon_{it} \tag{1}$$

where d_{it} is downloads of album i in week t, fl_avgh_{it} is average freeleech hours for album i in week t, the w_t are week dummies, and the a_i are album fixed effects.[18]

The results are stark for the full sample. I find no evidence whatsoever of freeleech anticipation, but I do find a large positive concurrent effect and small positive lagging effects. I find similar patterns in the matched sample, though the test's statistical power is low. I divide the albums into artist best–rank tiers (see Table 2) and repeat the estimation, again with similar results. I conclude that there is no evidence of anticipation, whatever substitution from future consumption occurs is outweighed by the positive supply shock generated by the freeleech, and that the contemporary effects of a freeleech are the most significant for our purposes.

These results shed light on how the file sharing network under study operates, and how its users respond to the incentives presented by a freeleech. In Section 4, I draw on these insights and use the merged dataset to estimate the effect of an exogenous change in file sharing activity on legitimate album sales.

[18] I depart from Poisson regressions here and use average freeleech hours instead of a freeleech dummy to more accurately match the estimation strategy used in Section 4, so that the results may be used as a test of common trends.

4. Estimating Causal Effects

Drawing on intuition from Section 1.1, Section 4.1 proposes an empirical framework to estimate the net effect of private–network downloads on album sales. Section 4.2 motivates the instrumental variables approach used here. Section 4.3 presents estimation results.

4.1. Empirical Strategy

I estimate the parameters of the following equation:

$$\ln(s_{it}) = \alpha \ln(s_{it-1}) + \delta \ln\left(\hat{d}_{it}\right) + \mathbf{g_i}\boldsymbol{\gamma} + \tau(\boldsymbol{\omega}, t) + u_i + \epsilon_{it} \qquad (2)$$

where s_{it} are sales of album i in period t, $\hat{d}_{it}$ are exogenous downloads of album i in period t, $\mathbf{g_i}$ is a vector of genre dummies for album i, $\tau()$ is some function of time t parameterized by the vector $\boldsymbol{\omega}$, the u_i are album fixed effects, and the errors $\epsilon_{it} \sim iid(0, \sigma_\epsilon^2)$.

As Section 1.1 suggests, the model should capture the *ceteris paribus* word–of–mouth effects and capacity effects of downloading, controlling for other covariates. In equation (2), α measures the geometric decay of album sales observed in Figure 1. Through a word–of–mouth effect and market saturation, album sales begin high and decay as consumers learn of the album, make a consumption decision, and leave the market; see Figure 6. Other coefficients will therefore measure deviations from this geometric sales trend.[19]

The factor δ measures the percentage change in sales due to a contemporaneous percentage change in private–network file sharing: the elasticity of sales with respect to private–network file sharing.[20] However, a transient file sharing shock also affects later sales through $s_{i,t-1}$ and α. In percentage terms, this effect will be less than δ as the shock decays, but the corresponding effect in levels will be larger than what δ alone would imply contemporaneously. For the remainder of the analysis, I focus on δ as a measure of the

[19] An alternative specification, which uses album age indicator variables to control for an album's life–cycle patterns instead of lagged sales, yields qualitatively similar results which are available upon request.

[20] The log–transformation of sales and downloads is appropriate for two reasons. First, the capacity and word–of–mouth effects change the behavior of a share of the population, not a fixed number of consumers, so estimated effects should be scale–free. Second, sales and downloads exhibit considerable skew. The transformation greatly reduces that skew, and therefore the distribution of ϵ_{it} should be much closer to Gaussian.

contemporaneous elasticity and as an upper bound on the lifetime elasticity of album sales with respect to file sharing.

I reiterate that the file sharing network I observe is *not* representative of aggregate file sharing patterns. The network is small and private, so its behavior will likely differ from the large public networks that most file sharing occurs on. Strictly speaking, then, δ measures the effect of a private tracker's file sharing activity on aggregate album sales. File sharing activity here initiates a word–of–mouth effect that propagates across the whole social network (not just the private sharing network), and the capacity effect will be felt on public networks if downloaders share their files on these other networks.[21]

Specification (2) requires that downloads are not associated with unobserved covariates of sales. Simply including all downloads in $\hat{d}_{it}$ clearly violates this requirement. Marketing campaigns, album quality, or word–of–mouth effects from previous consumption will all influence sales and downloads alike. To obtain consistent estimates of δ, I use instrumental variables to ensure that the variation in $\hat{d}_{it}$ is due solely to shocks that do not influence album sales directly.

4.2. Instrument Validity

I propose that freeleeches and shifts in file sharing wealth measures serve as suitable instruments. Below, I specify exactly how these shifts are quantified.

Freeleeches

Freeleech status differs across albums and across time. Further, any given album could be available for downloading in different file formats, each of which could differ in their freeleech status. Table 6 lists various ways I am able to quantify freeleech activity. The binary variable *fl* equals one if the album was on freeleech at all during the week and zero otherwise, and the other measures provide more information by counting the number of freeleech hours or available formats. These measures will obviously exhibit a high degree of correlation, so not all measures can be employed simultaneously.

[21] Hammond (2014) provides some observational evidence that albums first appear on private file sharing networks, but are made available on public networks quickly thereafter. In Section 5, I determine what size of spillover is consistent with this paper's results.

It is important to understand the proposed chain of causation that comes from using freeleeches as instruments here. An exogenous freeleech on one private network will increase sharing and downloading activity on that network. Some of these files will be re–shared on other networks, including large public ones, acting as a positive supply shock for the aggregate file sharing environment. In other words, a price shock in one part of the market (*i.e.*, a freeleech on one private network) leads to a supply shock in the rest of the file sharing market (*i.e.*, increased file sharing in aggregate), and this amplified supply shock is the one that directly affects the legitimate market for music.

Wealth Measures

As discussed in Section 1, the network imposes a sharing rule on its users and bans them from the network if this requirement is not satisfied. Let UL and DL represent the total data a user has ever uploaded (shared) or downloaded (received), measured in gigabytes. I define a user's *ratio* as $\frac{UL}{DL}$ and her *buffer* as $UL - DL$. Both measure how much more a user has given than she has received, the former as a share and the latter as a quantity. The network's sharing rule is formulated as a minimum ratio requirement that varies slightly by user and can be mitigated by actively offering files to share, but generally a user's ratio should not fall below 0.6. Thus I define a user's *minimum buffer* as $UL - 0.6DL$, which is the quantity of data she can download before the ratio requirement binds. I interpret a user's ratio, buffer, and minimum buffer as measures of her wealth or ability to download. Again, these measures will exhibit some degree of correlation, but they do differ in salience and relevance. The minimum buffer is the most accurate measure of how much a user can freely download, but only the user's ratio and requirement are prominently displayed to the user. For each week in my sample, I observe the mean ratio and the medians of the buffer measures. The mean ratio and median minimum buffer are graphed across the sample period in Figure 5.

As noted above, the freeleech measures from Table 6, as well as the wealth measures (ratio and minimum buffer), are highly correlated: the first three measures in Table 6 all have correlations of 0.83 or higher among themselves, the last two measures have a correlation of 0.85 among themselves, and the two wealth measures have a correlation of 0.31. To avoid multicollinearity and over–fitting, I can only employ a few of these instruments at any given time.

These instruments are prone to a more serious problem than multicollinearity, however, and one without a particularly clean solution: the majority of their variation is secular, not cross–sectional. The wealth measures are network aggregates and do not change across albums for a given week, and the freeleech measures only exhibit minimal variation across albums: the largest freeleech affected every album on the network. Thus simple weekly dummy variables (a possible choice for $\tau(\boldsymbol{\omega}, t)$ in equation (2)) would be highly collinear with these instruments. To overcome this unfortunate characteristic of the data I have devised other measures of secular variation. Instead of finding the one measure that yields the lowest p–value on δ and trying to defend its "correctness", I will present results for all of the possible secular controls I can construct. These different measures will lead to broadly similar results, so it is less likely that the relationships are a result of spurious secular correlation.

The results that follow include six different secular controls. The first three are (1) week indicator variables, (2) a six–degree polynomial time trend, and (3) a holiday season spline. The holiday spline equals zero for weeks prior to November and counts weeks linearly thereafter to account for the holiday retail season. The indicator variables and the polynomial have the advantages of generality and require few assumptions, but are more likely to hamstring the available IVs in the dataset. The spline is defined somewhat arbitrarily, but is informed by sales patterns in Figures 2 and 7 and has fewer parameters to interfere with the available IVs. For each of these three measures, I also calculate the weekly album sales trend implied by each; these conditional sales averages make up the second three controls I employ.[22] Again, none of these measures is ideal, but the ideal measure does not exist in the dataset. Each measure leads to broadly similar conclusions.[23]

4.2.1. First–Stage Regressions

Table 10 reports first–stage estimation results. The independent variable is the logarithm of downloads, and the dependent variables are the possible

[22]The weekly sales averages will yield functionally different instruments from the simple controls themselves. Intuitively, regressing a variable on the simple controls will control for all secular variation, but regressing on the average sales will only control for secular variation in sales. This distinction is important for inference on other parameters.

[23]For some secondary tests and alternative specifications, I only report results using the holiday spline in the interest of brevity. Results for other measures are consistent with the ones presented, and are available upon request.

instruments. The first two columns give results by pooled regression, the next two include album fixed effects, and the next eight include lagged values of the logarithm of sales (which is a regressor in the second stage and therefore should be included in the instrument matrix). Results are reported for all eight different secular control schemes.

Since many of the instruments are collinear, some coefficients may be insignificant on their own. Table 10 reports p–values for tests of joint significance for all the freeleech variables taken together, for the wealth measures together, and for the full model; every test rejects the hypothesis of insignificance in the full specification. These variables explain about 40% of an album's variation in downloading across the sample period.

