Queen's Economics Department Working Paper No. 1354

Purchase, Pirate, Publicize: The Effect of Private-Network File Sharing on Album Sales

Jonathan Lee
Queen's University

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

9-2016

Purchase, Pirate, Publicize: The Effect of Private–Network File Sharing on Album Sales

Jonathan Lee*

September 2016

Abstract

I quantify the relationship between private–network file sharing activity and album sales in the BitTorrent era using a panel of 2,109 albums' U.S. sales and file sharing downloads during 2008. Exogenous shocks to file sharing capacity address the simultaneity problem. In theory, piracy could crowd out sales by building file sharing capacity or increase them through word of mouth. I find evidence that file sharing decreases album sales for top–tier artists, but the economic impact is quite modest. However, file sharing seems to help mid–tier artists and have little effect on lower–tier artists. The results are consistent with the claim that word of mouth is stronger for lesser–known artists and that digital sales are the most vulnerable to increases in file sharing capacity.

JEL–Classification: L82, L86, O34
Keywords: intellectual property, copyright, file sharing, piracy, digital music

---

*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 for helpful comments from faculty and colleagues at Queen's University and Vanderbilt University, as well as feedback from attendees of the American Law and Economics Association 2016 Annual Meeting. I am especially indebted to the John Deustch Institute, whose 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 file sharing on album sales. I find that the file sharing elasticity of sales is -0.02 for physical albums, -0.04 for digital albums, and -0.02 in aggregate. I interpret these results as evidence that piracy crowds out legitimate sales and that this crowd–out is more salient for digital consumers, but that the practical extent of these effects is quite small. The effects differ by artist popularity as well: elasticities are largest for top–tier artists, may be positive for mid–tier artists, and are zero for lower–tier artists. I take these results as evidence that file sharing crowds out 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 small.

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. Now, 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 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. How users choose among 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 researchers. 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) spend considerable effort and resources to deter piracy and shut down file sharing networks like the one studied in this paper. If the effect of file sharing on sales is small, this expense may not be worth it.[1] The results of this paper should help to inform such cost–benefit analysis 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. Strong protection of intellectual property is needed to inhibit piracy and provide adequate incentives to create new music. The other argument claims 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

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

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, 2011a) and music quality (Waldfogel, 2011b). 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).

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 2 describes the data, which are used in Section 3 to estimate the sales–piracy relationship. Section 4 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 the usual 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 them together to build that desired file. As a result, 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 to increase overall transfer speeds. These benefits have been amplified by increased broadband penetration over time. Further, BitTorrent trackers only host the torrent files, not the actual copyrighted content, so legal challenges were made more difficult.

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). Dozens of active trackers exist, ranging from general interest networks to those that specialize in particular genres of music, TV, or movies. Trackers can be public, to which any user can connect; or private, to which only certain authorized users are allowed access. Public trackers are likely to be to be of lower reliability than private trackers: since file sharing relies on users to expend time, bandwidth, and risk in providing files to others, these networks must contend with the free–rider problem.

Private networks employ various methods to discourage free–riding, 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. These private networks tend to be smaller, more reliable, and maintain higher quality standards than public file sharing networks.

Of central interest is how file sharing 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 reputation is 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 *e.g.* 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.[2] 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 network, an exogenous change that impacts one set of consumers can nevertheless affect a much larger group; 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 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?

[2] 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.

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.[3] 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.[4]

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 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

[3]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.

[4]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.

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.[5] 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.[6] 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 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

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

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

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.[7] An album is downloaded most in the first three weeks of its being posted on the site, and then only moderately downloaded thereafter.

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 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.

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, 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 60 hours, 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.

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

The final phase was a celebration of the network's birthday. During this period, all newly uploaded files were freeleech for six hours. The celebration ran from October 31st to November 2nd (weeks 20 and 21 of the sample period). This freeleech 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.[8]

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 3.1.