The effects of these variables may differ across an artist's sales tier, as defined in Table 2. Table 11 presents estimates of each variable's effects across different sales tiers, as well as joint tests by tier and tests of effect equality across tiers (*i.e., Chow tests*). Each group, freeleech and wealth, is jointly significant within each tier, but tests provide strong evidence that their effects differ across tiers (especially for freeleech variables.) I take this heterogeneity into account in my second–stage estimation in Section 4.3.

4.2.2. Freeleech Anticipation

If users of the file sharing network anticipate freeleech periods, their usefulness as exogenous instruments would be questionable. For example, increased downloading from a freeleech could simply be due to intertemporal substitution instead of file sharing that would otherwise have not occurred. The data provide no evidence for such a possibility, however. First, recall the estimation of equation (1), the results of which are given in Table 9. The estimation finds no evidence whatsoever for freeleech anticipation.

Equation (1) is formulated in the spirit of a common trends assumption test. To further allay fears, I also test for anticipation in the first–stage relationship directly. Table 12 shows first–stage regression results but with three forward lags of the freeleech variables included as additional regressors. None of these forward lags has a significant effect on file sharing activity, whether taken separately or grouped together, but the contemporaneous variables remain highly significant. Given the results of Tables 9 and 12 together, no evidence exists that users might anticipate freeleech periods.

4.3. Results

To estimate (2), I employ the two–step system GMM variety of the Arellano–Bond dynamic panel estimation technique.[24] In the full specification, I include album fixed effects, genre dummies, and the various secular controls described in Section 4.2. File sharing downloads are instrumented by the album's average hours on freeleech and the median user's minimum buffer.[25] Table 13 reports results for total album sales, Table 14 reports results for physical sales alone, and Table 15 reports results for digital sales alone. I find that a transient (*i.e.*, one–period) percentage increase in private–network file sharing results in a 0.02% decrease in physical sales, a 0.04% decrease in digital sales, and a 0.02% decrease overall. These results hold even when different measures of secular controls are used, presenting a consistent picture across specifications.

However, the small magnitude of these estimates may mask larger, opposing effects. As shown in Section 4.2, the effects of exogenous changes in the determinants of file sharing differ by artist popularity, at least as measured by the tiers defined in Table 2. It follows that exogenous changes in file sharing should also differ in their impact on album sales. To investigate, I estimate the model using tier–download interactions and report results in Table 16 for different sales categories and artist tier definitions.

Elasticity estimates are considerably larger here, though some lack statistical significance. This may be due to a true zero effect or a lack of statistical power, so I interpret these results as weak evidence. With that caveat, I find that private–network file sharing slightly decreases album sales for top–tier artists, and that the effect is larger for digital sales. There is weak evidence for a *positive* effect of private–network file sharing on sales for middle–tier artists, and no effect is consistently present for lower–tier artists. These results suggest that the capacity effect defined in Section 1.1 may be stronger for well–known artists, while the word–of–mouth effect can play a larger role for lower–ranked artists who are less well–known. Again, these results are not definitive, but they do indicate that top–tier artists are consistently, modestly harmed by file sharing.

Finally note that even though the Chow tests in Section 4.2 imply that

[24]For a full description of the estimation technique and the software package used, see Roodman (2006).

[25]Results are similar for other instrument sets. I cannot employ all measures at once due to over–fitting, as discussed in Section 4.2.

first–stage relationships vary by artist tier, I cannot use tier–specific IVs: tiers are obviously correlated with sales, and therefore tier–instrument interactions are not valid instruments. Still, the second–stage results take first–stage heterogeneity into account because instrumenting relationships are not constrained to be identical across tier–download interaction terms.

4.4. Robustness

One might be concerned that even though a freeleech is a transient occurrence, its effects may persist if word–of–mouth is slow to spread or if the within–network supply shock in the weeks following new downloading is large. The common trends estimation, given in Table 9, shows that whatever post–freeleech supply effect exists is small in comparison to the contemporary effects, but a more direct test is warranted. Table 17 estimates equation 2, but adds three lagged values of downloading as endogenous, instrumented variables. I control for secular trends using a holiday dummy, mirroring the sixth column in tables 13, 14, and 15. In all cases, the coefficient on contemporary downloads is the same as the baseline case. For overall and physical sales, no dynamic effect exists. For digital sales, a persistent effect of downloading is found, but I interpret these results with caution. The difference–in–Hansen test (reported as "H p–val" in the tables) returns a p–value close to one, which could indicate an instrument proliferation problem in this class of estimation models (Roodman, 2009). Broadly speaking, however, evidence for a persistent dynamic effect of downloads is scant.

Another concern is that even if the instruments are valid and exogenous, their effect may be contemporaneous with other trends in activity for certain kinds of music. If so, then the final estimates will systematically overweight the influence of these albums: downloads attributed to the freeleech may actually just occur because users have a systematic preference for certain kinds of music. These effects will be especially worrisome when trying to separate results across different kinds of music, such as artist popularity. One way to test for such group effects is to use a shuffling test, in which the freeleech data are randomly reassigned to different albums, new estimates are obtained from the shuffled data, and are compared to those from the original dataset.[26] I carry out two different shuffling tests on the first–stage

[26] I thank an anonymous area editor for this helpful suggestion. Shuffling tests are often used to tease out social influences by shuffling the connections in a social network; see Anagnostopoulos et al. (2008) for an example.

regression for the various freeleech measures available: one shuffles across all albums indiscriminately, and one shuffles only within artist best–rank tiers. Table 18 reports results.

First, note that the joint hypothesis that all freeleech variables are insignificant is strongly rejected in the original sample, both for pooled estimates and across tiers. For the shuffled sample, that hypothesis cannot be rejected at any reasonable significance level. The p–values from these tests are reported as "FL Joint p–value" in Table 18. Second, note that the estimates from the shuffled sample are significantly different from the original estimates: the hypothesis of equality is strongly rejected for the pooled sample and for all artist tiers. Given these results, I contend that the effects of the freeleeches observed here do not systematically affect all albums independent of each album's freeleech status, nor are there systematic effects by tiers of artist popularity that would alter the main estimates of interest.

5. Interpretation and Implications for Policy

The results provide evidence that music piracy on private networks has a statistically significant effect on legitimate music sales. However, its economic significance is less clear. An elasticity of 0.02 is quite small, but its effects could be large if amplified throughout the whole market. Hence in this section I connect my results to other findings in the literature for comparison. I also use a counterfactual exercise (eliminating the week 15 freeleech) to demonstrate the policy implications of the results.

5.1. Amplified Effects: The Link Between Private and Public File Sharing

The proposed mechanism behind the results is that additional file sharing activity on private networks (such as the one under study here) spills over and augments the sharing capacity on *all* networks, and that additional activity in aggregate affects the entire market for recorded music. I observe activity on private networks and on sales, but I cannot observe file sharing on other networks; *i.e.*, in aggregate. Other work has estimated the aggregate effects of file sharing on sales, however; the results of this paper can provide indirect evidence on the extent of the private–to–public spillover when those previous estimates are accounted for.

Table 19 lists five estimates of the elasticity of sales with respect to total piracy, which imply an average estimate of about -0.13: a 10% increase in aggregate file sharing leads to a 1.3% decrease in music sales on average.

Table 13 reports estimates of the elasticity of sales with respect to private–network piracy, with a median estimate of -0.02 across specifications. To reconcile this estimate with the literature, it must be that the elasticity of total piracy with respect to private–network piracy equals 0.15.[27] In other words, a 10% increase in private–network piracy should increase total piracy by 1.5%, which is evidence of a positive spillover if private sharing constitutes less than 15% of total sharing. Even a small private–to–public spillover can therefore reconcile effects of the size estimated in this paper with previous results in the literature.

5.2. A Counterfactual: Eliminating a Freeleech

To ground the results in concrete numbers and to better understand the importance of the results for policy makers and for those in the music industry, I use the estimates to outline a counterfactual: what if the sitewide freeleech in week 15 had not occurred?[28] First, I use the first–stage estimates from Table 13 (using the holiday spline specification) to predict downloads for each album during week 15, but with freeleech activity set to zero. The difference between this prediction and unconditionally–predicted downloads equals the number of new downloads attributable directly to the freeleech. I then use the second–stage estimates to infer the change in that week's sales that the procedure attributes to instrumented downloads for each album. Finally, I sum the predicted change in sales over all albums and used the average retail album price in 2008 to calculate the aggregate effect of the freeleech on sales and revenue in the industry. I carry out this exercise for the full sample, as well as for Tier 1 and Tier 2 albums separately.[29] Note that these calculations are fairly ad hoc; they should not be taken as rigorous econometric evidence and are simply for illustrative purposes. Table 20 reports results.

As with the data themselves, the results exhibit considerable dispersion and skew. Most albums are downloaded a few dozen times more, but a few albums experience hundreds of additional downloads. The skew is most prominent for Tier 1 albums. As a result, the change in sales are skewed as well, where some albums lose over a thousand sales while others are relatively

[27]This follows from $-0.13 \times 0.15 = -0.02$.

[28]I thank an anonymous referee for suggesting this procedure.

[29]As estimates for Tiers 3 and 4 are insignificant for overall sales, I omitted them in this exercise.

unaffected. There are considerable differences in how Tier 1 and Tier 2 albums change with the freeleech. Downloading is relatively less responsive to the freeleech for Tier 2 albums, but sales are considerably more responsive in Tier 2. Summing the pooled distribution implies that the freeleech reduced that week's sales by 94,216 albums total (4.1% of that week's 2,322,575 observed sales), while summing the tiers separately implies a sales reduction of 310,809 albums for Tier 1 and a sales increase of 539,657 albums for Tier 2, a net increase of 228,848 album sales (10.3% of that week's 2,219,644 observed Tier 1 and Tier 2 sales). Using a CD's average retail price of $12.96 in 2008[30], the pooled figures imply an aggregate revenue loss of $1.22 million; while the tiered figures imply a $4.03 million revenue loss for Tier 1 albums, a $7.00 million revenue gain for Tier 2 albums, and a $2.97 million revenue gain on net. For context, note that in the same year, the music industry reported $7.01 billion in revenue from album sales.[31]

How should one interpret these back–of–the–envelope calculations? First, note that the revenue changes are quite small when compared to the total revenue in the industry, and that events such as the freeleech can only happen a handful of times annually without significant increases in free riding on the file sharing network. A few million dollars is a large amount of money, to be sure, but other forces are likely more important determinants of industry revenue. However, the relevant figure of comparison is the cost of intervention: if file sharing reduces revenue but policies to combat file sharing are costly, are the policies worth the cost? The costs of enforcing copyright policy are difficult to quantify, but are likely considerable if the goal is to shut down a private file sharing network like the one studied here: one popular private music sharing network, OiNK, was shut down only after a two–year joint investigation by national and international industry groups, culminating in a series of raids conducted by British police, Dutch police, and Interpol (BBC, 2007). It is worth noting that following that shutdown, two new private music sharing networks were up and running within two weeks (Fisher, 2007).

[30] Figures are sourced from Friedlander (2008). The average retail price is calculated as total album equivalents divided by total revenue, where an album equivalent is one physical CD album, one digital albums, or ten digital single downloads.

[31] I should note that this procedure is necessarily limited; the only reason the numbers listed in Table 20 vary at all is because of the log–log specification of the model. Thus the results reflect a scale effect to some degree, since a unit change in the log of downloads represents a larger quantity for more popular albums.

One of these, what.cd, grew larger than OiNK had ever been, and was active for nine years before being shut down itself in a police raid and spawning at least three successor networks in the aftermath (Van der Sar, 2016).

If network shutdowns are of limited long–term effectiveness and the costs of private network file sharing are relatively small (at least as estimated here), then what other policies might be more appropriate here? A single overarching policy will likely be inadequate, since the effects of file sharing vary across albums. As noted above, the heterogeneity of effects across artist tiers is striking. The counterfactual exercises suggest that small reductions in the sale of more popular albums could be masking increases in the sale of less popular albums, since top–tier albums account for the majority of overall sales. Another source of heterogeneity is the consumption method; *i.e.*, the purchase of a physical copy or a digital download. The results in this paper consistently show that whatever effects exist, they are of a larger magnitude for digital consumption than for physical. A straightforward economic interpretation of these results is that consumers are more willing to substitute between digital purchases and digital piracy than between physical purchases and digital piracy. Effective policy must discriminate between these different effects. Below, I present some policy suggestions that account for heterogeneous effects using two broad principles: first, that deterrence efforts should concentrate on popular artists' music; and second, that digital consumption should be made easier.

To the extent that it is possible, policy makers and industry groups should focus their efforts on the piracy of top–tier artists whose reputation is already established. It is likely that piracy's effects are negative here, so targeted piracy reduction is advisable. However, I urge caution when considering efforts to shut down entire networks of file sharing, since this could have an adverse effect on mid–tier artists' sales. Agencies might decide not to pursue cases unless top–tier artists' music has been pirated, or might even consider releasing up–and–coming artists on file sharing networks themselves alongside a more traditional retail launch. Similar strategies have been used occasionally by high–profile artists: Radiohead released "In Rainbows" online using a pay–what–you–want scheme, and Nine Inch Nails released "Ghosts I–IV" under a copyright license that specifically permits free non–commercial distribution. This paper's results suggest that such schemes are not likely to lead to commercial success for top–tier artists once their novelty has worn off, but that it might be helpful for artists early in their careers.

Digital consumption of music should also be made easier and more convenient to elicit substitution away from piracy. The widespread advent of streaming music services helps here, as it allows access to millions of songs digitally for a fee. However, industry groups and record labels should be cautious about fragmentation. As a record label or artist, it may be tempting to consolidate and start a distribution platform of their own (*e.g.*, Jay-Z's TIDAL service or Garth Brooks' GhostTunes), or to make their music available only on a few streaming platforms (*e.g.*, Taylor Swift's absence from the streaming service Spotify from 2014—17), but streaming platforms are not only competing with each other for listeners. When faced with a choice between a fragmented market requiring multiple subscriptions or using a single unified file sharing network where all content is available, users may opt for simplicity. Still, copyright holders are on one side of a two–sided market when deciding which streaming platforms to license to, and committing to licensing to all platforms reduces a record label's bargaining power. Belleflamme (2016) notes that in two–sided markets where one side multihomes (here, copyright holders license to multiple platforms) and one side singlehomes (consumers typically use just one platform), the "platform is in a monopoly position on the multihoming side." In such a model, file sharing networks compete with streaming platforms, but the networks do not have to license content and users do not have to pay a subscription. Since consumers prefer to singlehome, a copyright holder that licenses exclusively to one streaming platform is at greatest risk of losing customers to piracy. One solution might be to grant licenses rights through umbrella industry groups like the RIAA and demand that streaming platforms license to everyone under that umbrella, monopolizing the supply side of the platform and increasing the likelihood that consumers will singlehome with a streaming platform rather than a file sharing network. This monopolization also allows the copyright holders to charge higher licensing fees for less popular artists that the streaming services would otherwise reject. If file sharing increases demand for these artists' music, as the results of this paper suggest, price discrimination in licensing royalties can be structured to amplify the positive effects of file sharing and mitigate the negative ones.

6. Conclusions

Private file sharing networks present unique concerns when studying the relationship between piracy and legitimate sales. These private networks facilitate a small share of total file sharing, but their reliability, quality standards, and policing of free riding allows them to punch above their weight. Law enforcement agencies and industry groups exert considerable effort to curtail and eliminate private file sharing networks, but there is relatively little available evidence on how these networks affect the legitimate market for music. This paper combines a unique, album–level dataset obtained from one of the most active private music file sharing networks at the time, combines it with retail sales data, and uses exogenous variation in the sharing capacity of users on the network to understand how file sharing occurs in this environment and how it might affect the legitimate market.

I find that private file–sharing networks behave much like a market; users indeed respond to incentives akin to price and wealth shocks. During periods in which users are credited for sharing music but not debited for receiving music (a freeleech), users increase downloading activity significantly as would be expected if a good's price was drastically reduced. After such periods, when users have accumulated excess sharing credit and have more content on offer to share, activity is significantly higher than baseline and the effect decays over time as users spend down their wealth.

I also find that private network file sharing has a statistically significant effect on the retail market for music. Additional file sharing activity has a negative effect on album sales, and the effect is larger for digital sales than for physical sales. I attribute the difference to a higher willingness to substitute digital piracy for digital purchases, rather than physical purchases. However, the effects differ for artists of different quality and popularity. Top–tier artists' sales decrease, while mid–tier artists' sales actually increase with additional file sharing activity. The magnitude of these effects is larger for digital sales, as before. These results are consistent with the combination of a capacity effect, in which file sharing activity subsequently increases users' supply of shareable files and a substitution away from retail markets, and a word–of–mouth effect, in which file sharing increases awareness, decreases quality uncertainty, and therefore increases demand in all markets. Thus for artists whose reputation is already established (*i.e.*, top–tier artists), the capacity effect dominates the word–of–mouth effect and the overall change is negative, while up–and–coming artists stand to gain more from increased word–of–mouth. These hypothesized effects closely mirror the direct substitution and sampling effects already documented in the literature, but the capacity and word–of–mouth effects proposed here are the culmination of a chain of effects that propogates as files are shared across the mesh of public and private file sharing networks. Because of the infinitely replicable nature of digital goods, changes in one small corner of the file sharing market can influence the market as a whole. The results of this paper are consistent with a private–to–public spillover elasticity of 0.15, a relatively inelastic spillover response.

It is interesting to note how the effects found here parallel those found in studies of digital streaming services for music and for piracy of other media. For instance, Hiller (2016), using data that are contemporary to this paper's, finds that a music video's presence on YouTube may reduce album sales for chart–topping albums, but that a "promotional" effect similar to the word–of–mouth effect here may moderate reductions to sales for lower–ranked albums. Kretschmer and Peukert (2014) also find evidence of countervailing "promotional" and "displacement" effects in a similar market. And Peukert et al. (2017) find that a negative capacity shock for pirated movies led to an increase in box office revenues for movies with a wide release but a decrease in revenues for more limited–release movies, reflecting the greater importance of word–of–mouth for these less–popular offerings. These conclusions are paralleled in the results of this paper: increased file sharing capacity lowers sales more for top–tier artists, but increased word–of–mouth may offset such decreases for less–popular artists.

I present a few suggestions for policy makers and industry groups to consider as they address file sharing's effects on the legitimate market, guided by two general principles: first, any attempts at file sharing deterrence should focus on top–tier artists' music; second, digital consumption should be made easier to facilitate substitution away from file sharing. I caution against attempts to destroy whole file sharing networks, as the benefits of doing so could also destroy positive effects of file sharing and may not be worth the cost. It is worth experimenting with providing albums directly to the file sharing networks themselves to see if the benefits of file sharing could be directly harnessed. Finally, a fracturing of the market for music should be avoided if at all possible. Consumers are hesitant to use more than one music platform, and file sharing can be viewed as a music platform where all content is available and the only costs are non–monetary. Music copyright holders should consider joint licenses to music platforms instead of exclusive arrangements, which increases their bargaining power vis–à–vis the music platforms', and can facilitate royalty fee structures to amplify the benefits of file sharing and mitigate the costs.