The data include the mean user's ratio each week, as well as the median user's buffer each week.[9] Figure 5 plots these measures as they change during the sample period. The plot shows clear trends. During and after the freeleech periods, user wealth increases significantly, declining steadily afterward.[10] I analyze the effects of these measures on file sharing activity,

[8]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 3.2.2 and find no evidence to suggest it occurred here.

[9]User data were anonymized as a condition of data access, so I cannot measure individual wealth effects.

[10]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.

alongside freeleeches, in Section 3.

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.[11] 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, 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.

[11] It was, however, exhausting.

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.[12] 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 estimate effect of an exogenous change in file sharing activity on legitimate album sales.

3 Estimation

Drawing on insights from Section 1.1, Section 3.1 proposes an empirical framework to estimate the net effect of private–network downloads on album sales. Section 3.2 motivates the instrumental variables approach used here. Section 3.3 presents estimation results.

3.1 Empirical Strategy

I estimate the parameters of the following equation:

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

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 (1), α measures the geometric decay of album

[12]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.

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.

The factor δ measures the percentage change in sales due to a contemporaneous percentage change in file sharing: the file sharing elasticity of sales.[13] 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 contemporaneous elasticity and as an upper bound on the lifetime elasticity of album sales with respect to file sharing.

It should be noted 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, and the capacity effect will be felt on public networks if downloaders share their files on these other networks.[14]

Specification (1) 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.

[13] 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.

[14] Hammond (2014) provides some observational evidence that albums first appear on private file sharing networks, but are made available on public networks quickly thereafter.

3.2 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.

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 site. Thus simple weekly dummy variables (a possible choice for $\tau(\boldsymbol{\omega}, t)$ in equation (1)) 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 not likely that the relationships are a result of spurious secular correlation.

The results that follow include eight different secular controls. The first four are (1) week indicator variables, (2) a six–degree polynomial time trend, (3) a holiday season spline, and (4) a holiday and summer spline. The holiday spline equals zero for weeks prior to November and counts weeks linearly thereafter to account for the holiday retail season, and the holiday and summer spline additionally counts down weeks through the first weekend in July (a holiday weekend in North America). 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 splines are defined somewhat arbitrarily, but are informed by sales patterns in Figures 2 and 7 and have fewer parameters to interfere with the available IVs. For each of these four measures, I also calculate the weekly album sales trend implied by each; these conditional sales averages make up the second four controls I employ.[15] Again, none of these measures is ideal, but the ideal measure does not exist in the dataset, and each measure leads to broadly

[15] 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.

similar conclusions.[16]

3.2.1 First–Stage Regressions

Table 7 reports first–stage estimation results. The independent variable is the logarithm of downloads, and the dependent variables are the possible 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 7 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 8 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 3.3.

3.2.2 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. Table 9 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

[16] 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.

effect on file sharing activity, whether taken separately or grouped together, but the contemporaneous variables remain highly significant.

3.3 Results

To estimate (1), I employ the two–step system GMM variety of the Arellano–Bond dynamic panel estimation technique.[17] In the full specification, I include album fixed effects, genre dummies, and the various secular controls described in Section 3.2. File sharing downloads are instrumented by the album's average hours on freeleech and the median user's minimum buffer.[18] Table 10 reports results for total album sales, Table 11 reports results for physical sales alone, and Table 12 reports results for digital sales alone. I find that a transient (*i.e.*, one–period) percentage increase in 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.

While these results are statistically significant, their economic significance is lacking. An elasticity of 0.04 is extremely small, suggesting that these private file sharing networks have a limited effect on the market as a whole. These baseline estimates also suggest that what effect does exist is more concentrated in the digital music market, since the elasticity estimate is twice as high there.

However, the small magnitude of these estimates may mask larger, opposing effects. As shown in Section 3.2, 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 13 for different sales categories and artist tier definitions.

Elasticity estimates are an order of magnitude 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 file sharing slightly decreases album sales for top–tier

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

[18]Results are similar for other instrument sets. I cannot employ all measures at once due to over–fitting, as discussed in Section 3.2.

artists, and that the effect is larger for digital sales. There is weak evidence for a *positive* effect of file sharing on sales for middle–tier artists, and no effect is 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 suggestive at best, 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 3.2 imply that 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.

3.4 An Alternative Specification

The dynamic panel methods used to estimate (1) require the data to satisfy additional moment conditions, since exotic instruments are synthesized from the lagged values of the dependent variable. These conditions would be a central concern of the paper if α were the central variable of interest. However, in this paper the lagged dependent variable serves mostly to control for life–cycle patterns in album sales, and its exact specification is not crucial. In this section, I test an alternative specification without the lagged dependent variable. Instead, I include indicator variables for the age of the album in weeks up to age ten and one for age greater than ten. Table 14 reports results for total album sales, Table 15 reports results for physical sales alone, and Table 16 reports results for digital sales alone.

The results are largely consistent with the baseline results above, with a few notable exceptions. For each sales class, the specification with simple week controls reports an outlandishly large, positive coefficient inconsistent with the other estimates in the table and is even close to the pooled OLS estimate. I believe these results are due to over–fitting, since the age and week controls together comprise 260 age–week possibilities by themselves. The observant reader will also note that the coefficients are smaller for digital sales than for physical and overall sales here, contrary to my interpretation of the baseline results. However, the Hansen tests (reported as "H p–val" in the tables) reject the null hypothesis that the overidentifying restrictions are valid in Table 16, and reject quite strongly. This is not the case for the other results, which suggests that secular and life–cycle trends are not being sufficiently controlled for in this particular specification.[19] Overall, the results from this alternative control of the sales life–cycle support the core conclusion of the baseline model: file sharing yields a statistically significant but economically small decrease in album sales.

4 Conclusion

Researchers have debated the nature of file sharing's effects on legitimate markets *ad nauseum*, but consensus has been hard to come by. This paper alone cannot close that debate, but it does shed new light on the nature of the relationship. From the results, I conclude that file sharing activity has a statistically significant but economically small negative effect on legitimate music sales. This relationship varies by medium: file sharing decreases sales of digital music more than sales of physical copies, especially for top–tier artists. The results also suggest that file sharing may increase sales for mid–tier artists but give no evidence for an effect on lower–tier artists. These results are consistent with the claim that the word–of–mouth effect is stronger for lesser–known artists, and that the capacity effect is strongest for consumers who are prone to digital consumption. The results are robust to different instrument sets, secular controls, and alternative model specifications.

It is important to remember that these results come from a time when file sharing was in its prime, at least for music. Previous work in the literature focuses on file sharing technology in its nascent stages, where quality, reliability, and ease of use were not guaranteed. Users downloaded one song at a time, not the entire album, so the sampling effect invoked by some authors was plausible. Using the technology under study in this paper, file sharers downloaded entire albums as a unit instead of individual songs, so the sampling effect seems less likely. The word–of–mouth effect instead acts as free publicity, and non–pirates increase their consumption *ceteris paribus*. The results support the contention that this effect is strongest for artists whose reputation is less established.

Piracy not only affects the market for music, it affects markets for all

[19]Note that difference–in–Hansen tests for the main specification are harder to cleanly interpret. Instrument proliferation not only risks overfitting, it also reduces the power of these tests. See Roodman (2009) for a discussion.

media that can be digitized: television, film, and e–book markets all compete with file sharing networks that distribute their content freely. Further work is needed to determine the nature of these relationships in their specific markets, but the current paper provides insight into what economic forces might be at play and how they might break down across different classes of goods. At least in the case of music, the results suggest that if customers are being lost to piracy, content producers should consider making digital consumption easier (where elasticity is highest) instead of expending resources on shutting down file sharing networks. Doing so could crowd out file sharing and harness the word–of–mouth effect, both of which would lead to digital purchases instead of copyright violations from marginal consumers. If the same effects manifest in markets for film or television, digital distribution should be a primary focus for film studios and television networks as well.