Piracy affects markets for all media that can be digitized, and the mechanisms proposed here are not unique to the market for music. Structurally similar file sharing networks exist for film, television, academic textbooks, and more. Further, these markets all rely on the perceived quality of the good as spread between customers through word of mouth, so the capacity and word–of–mouth effects could be important factors in these markets as well. Further work is needed to quantify the nature of the sharing–sales relationship in each of these markets, but this paper provides insight into what economic forces might be at play, how they might break down across different classes of goods, and the economic significance of their effects. If similar results hold in these markets, the lessons learned by policy makers and industry groups could be applied to these markets as well.

Anagnostopoulos, A., Kumar, R., Mahdian, M., 2008. Influence and correlation in social networks. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, pp. 7–15.

Andersen, B., Frenz, M., 2010. Don't blame the P2P file-sharers: The impact of free music downloads on the purchase of music CDs in Canada. Journal of Evolutionary Economics 20 (5), 715–740.

BBC, 2007. Huge pirate music site shut down.
URL http://news.bbc.co.uk/2/hi/uk_news/england/tees/7057812.stm

Belleflamme, P., 2016. The economics of digital goods: A progress report. Review of Economic Research on Copyright Issues 13, 1–24.

Blackburn, D., 2006. The heterogeneous effects of copying: The case of recorded music, unpublished Manuscript. Available at http://www.davidjhblackburn.com/papers/blackburn_fas.pdf.

Cohen, B., 2003. Incentives build robustness in BitTorrent. In: Workshop on Economics of Peer-to-Peer systems. Vol. 6. pp. 68–72.

Connolly, M., Krueger, A. B., 2006. Rockonomics: The economics of popular music. Handbook of the Economics of Art and Culture 1, 667–719.

Danaher, B., Smith, M. D., Telang, R., Chen, S., 2014. The effect of graduated response anti-piracy laws on music sales: Evidence from an event study in France. The Journal of Industrial Economics 62 (3), 541–553.

Dejean, S., 2009. What can we learn from empirical studies about piracy? CESifo Economic Studies 55 (2), 326–352.

Fisher, K., 2007. Oink's new piglets proof positive that big content's efforts often backfire.
URL http://arstechnica.com/?post_type=post&p=79091

Friedlander, J. P., 2008. RIAA 2008 year–end shipment statistics.

Gopal, R. D., Bhattacharjee, S., Sanders, G. L., 2006. Do artists benefit from online music sharing? The Journal of Business 79 (3), 1503–1533.

Hammond, R. G., 2014. Profit leak? Pre-release file sharing and the music industry. Southern Economic Journal 81 (2), 387–408.

Hausman, J., Hall, B. H., Griliches, Z., 1984. Econometric models for count data with an application to the Patents—R & D relationship. Econometrica: Journal of the Econometric Society, 909–938.

Hiller, R. S., 2016. Sales displacement and streaming music: Evidence from YouTube. Information Economics and Policy 34, 16–26.

Hong, S.-H., 2013. Measuring the effect of Napster on recorded music sales: Difference–in–differences estimates under compositional changes. Journal of Applied Econometrics 28 (2), 297–324.

Hui, K.-L., Png, I., 2003. Piracy and the legitimate demand for recorded music. Contributions in Economic Analysis & Policy 2 (1).

Kretschmer, T., Peukert, C., 2014. Video killed the radio star? Online music videos and digital music sales. Centre for Economic Performance, London School of Economics and Political Science.

Lee, J., 2009. Private filesharing markets and album sales, undegraduate Honor's Thesis, Vanderbilt University Department of Economics. Available at http://as.vanderbilt.edu/econ/documents/honor09Lee.pdf.

Leung, T. C., 2008. Should the music industry sue its own customers? Impacts of music piracy and policy suggestions. University of Minnesota, retrieved Nov 28, 2012.

Liebowitz, S. J., 2005a. Economists' topsy–turvy view of piracy. Review of Economic Research on Copyright Issues 1, 5–17.

Liebowitz, S. J., 2005b. Pitfalls in measuring the impact of file–sharing on the sound recording market. CESifo Economic Studies 51 (2-3), 435–473.

Liebowitz, S. J., 2006a. Economists examine file–sharing and music sales. Industrial Organization And The Digital Economy, 145–174.

Liebowitz, S. J., 2006b. File sharing: Creative destruction or just plain destruction? Journal of Law and Economics 49 (1), 1–28.

Liebowitz, S. J., 2008. Research note: Testing file sharing's impact on music album sales in cities. Management Science 54 (4), 852–859.

Mortimer, J. H., Nosko, C., Sorensen, A., 2012. Supply responses to digital distribution: Recorded music and live performances. Information Economics and Policy 24 (1), 3–14.

Oberholzer-Gee, F., Strumpf, K., 2007. The effect of file sharing on record sales: An empirical analysis. Journal of Political Economy 115 (1), 1–42.

Oberholzer-Gee, F., Strumpf, K., 2010. File sharing and copyright. In: Innovation Policy and the Economy, Volume 10. University of Chicago Press, pp. 19–55.

Palo Alto Networks, June 2014. Application usage and threat report. Technical Report.

Peitz, M., Waelbroeck, P., 2004. The effect of internet piracy on cd sales: Cross-section evidence. Tech. rep., CESifo Working Paper.

Peitz, M., Waelbroeck, P., 2006a. Piracy of digital products: A critical review of the theoretical literature. Information Economics and Policy 18 (4), 449–476.

Peitz, M., Waelbroeck, P., 2006b. Why the music industry may gain from free downloading — the role of sampling. International Journal of Industrial Organization 24 (5), 907–913.

Peukert, C., Claussen, J., Kretschmer, T., 2017. Piracy and box office movie revenues: Evidence from megaupload. International Journal of Industrial Organization 52, 188–215.

Piatek, M., Isdal, T., Anderson, T., Krishnamurthy, A., Venkataramani, A., 2007. Do incentives build robustness in BitTorrent. In: 4th USENIX Symposium on Networked Systems Design & Implementation. pp. 1–14.

Rob, R., Waldfogel, J., April 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.

Roodman, D., 2006. How to do xtabond2: An introduction to difference and system GMM in Stata. Center for Global Development Working Paper (103).

Roodman, D., 2009. A note on the theme of too many instruments. Oxford Bulletin of Economics and statistics 71 (1), 135–158.

Van der Sar, E., 2016. What.cd is dead, but the torrent hydra lives on.
URL https://torrentfreak.com/what-cd-is-dead-but-the-torrent-hydra-lives-on-16

Waldfogel, J., 2010. Music file sharing and sales displacement in the iTunes era. Information economics and policy 22 (4), 306–314.

Waldfogel, J., 2011. Bye, bye, miss American pie? The supply of new recorded music since Napster. Tech. rep., National Bureau of Economic Research.

Waldfogel, J., 2012. Copyright protection, technological change, and the quality of new products: Evidence from recorded music since napster. The journal of law and economics 55 (4), 715–740.

Zentner, A., 2006. Measuring the effect of file sharing on music purchases. Journal of Law and Economics 49 (1), 63–90.

Tables

Table 1: Summary Statistics for Sales

Statistic	Mean	Percentiles 25th	Percentiles 50th	Percentiles 75th	Skew
Total Sales	45,596	2,714	11,427	38,687	9.4
Log Total Sales	9.3	7.9	9.3	11	.27
Weekly Sales	4,552	1,202	1,667	2,771	12
Log Weekly Sales	7.7	7.1	7.4	7.9	2.2
Digital Share	19%	2.2%	9.4%	24%	2
Weeks on Chart	149	6	30	177	2.4
Number Sales	131,087,702				
Mean Weekly Sales	195,910				
Albums	2,875				
Artists	1,793				

Table 2: Division of Albums into Sales Tiers

	Best–Rank Tiers			Mean–Rank Tiers		
Tier	Best Rank	Albums	Share	Mean Rank	Albums	Share
1	1–103	1,113	39%	1–202	536	19%
2	104–329	698	24%	204–426	976	34%
3	330–639	573	20%	427–663	820	29%
4	640 +	491	17%	664 +	543	19%

Table 3: Summary Statistics for Downloads

Statistic	Mean	Percentiles			Skew
		25th	50th	75th	
Total DLs	26	3	7	19	30
Log Total DLs	2	1.1	1.9	2.9	.5
Weekly DLs	1.3	.15	.37	.89	81
Log Weekly DLs	-.95	-1.9	-.99	-.12	.46
Number of DLs	5,099,397				
Mean Weekly DLs	195,910				
Albums	264,672				
Active Albums	195,164				

Statistics are calculated for albums with at least one download.

Table 4: Albums on Freeleech by Sample

Week	Full Sample		Merged Sample	
	Share	Avg. Hours	Share	Avg. Hours
13	0.6%	6.1	0.3%	7
14	100%	19.5	100%	52.3
15	100%	80.6	100%	222.4
20	0.8%	7.2	2.0%	6.3
21	2.1%	7.4	4.6%	7.9
Others	None			

The full sample includes all albums available on the network, while the matched sample includes only albums with successful matches in the retail sales data.