References

Andersen, Birgitte and Marion Frenz (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, 715–740.

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

Blackburn, David (2006), "The heterogeneous effects of copying: The case of recorded music." Unpublished Manuscript. Available at `http://www.davidjhblackburn.com/papers/blackburn_fas.pdf`.

Connolly, Marie and Alan B Krueger (2006), "Rockonomics: The economics of popular music." *Handbook of the Economics of Art and Culture*, 1, 667–719.

Danaher, Brett, Michael D Smith, Rahul Telang, and Siwen Chen (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, 541–553.

Dejean, Sylvain (2009), "What can we learn from empirical studies about piracy?" *CESifo Economic Studies*, ifp006.

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

Gopal, Ram D, Sudip Bhattacharjee, and G Lawrence Sanders (2006), "Do artists benefit from online music sharing?" *The Journal of Business*, 79, 1503–1533.

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

Hong, Seung-Hyun (2013), "Measuring the effect of napster on recorded music sales: Difference–in–differences estimates under compositional changes." *Journal of Applied Econometrics*, 28, 297–324.

Hui, Kai-Lung and Ivan Png (2003), "Piracy and the legitimate demand for recorded music." *Contributions in Economic Analysis & Policy*, 2.

Lee, Jonathan (2009), "Private filesharing markets and album sales." Undegraduate Honor's Thesis, Vanderbilt University Department of Economics. Available at `http://as.vanderbilt.edu/econon/documents/honor09Lee.pdf`.

Leung, Tin Cheuk (2008), "Should the music industry sue its own customers? impacts of music piracy and policy suggestions." *University of Minnesota, retrieved Nov*, 28, 2012.

Liebowitz, Stan J (2005a), "Economists' topsy–turvy view of piracy." *Review of Economic Research on Copyright Issues*, 1, 5–17.

Liebowitz, Stan J (2005b), "Pitfalls in measuring the impact of file–sharing on the sound recording market." *CESifo Economic Studies*, 51, 435–473.

Liebowitz, Stan J (2006a), "Economists examine file–sharing and music sales." *Industrial Organization And The Digital Economy*, 145–174.

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

Liebowitz, Stan J (2008), "Research note–testing file sharing's impact on music album sales in cities." *Management Science*, 54, 852–859.

Mortimer, Julie Holland, Chris Nosko, and Alan Sorensen (2012), "Supply responses to digital distribution: Recorded music and live performances." *Information Economics and Policy*, 24, 3–14.

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

Oberholzer-Gee, Felix and Koleman Strumpf (2010), "File sharing and copyright." In *Innovation Policy and the Economy, Volume 10*, 19–55, University of Chicago Press.

Palo Alto Networks (2014), "Application usage and threat report." Technical Report.

Peitz, Martin and Patrick Waelbroeck (2004), "The effect of internet piracy on CD sales: Cross-section evidence."

Peitz, Martin and Patrick Waelbroeck (2006a), "Piracy of digital products: A critical review of the theoretical literature." *Information Economics and Policy*, 18, 449–476.

Peitz, Martin and Patrick Waelbroeck (2006b), "Why the music industry may gain from free downloading — the role of sampling." *International Journal of Industrial Organization*, 24, 907–913.

Rob, Rafael and Joel Waldfogel (2006), "Piracy on the high C's: Music downloading, sales displacement, and social welfare in a sample of college students." *Journal of Law and Economics*, 49, 29–62.

Roodman, David (2006), "How to do xtabond2: An introduction to difference and system gmm in stata." *Center for Global Development Working Paper*.

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

Waldfogel, Joel (2010), "Music file sharing and sales displacement in the itunes era." *Information economics and policy*, 22, 306–314.

Waldfogel, Joel (2011a), "Bye, bye, miss American pie? the supply of new recorded music since Napster." Technical report, National Bureau of Economic Research.

Waldfogel, Joel (2011b), "Copyright protection, technological change, and the quality of new products: Evidence from recorded music since Napster." Technical report, National Bureau of Economic Research.