Table 5: Summary Statistics for Matched Albums

Statistic	**Mean**	**Percentiles** 25th	50th	75th	**Skew**
Total DLs	360	36	126	369	5.8
Log Total DLs	4.8	3.7	4.9	5.9	-.28
Total Sales	57,085	3,591	17,593	49,386	8.3
Log TotalSales	9.6	8.2	9.8	11	.1
Number of DLs	773,787				
Number of Sales	122,675,525				
Albums	2,109				
Match Rate, DLs	1.1%				
Match Rate, Sales	73.4%				

Table 6: Various Freeleech Measures

Variable	Description
fl	Freeleech dummy
fl_pct	Share of formats on freeleech
fl_avgh	Average format–hours on freeleech
fl_sum	Number of formats on freeleech
fl_h	Number of format–hours on freeleech

Table 7: Relative Likelihood of Weekly Downloads by Genre

	Full Sample			Matched Albums		
	FL Dummy	Genre–FL Interactions		FL Dummy	Genre–FL Interactions	
Genre		No FL	FL		No FL	FL
Decades (eg: 80s)	1.21***	1.05***	3.22***	1.00	0.82***	2.51***
	(0.00)	(0.00)	(0.00)	(0.98)	(0.00)	(0.00)
Alternative	1.12***	1.01**	2.50***	0.80***	0.74***	1.29***
	(0.00)	(0.02)	(0.00)	(0.00)	(0.00)	(0.00)
Rap, Hip-Hop	0.86***	0.76***	1.57***	0.83***	0.77***	1.11***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
R&B, Soul	0.92***	1.01	0.88***	0.73***	0.73***	0.94**
	(0.00)	(0.45)	(0.00)	(0.00)	(0.00)	(0.04)
Rock	0.90***	0.90***	1.10***	0.89***	0.85***	1.12***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Pop, Dance	1.11***	1.09***	1.19***	1.14***	1.09***	1.48***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Country	0.89***	0.80***	2.03***	0.85***	0.69***	1.58***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Religious	0.68***	0.72***	0.19***	0.56***	0.60***	0.11***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Holiday	16.73***	16.69***	2.87***	13.43***	12.89***	3.54***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Classical	0.81***	0.51***	1.32***	1.34***	0.93*	2.41***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.07)	(0.00)
Soundtrack	0.78***	0.77***	0.65***	0.60***	0.72***	0.57***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Spoken Word	1.02	0.93	3.04***	1.68***	1.39**	6.21***
	(0.81)	(0.39)	(0.00)	(0.00)	(0.04)	(0.00)
Metal	1.44***	1.34***	3.49***	1.71***	1.59***	2.75***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Electronic	1.01	0.85***	1.82***	1.05***	1.04***	1.41***
	(0.12)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Jazz, Blues	1.00	0.88***	2.10***	1.19***	1.12***	1.70***
	(0.84)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Cultural	1.67***	1.47***	2.65***	0.71***	0.62***	0.67***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Children	1.22**	1.03	0.94	0.96	0.83	1.26
	(0.04)	(0.78)	(0.71)	(0.76)	(0.14)	(0.76)
Ambient	0.98***	0.93***	1.37***	0.94***	0.94***	0.75***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Time Trend	Yes	Yes		Yes	Yes	
FL Dummy	Yes	No		Yes	No	
Age Controls	Yes	Yes		Yes	Yes	
User Buffer	Yes	Yes		Yes	Yes	
Fizesize	Yes	Yes		Yes	Yes	
AIC	1.05×10^7	1.09×10^7		780350.63	807732.56	
Observations	4,266,217	4,266,217		47,291	47,291	
Groups	194,200	194,200		2,105	2,105	

Exponentiated coefficients; p-values in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 8: Relative Likelihood of Downloads by Freeleech Status and Age

	Full Sample			Matched Albums		
	Baseline	Age-FL Interactions		Baseline	Age-FL Interactions	
Album Age		No FL	FL		No FL	FL
Release Week	1	1	6.39***	1	1	1.12***
	(—)	(—)	(0.00)	(—)	(—)	(0.00)
Age = 1	0.52***	0.80***	1.04***	0.93***	0.98***	0.77***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 2	0.58***	0.70***	2.03***	0.76***	0.70***	1.57***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 3	0.41***	0.48***	1.26***	0.56***	0.53***	1.07***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 4	0.35***	0.40***	1.15***	0.43***	0.39***	0.82***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 5	0.32***	0.36***	1.10***	0.35***	0.32***	0.73***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 6	0.30***	0.34***	0.96***	0.34***	0.32***	0.52***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 7	0.27***	0.29***	0.92***	0.28***	0.24***	0.48***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 8	0.32***	0.35***	0.93***	0.29***	0.26***	0.40***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 9	0.28***	0.31***	0.95***	0.26***	0.23***	0.51***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age = 10	0.24***	0.25***	0.97***	0.26***	0.22***	0.53***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)
Age > 10	0.20***	0.20***	1.04***	0.18***	0.15***	0.63***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
FL Dummy	4.60***			3.25***		
	(0.00)			(0.00)		
Size in MB	1.00***	1.00***		1.00***	1.00***	
	(0.00)	(0.00)		(0.00)	(0.00)	
Time Trend	Yes	Yes		Yes	Yes	
Genre Controls	Yes	Yes		Yes	Yes	
AIC	1.05×10^7	1.05×10^7		780350.63	758759.40	
Observations	4,266,217	4,266,217		47,291	47,291	
Groups	194,200	194,200		2,105	2,105	

Exponentiated coefficients; p-values in parentheses

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 9: Leading and Lagging Effects of Freeleech Activity on Downloading

	Full Sample	Matched Sample				
			Artist Best–Rank Tiers			
	All Albums	All Albums	Tier 1 Albums	Tier 2 Albums	Tier 3 Albums	Tier 4 Albums
Avg. FL Hours, t+3	0.00	0.05	0.05	0.05	0.05	0.06
	(0.97)	(0.50)	(0.48)	(0.49)	(0.47)	(0.45)
Avg. FL Hours, t+2	-0.01	0.01	0.00	0.00	0.00	0.01
	(0.32)	(0.94)	(0.98)	(0.99)	(0.98)	(0.94)
Avg. FL Hours, t+1	-0.00	0.01	0.01	0.01	0.01	0.01
	(0.75)	(0.87)	(0.89)	(0.91)	(0.88)	(0.92)
Avg. FL Hours, t	0.28^{***}	0.13^{+}	0.12^{+}	0.12^{+}	0.12^{+}	0.13^{+}
	(0.00)	(0.11)	(0.12)	(0.12)	(0.13)	(0.10)
Avg. FL Hours, t-1	0.08^{***}	-0.10	-0.10^{+}	-0.10^{+}	-0.10^{+}	-0.10^{+}
	(0.00)	(0.20)	(0.19)	(0.19)	(0.19)	(0.18)
Avg. FL Hours, t-2	0.02^{***}	0.07^{*}	0.07^{*}	0.07^{*}	0.07^{*}	0.07^{*}
	(0.00)	(0.07)	(0.07)	(0.08)	(0.07)	(0.05)
Avg. FL Hours, t-3	0.01^{***}	0.05^{**}	0.05^{**}	0.05^{*}	0.05^{**}	0.05^{**}
	(0.00)	(0.05)	(0.05)	(0.06)	(0.05)	(0.04)
Week Controls	Yes	Yes	Yes			
Observations	993,174	13,236	13,236			
Groups	162,131	1,383	1,383			

p-values in parentheses
$^{+}$ $p < .2$, * $p < .1$, ** $p < .05$, *** $p < .01$

Table 10: Estimation of First–Stage Relationship

	OLS		Panel FE					Panel FE, Secular Means		
fl	1.21***	1.09***	0.26**	0.14	0.11	0.17	0.23**	0.25**	0.25**	0.23**
	(0.00)	(0.00)	(0.03)	(0.39)	(0.48)	(0.14)	(0.05)	(0.03)	(0.03)	(0.04)
fl_pct	-2.65***	-2.54***	-0.86***	0.13	0.34	-0.77***	-0.78***	-0.78***	-0.78***	-0.78***
	(0.00)	(0.00)	(0.00)	(0.87)	(0.65)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
fl_avgh	0.04***	0.04***	0.04***	0.05	0.00	0.04***	0.04***	0.04***	0.04***	0.04***
	(0.00)	(0.00)	(0.00)	(0.52)	(0.98)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
fl_sum	0.14***	0.14***	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
	(0.00)	(0.00)	(0.16)	(0.19)	(0.11)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)
fl_h	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Ratio		-0.53***	-0.03	0.83	4.86***	5.83***	-0.23***	-0.15***	-0.15***	-0.20***
		(0.00)	(0.56)	(0.57)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Buffer		0.05***	0.01***	-0.41	-1.86***	-0.55***	0.05***	0.04***	0.04***	0.04***
		(0.00)	(0.01)	(0.49)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Genre Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Sales	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Week Controls	No	No	No	Yes	Yes	No	No	Yes	No	No
Time Trend	No	No	No	No	No	Yes	No	No	Yes	No
Holiday Trend	No	No	No	No	No	No	Yes	No	No	Yes
FL Joint p-value	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Wealth Joint p-value		0.00	0.03	0.78	0.00	0.00	0.00	0.00	0.00	0.00
Model F-stat	395.41	302.93	1,112.88	178.72	237.64	345.23	396.56	396.29	396.27	396.28
Within-R^2	0.11	0.11	0.34	0.37	0.44	0.42	0.41	0.41	0.41	0.41
Observations	16,800	16,800	16,800	16,800	16,800	16,800	16,800	16,800	16,800	16,800
Albums			1,589	1,589	1,589	1,589	1,589	1,589	1,589	1,589

p-values in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 11: Differences in First–Stage Relationships Across Tiers

	Artist's Best Tier					Artist's Mean Tier				
	Tier 1	Tier 2	Tier 3	Tier 4	Equality	Tier 1	Tier 2	Tier 3	Tier 4	Equality
fl	0.180	0.284	0.843	-1.060*	0.158	0.351	0.158	0.212	-0.389	0.164
	(0.205)	(0.153)	(0.376)	(0.072)		(0.111)	(0.364)	(0.319)	(0.138)	
fl_sum	-0.021**	0.005	-0.027	0.029	0.299	-0.013	-0.020	-0.001	-0.034	0.550
	(0.038)	(0.640)	(0.210)	(0.769)		(0.482)	(0.121)	(0.936)	(0.227)	
fl_pct	-0.647***	-0.974***	-1.260	0.000	0.424	-0.924***	-0.626***	-0.842***	0.000	0.598
	(0.000)	(0.000)	(0.192)	—		(0.000)	(0.001)	(0.000)	—	
fl_h	0.001***	0.001***	0.002***	0.000	0.221	0.001***	0.002***	0.001***	0.001*	0.309
	(0.000)	(0.001)	(0.000)	(0.954)		(0.002)	(0.000)	(0.001)	(0.050)	
fl_avgh	0.041**	0.046***	0.035***	0.059***	0.035	0.041***	0.040***	0.047***	0.044***	0.101
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)	
Ratio	-0.159**	-0.274**	-0.289**	-0.635**	0.309	-0.196**	-0.138*	-0.329***	-0.357	0.373
	(0.029)	(0.003)	(0.038)	(0.027)		(0.045)	(0.081)	(0.001)	(0.108)	
Buffer	0.041***	0.051***	0.041***	0.066***	0.153	0.034***	0.045***	0.053***	0.046***	0.046
	(0.000)	(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)	(0.000)	
Joint FL	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.000	0.000
Joint Buffer	0.000	0.000	0.000	0.001	0.301	0.000	0.000	0.000	0.001	0.089

p-values in parentheses: * $p < .1$, ** $p < .05$, *** $p < .01$

Joint p–values are from a test of group significance within a tier.

Chow p–values are from a test of equality across tiers.

Table 12: Test of User Anticipation of Freeleeches

	Freeleech or Wealth Measure						
Timing	*fl*	*fl_pct*	*fl_avgh*	*fl_sum*	*fl_h*	Ratio	Buffer
t	0.11	0.11	0.03**	-0.02	0.00	-0.23***	0.05***
	(0.55)	(0.87)	(0.01)	(0.77)	(0.29)	(0.00)	(0.00)
t+1	-0.07	-0.26	-0.01	0.09	-0.00		
	(0.72)	(0.79)	(0.80)	(0.31)	(0.51)		
t+2	-0.10	-0.25	0.01	0.10	-0.00		
	(0.59)	(0.77)	(0.52)	(0.26)	(0.19)		
t+3	-0.01	-0.14	0.01	0.04*	-0.00		
	(0.92)	(0.54)	(0.67)	(0.09)	(0.28)		
Genre Controls			Yes				
Holiday Trend			Yes				
Lagged Sales			Yes				
FL Joint p-value			0.00				
FL t+1 Joint p-value			0.63				
FL t+2 Joint p-value			0.87				
FL t+3 Joint p-value			0.14				
Model F-stat			245.17				
Within-R^2			0.44				
Observations			14,915				
Albums			1,514				

p-values in parentheses: * $p<.1$, ** $p<.05$, *** $p<.01$

Joint p–values are from a test of group significance within a tier.

Table 13: Estimation of (2) for Overall Sales.

	OLS	IV	Panel IV	Dyn. PIV, Direct Secular Controls			Dyn. PIV, Secular Mean Controls		
DLs	0.01***	-0.01**	-0.01***	-0.01**	-0.02	-0.02***	-0.02***	-0.00	-0.02***
	(0.00)	(0.02)	(0.00)	(0.02)	(0.69)	(0.00)	(0.00)	(0.46)	(0.00)
Lag Sales	0.90***	0.91***	0.69***	0.88***	0.90***	0.90***	0.89***	0.88***	0.90***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Genre Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week Polynomial	Yes	Yes	Yes	Yes	No	No	Yes	No	No
Week Controls	No	No	No	No	Yes	No	No	Yes	No
Holiday Trend	No	No	No	No	No	Yes	No	No	Yes
Total IVs		2	2	27	24	27	27	27	27
H p-val		0.26	0.81	0.01	0.11	0.37	0.74	0.78	0.33
Observations	16,800	16,800	16,588	16,800	16,800	16,800	16,800	16,800	16,800
Albums			1,377	1,589	1,589	1,589	1,589	1,589	1,589

p-values in parentheses: * $p < .1$, ** $p < .05$, *** $p < .01$

Instruments used: *fl_avgh*, buffer, and those derived from lagged values of the dependent variable.

"H p–val" is the difference–in–Hansen test p–values from excluding *fl_avgh* and buffer from the instrument set.

Table 14: Estimation of (2) for Physical Sales.

	OLS	IV	Panel IV	Dyn. PIV, Direct Secular Controls			Dyn. PIV, Secular Mean Controls		
DLs	0.01***	-0.00	-0.01**	-0.01	-0.01	-0.02***	-0.01*	0.00	-0.02***
	(0.00)	(0.64)	(0.01)	(0.25)	(0.73)	(0.00)	(0.08)	(0.59)	(0.00)
Lag Sales	0.90***	0.91***	0.67***	0.88***	0.90***	0.90***	0.88***	0.88***	0.89***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Genre Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week Polynomial	Yes	Yes	Yes	Yes	No	No	Yes	No	No
Week Controls	No	No	No	No	Yes	No	No	Yes	No
Holiday Trend	No	No	No	No	No	Yes	No	No	Yes
Total IVs		2	2	27	24	27	27	27	27
H p-val		0.97	0.05	0.16	0.17	0.97	0.79	0.45	0.80
Observations	16,787	16,787	16,580	16,787	16,787	16,787	16,787	16,787	16,787
Albums			1,375	1,582	1,582	1,582	1,582	1,582	1,582

p-values in parentheses: * $p < .1$, ** $p < .05$, *** $p < .01$

Instruments used: *fl_avgh*, buffer, and those derived from lagged values of the dependent variable.

"H p–val" is the difference–in–Hansen test p–values from excluding *fl_avgh* and buffer from the instrument set.

Table 15: Estimation of (2) for Digital Sales.

	OLS	IV	Panel IV	Dyn. PIV, Direct Secular Controls			Dyn. PIV, Secular Mean Controls		
DLs	0.03***	-0.03***	-0.02***	-0.03***	-0.06	-0.04***	-0.04***	-0.04***	-0.04***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.38)	(0.00)	(0.00)	(0.00)	(0.00)
Lag Sales	0.89***	0.92***	0.67***	0.84***	0.87***	0.85***	0.79***	0.80***	0.85***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Genre Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week Polynomial	Yes	Yes	Yes	Yes	No	No	Yes	No	No
Week Controls	No	No	No	No	Yes	No	No	Yes	No
Holiday Trend	No	No	No	No	No	Yes	No	No	Yes
Total IVs		2	2	27	24	27	27	27	27
H p-val		0.00	0.00	0.01	0.01	0.05	0.16	0.05	0.04
Observations	15,537	15,537	15,343	15,537	15,537	15,537	15,537	15,537	15,537
Albums			1,270	1,464	1,464	1,464	1,464	1,464	1,464

p-values in parentheses: * $p < .1$, ** $p < .05$, *** $p < .01$

Instruments used: *fl_avgh*, buffer, and those derived from lagged values of the dependent variable.

"H p–val" is the difference–in–Hansen test p–values from excluding *fl_avgh* and buffer from the instrument set.

Table 16: Estimation of (2) with Artist Tiers.

	Best Rank Tiers			Mean Rank Tiers		
	Overall	Physical	Digital	Overall	Physical	Digital
Tier 1 DLs	-0.12**	-0.12**	-0.15***	-0.08***	-0.05*	-0.10**
	(0.02)	(0.02)	(0.00)	(0.00)	(0.05)	(0.02)
Tier 2 DLs	0.31**	0.26*	0.50**	0.06	0.08**	0.17
	(0.05)	(0.09)	(0.01)	(0.19)	(0.03)	(0.16)
Tier 3 DLs	-0.13	-0.04	-1.02***	-0.05	-0.13	-0.33*
	(0.60)	(0.85)	(0.00)	(0.54)	(0.12)	(0.08)
Tier 4 DLs	-0.57	-0.70	0.92	-0.06	0.09	0.42
	(0.40)	(0.26)	(0.15)	(0.68)	(0.63)	(0.15)
Lag Sales	0.88***	0.89***	0.82***	0.93***	0.92***	0.86***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Total IVs	48	48	48	48	48	48
H p-val	0.22	0.21	0.05	0.47	0.27	0.04
Observations	16,800	16,787	15,537	16,800	16,787	15,537
Albums	1,589	1,582	1,464	1,589	1,582	1,464

p-values in parentheses: * $p < .1$, ** $p < .05$, *** $p < .01$

IVs: *fl_avgh*, buffer, and those derived from lagged values of the dependent variable.

"H p–val" is the difference–in–Hansen test p–value from excluding *fl_avgh* and buffer.