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

Tables

Table 1: Summary Statistics for Sales

Statistic	**Mean**	**Percentiles**			**Skew**
		25th	50th	75th	
Total Sales	45,596	2,714	11,427	38,687	9.4
Log TotalSales	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

Week	**Share**		**Average Hours**	
	All	Matched	All	Matched
13	0.6%	0.3%	6.1	7
14	100%	100%	19.5	52.3
15	100%	100%	80.6	222.4
20	0.8%	2.0%	7.2	6.3
21	2.1%	4.6%	7.4	7.9
Others	None			

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: 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**	0.22*	0.25**	0.25**	0.23**	0.23**
	(0.00)	(0.00)	(0.03)	(0.39)	(0.48)	(0.14)	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)
fl_pct	-2.65***	-2.54***	-0.86***	0.13	0.34	-0.77***	-0.78***	-0.79***	-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)	(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***	0.04***	0.04***
	(0.00)	(0.00)	(0.00)	(0.52)	(0.98)	(0.00)	(0.00)	(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	-0.01	-0.01
	(0.00)	(0.00)	(0.16)	(0.19)	(0.11)	(0.11)	(0.12)	(0.12)	(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)	(0.00)	(0.00)	(0.00)	(0.00)
Ratio		-0.53***	-0.03	0.83	4.86***	5.83***	-0.23***	-0.40***	-0.15***	-0.15***	-0.20***	-0.20***
		(0.00)	(0.56)	(0.57)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
Buffer		0.05***	0.01***	-0.41	-1.86***	-0.55***	0.05***	0.05***	0.04***	0.04***	0.04***	0.04***
		(0.00)	(0.01)	(0.49)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Genre Controls	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lagged Sales	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week Controls	No	No	No	Yes	Yes	No	No	No	Yes	No	No	No
Time Trend	No	No	No	No	No	Yes	No	No	No	Yes	No	No
Holiday Trend	No	No	No	No	No	No	Yes	Yes	No	No	Yes	No
Summer Trend	No	No	No	No	No	No	No	Yes	No	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	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	0.00	0.00
Model F-stat	395.41	302.93	1,112.88	178.72	237.64	345.23	396.56	383.20	396.29	396.27	396.28	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	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	16,800	16,800
Albums			1,589	1,589	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 8: 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 9: 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 10: Estimation of (1) for Overall Sales.

	OLS	IV	Panel IV	Dynamic Panel IV, Direct Secular Controls				Dynamic Panel IV, Secular Mean Controls			
DLs	0.01***	-0.01**	-0.01***	-0.01**	-0.02	-0.02***	-0.02***	-0.02***	-0.00	-0.02***	-0.02***
	(0.00)	(0.02)	(0.00)	(0.02)	(0.69)	(0.00)	(0.00)	(0.00)	(0.46)	(0.00)	(0.00)
Lag Sales	0.90***	0.91***	0.69***	0.88***	0.90***	0.90***	0.90***	0.89***	0.88***	0.90***	0.90***
	(0.00)	(0.00)	(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	Yes	Yes
Week Polynomial	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
Week Controls	No	No	No	No	Yes	No	No	No	Yes	No	No
Holiday Trend	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Summer Trend	No	No	No	No	No	No	Yes	No	No	No	Yes
Total IVs		2	2	27	24	27	27	27	27	27	27
H p-val		0.26	0.81	0.01	0.11	0.37	0.48	0.74	0.78	0.33	0.33
Observations	16,800	16,800	16,588	16,800	16,800	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	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 11: Estimation of (1) for Physical Sales.

	OLS	IV	Panel IV	Dynamic Panel IV, Direct Secular Controls				Dynamic Panel IV, Secular Mean Controls			
DLs	0.01***	-0.00	-0.01**	-0.01	-0.01	-0.02***	-0.02***	-0.01*	0.00	-0.02***	-0.02***
	(0.00)	(0.64)	(0.01)	(0.25)	(0.73)	(0.00)	(0.00)	(0.08)	(0.59)	(0.00)	(0.00)
Lag Sales	0.90***	0.91***	0.67***	0.88***	0.90***	0.90***	0.89***	0.88***	0.88***	0.89***	0.89***
	(0.00)	(0.00)	(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	Yes	Yes
Week Polynomial	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
Week Controls	No	No	No	No	Yes	No	No	No	Yes	No	No
Holiday Trend	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Summer Trend	No	No	No	No	No	No	Yes	No	No	No	Yes
Total IVs		2	2	27	24	27	27	27	27	27	27
H p-val		0.97	0.05	0.16	0.17	0.97	0.73	0.79	0.45	0.80	0.80
Observations	16,787	16,787	16,580	16,787	16,787	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	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 12: Estimation of (1) for Digital Sales.

	OLS	IV	Panel IV	Dynamic Panel IV, Direct Secular Controls				Dynamic Panel IV, Secular Mean Controls			
DLs	0.03***	-0.03***	-0.02***	-0.03***	-0.06	-0.04***	-0.04***	-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)	(0.00)	(0.00)
Lag Sales	0.89***	0.92***	0.67***	0.84***	0.87***	0.85***	0.87***	0.79***	0.80***	0.85***	0.85***
	(0.00)	(0.00)	(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	Yes	Yes
Week Polynomial	Yes	Yes	Yes	Yes	No	No	No	Yes	No	No	No
Week Controls	No	No	No	No	Yes	No	No	No	Yes	No	No
Holiday Trend	No	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Summer Trend	No	No	No	No	No	No	Yes	No	No	No	Yes
Total IVs		2	2	27	24	27	27	27	27	27	27
H p-val		0.00	0.00	0.01	0.01	0.05	0.12	0.16	0.05	0.04	0.04
Observations	15,537	15,537	15,343	15,537	15,537	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	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 13: Estimation of (1) 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	0.04	0.04	0.04	0.04	0.04	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 14: Alternative Specification of (1) for Overall Sales.

	OLS	IV	Panel IV, Direct Secular Controls				Panel IV, Secular Mean Controls			
DLs	0.25***	0.01	0.22***	-0.03***	-0.07***	-0.06***	-0.04***	-0.04***	-0.07***	-0.07***
	(0.00)	(0.53)	(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	Yes
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week Polynomial	Yes	Yes	No	Yes	No	No	Yes	No	No	No
Week Controls	No	No	Yes	No	No	No	No	Yes	No	No
Holiday Trend	No	No	No	No	Yes	Yes	No	Yes	Yes	No
Summer Trend	No	No	No	No	No	Yes	No	No	No	Yes
H p-val		0.00	0.20	0.00	0.35	0.10	0.01	0.54	0.39	0.39
Within-R2	0.32	0.24	0.56	0.51	0.38	0.40	0.38	0.39	0.37	0.37
Observations	18,208	18,208	17,803	17,803	17,803	17,803	17,803	17,803	17,803	17,803
Albums			1,532	1,532	1,532	1,532	1,532	1,532	1,532	1,532

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: Alternative Specification of (1) for Physical Sales.

	OLS	IV	Panel IV, Direct Secular Controls				Panel IV, Secular Mean Controls			
DLs	0.22***	0.01	0.21***	-0.03***	-0.07***	-0.07***	-0.04***	-0.04***	-0.07***	-0.07***
	(0.00)	(0.72)	(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	Yes
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week Polynomial	Yes	Yes	No	Yes	No	No	Yes	No	No	No
Week Controls	No	No	Yes	No	No	No	No	Yes	No	No
Holiday Trend	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Summer Trend	No	No	No	No	No	Yes	No	No	No	Yes
H p-val		0.00	0.17	0.02	0.37	0.86	0.20	0.27	0.33	0.33
Within-R2	0.29	0.23	0.53	0.48	0.35	0.36	0.35	0.36	0.33	0.33
Observations	18,176	18,176	17,787	17,787	17,787	17,787	17,787	17,787	17,787	17,787
Albums			1,527	1,527	1,527	1,527	1,527	1,527	1,527	1,527

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: Alternative Specification of (1) for Digital Sales.