Table 17: Estimation of (2) with Lagged Downloads

	Overall	Physical	Digital
DLs	-0.02***	-0.02***	-0.04***
	(0.00)	(0.00)	(0.00)
DLs, t-1	-0.00	0.00	-0.03***
	(0.48)	(0.99)	(0.01)
DLs, t-2	0.01	0.01	-0.02*
	(0.34)	(0.33)	(0.08)
DLs, t-3	-0.01	-0.01	-0.01
	(0.23)	(0.23)	(0.36)
Lag Sales	0.91***	0.90***	0.91***
	(0.00)	(0.00)	(0.00)
Genre Controls	Yes	Yes	Yes
Holiday Trend	Yes	Yes	Yes
Total IVs	24	24	24
H p-val	0.24	0.21	0.90
Observations	12,059	12,058	11,242
Groups	1,236	1,235	1,158

p-values in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 18: Shuffling Regressions for First–Stage Estimates

	Original					Shuffled				
	Pooled	Tier 1	Tier 2	Tier 3	Tier 4	Pooled	Tier 1	Tier 2	Tier 3	Tier 4
FL	0.09	0.07	0.09	0.63	-3.15***	0.10	0.11	0.15	-0.13	-0.37
	(0.56)	(0.71)	(0.73)	(0.51)	(0.01)	(0.48)	(0.52)	(0.63)	(0.80)	(0.76)
Formats on FL	-0.01	-0.03**	0.01	-0.03	0.06	-0.01*	-0.00	0.02	0.05	0.11
	(0.21)	(0.03)	(0.73)	(0.39)	(0.60)	(0.08)	(0.96)	(0.37)	(0.33)	(0.35)
Percent on FL	-2.77***	-2.48**	-2.79**	-3.09**	—	-0.37	-0.06	-0.27	—	-0.14
	(0.01)	(0.02)	(0.01)	(0.03)	—	(0.35)	(0.91)	(0.67)	—	(0.92)
Total FL Hours	0.00***	0.00***	0.00**	0.00***	-0.00	0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.03)	(0.00)	(0.86)	(0.27)	(0.81)	(0.88)	(0.48)	(0.21)
Average FL Hours	0.52***	0.49***	0.50***	0.49***	0.52***	-0.02	-0.00	-0.00	-0.00	0.02
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.63)	(0.91)	(0.99)	(.93)	(0.67)
Genre Controls	Yes	Yes				Yes	Yes			
Album Age	Yes	Yes				Yes	Yes			
Week Controls	Yes	Yes				Yes	Yes			
Lagged Sales	Yes	Yes				Yes	Yes			
User Buffer	Yes	Yes				Yes	Yes			
Shuffled=Original p-value						0.00	0.00	0.00	0.00	0.00
FL Joint p-value	0.00	0.00	0.00	0.00	0.01	0.22	0.97	0.81	0.89	0.83
Wealth Joint p-value	0.00	0.00				0.00	0.00			
Model F-stat	197.68	110.77				192.81	106.97			
Within-Rsq	0.44	0.44				0.44	0.43			
Obs	13,838	10,437				13,744	10,337			
Groups	1,436	1,275				1,427	1,270			

p-values in parentheses
* $p < .1$, ** $p < .05$, *** $p < .01$

Table 19: Estimated Aggregate Effect of Piracy on Sales

Source	Detail	Estimate
Waldfogel (2010)	Inferred; see footnote	-0.195
Rob and Waldfogel (2006)	Inferred; see footnote	-0.130
Leung (2008)	Table 11, CDs	-0.040
Oberholzer-Gee and Strumpf (2007)	Inferred; see footnote	-0.00001
Blackburn (2006)	Table 5, Average	-0.28
Mean		-0.127
Median		-0.13

Waldfogel (2010): Combined data from Table 2 with estimates from Table 5 for all; average over years.

Rob and Waldfogel (2006): See footnote 14 of Leung (2008).

Oberholzer-Gee and Strumpf (2007): See footnote 14 of Leung (2008).

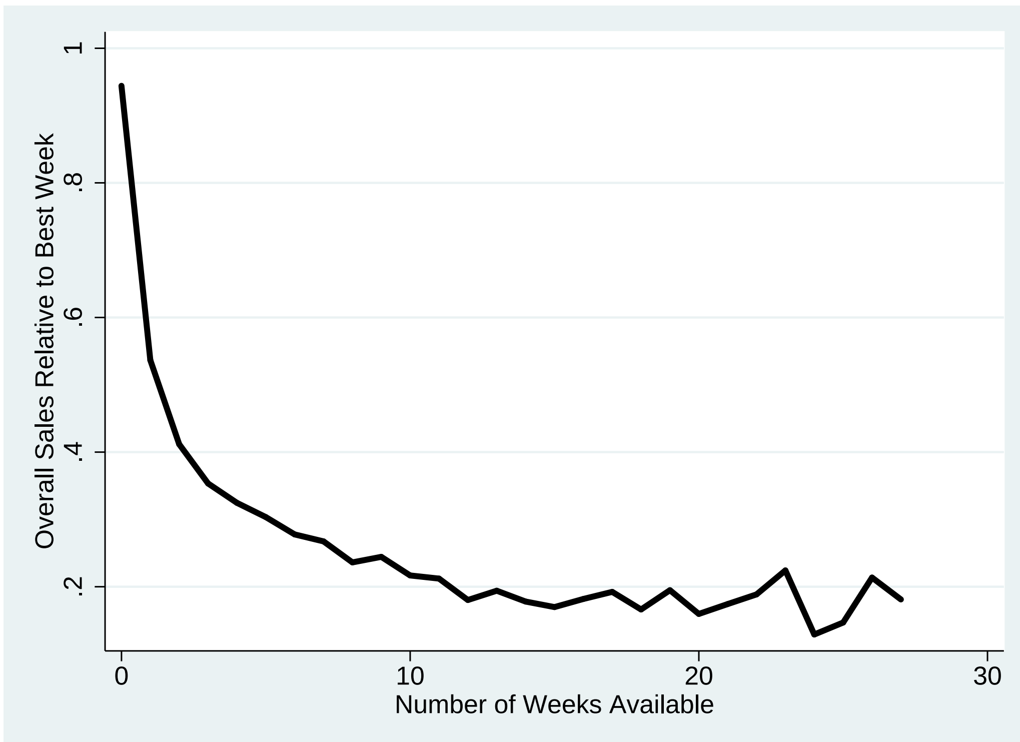
Table 20: Effects of the Sitewide Freeleech in Week 15

		Mean	Percentiles			Skew
			25th	50th	75th	
Change in Downloads	Pooled	79.54	13.19	32.58	83.97	6.05
	Tier 1	96.11	15.22	39.37	98.24	5.37
	Tier 2	57.45	11.34	23.12	55.26	2.58
Change in Sales	Pooled	-131.22	-109.34	-62.26	-44.97	-13.02
	Tier 1	-747.14	-724.05	-392.59	-229.48	-8.70
	Tier 2	2619.69	1340.84	2133.19	3114.60	3.35
Albums	Pooled			718		
	Tier 1			416		
	Tier 2			206		

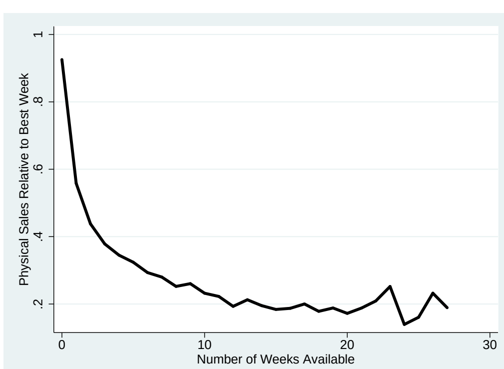
Figures

Figure 1: Sales as a Percentage of Best Week's Sales

(a) Total Sales



(b) Physical Sales



(c) Digital Sales

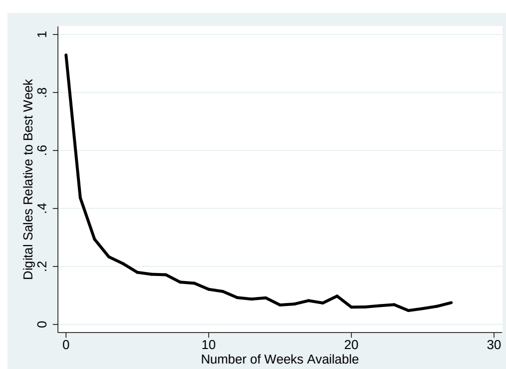
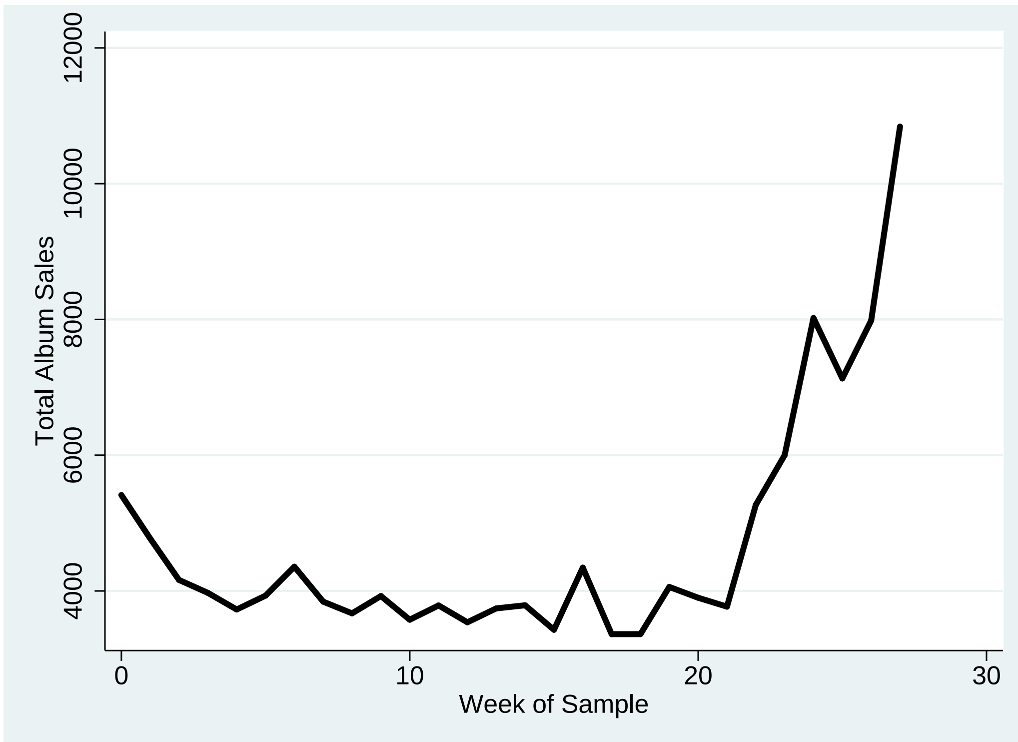
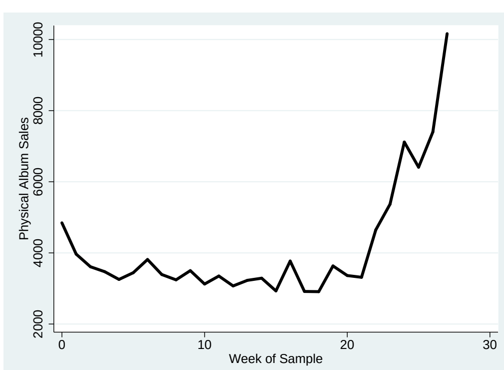


Figure 2: Average Album Sales per Week

(a) Total Sales



(b) Physical Sales



(c) Digital Sales

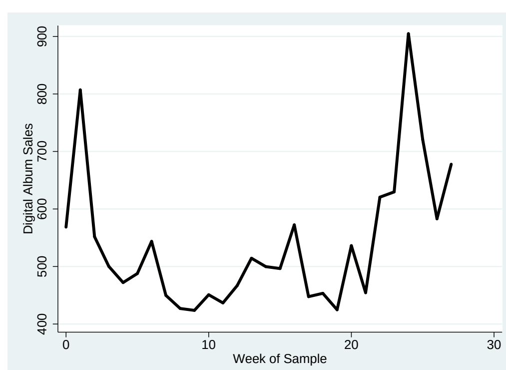


Figure 3: The Average Downloading Life–cycle of an Album

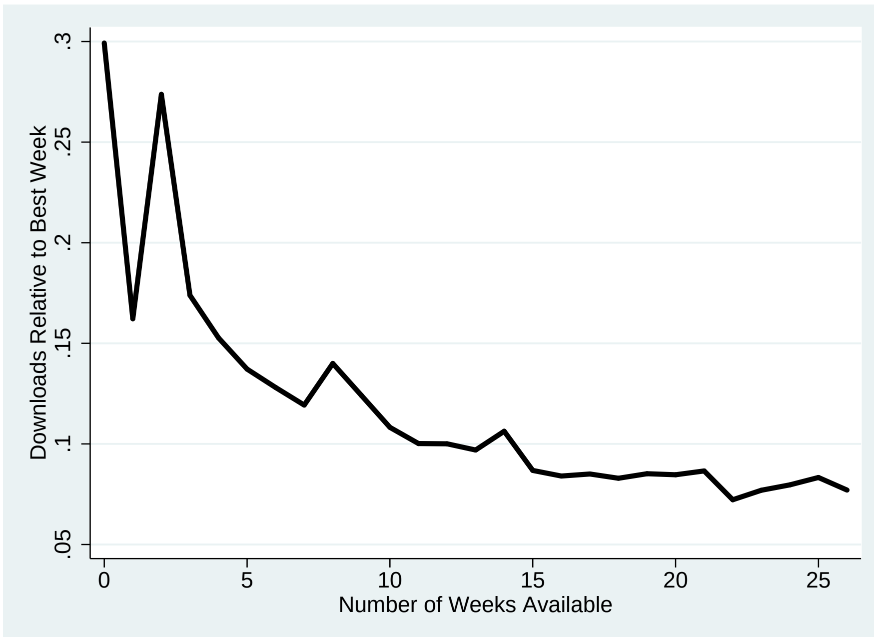


Figure 4: Average Downloads by Week, All Albums

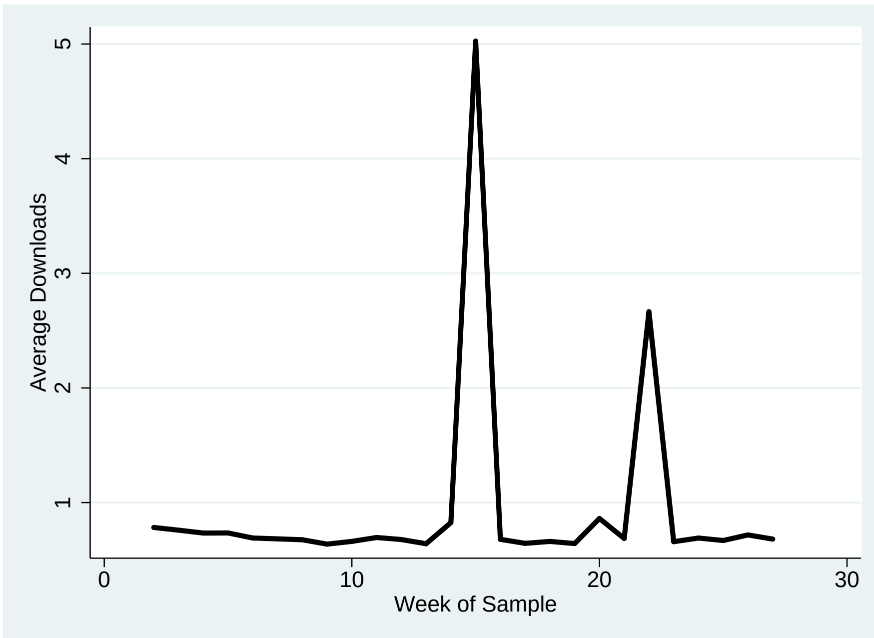


Figure 5: Average Ratio and Median Buffer

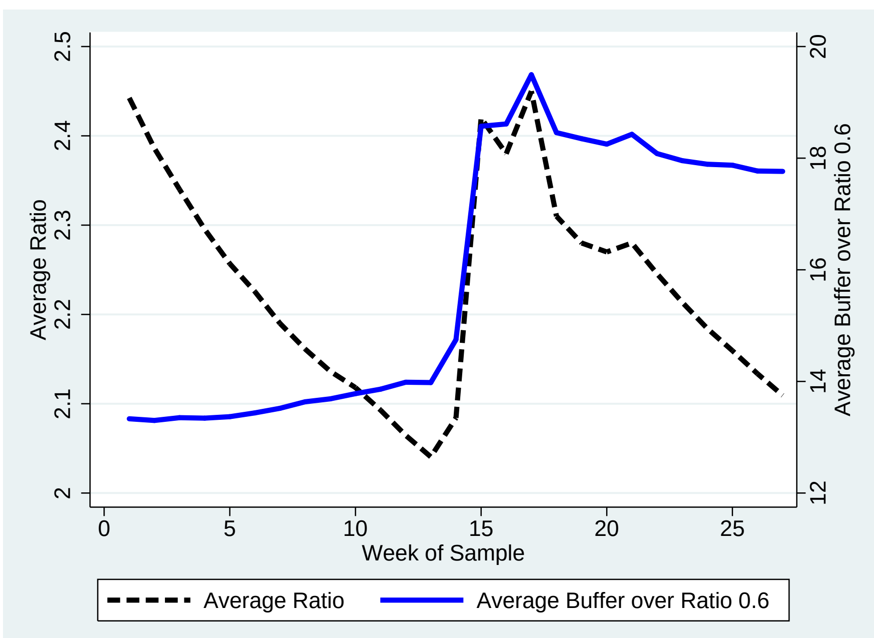


Figure 6: The Life–Cycle of a Matched Album

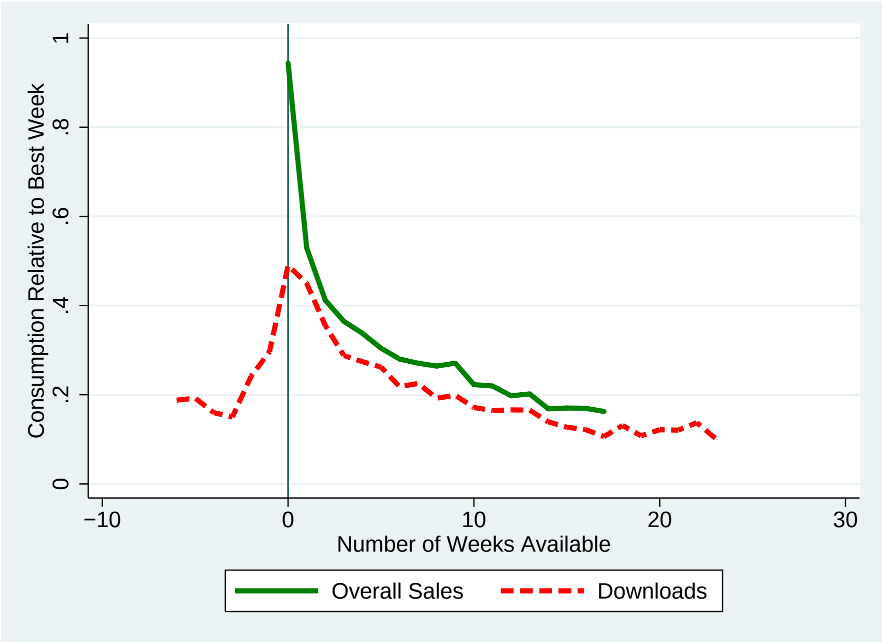


Figure 7: Average Consumption of a Matched Album