	OLS	IV	Panel IV, Direct Secular Controls				Panel IV, Secular Mean Controls			
DLs	0.48***	0.04*	0.29***	-0.01	-0.02***	-0.02**	-0.03***	-0.02***	-0.02***	-0.02***
	(0.00)	(0.08)	(0.00)	(0.19)	(0.00)	(0.04)	(0.00)	(0.01)	(0.00)	(0.00)
Genre Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Week Polynomial	Yes	Yes	No	Yes	No	No	Yes	No	No	No
Week Controls	No	No	Yes	No	No	No	No	Yes	No	No
Holiday Trend	No	No	No	No	Yes	Yes	No	No	Yes	Yes
Summer Trend	No	No	No	No	No	Yes	No	No	No	Yes
H p-val		0.00	0.42	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Within-R2	0.38	0.22	0.47	0.44	0.39	0.40	0.39	0.39	0.39	0.39
Observations	16,880	16,880	16,499	16,499	16,499	16,499	16,499	16,499	16,499	16,499
Albums			1,416	1,416	1,416	1,416	1,416	1,416	1,416	1,416

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

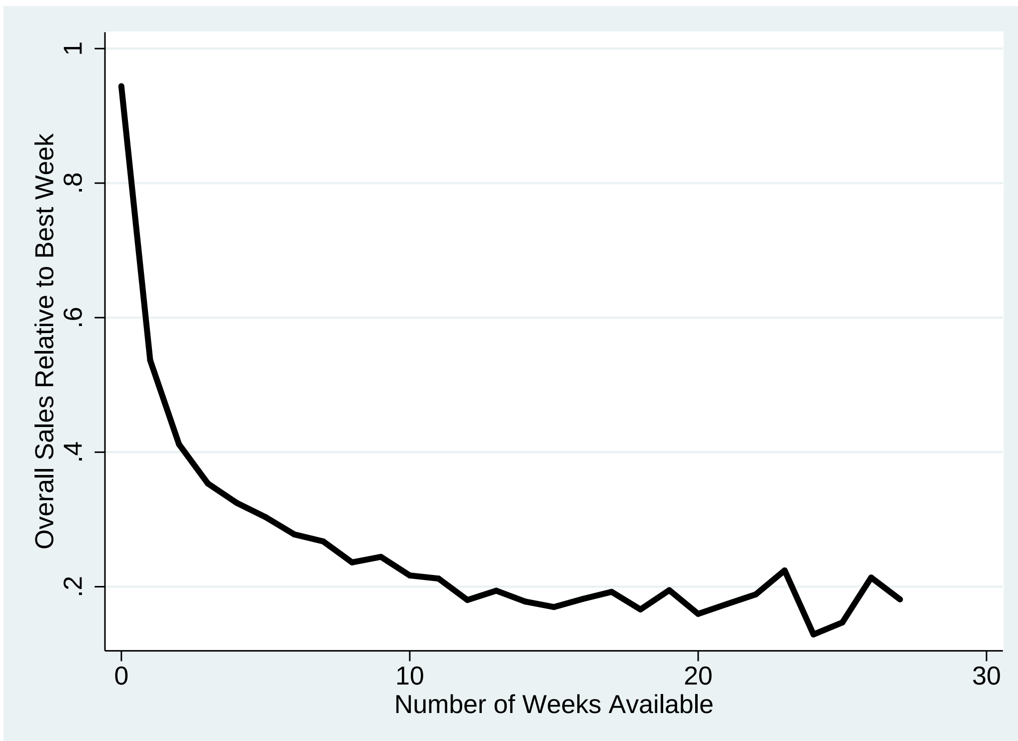
Instruments used: *fl_avgh*, buffer, and those derived from lagged values of the dependent variable.

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

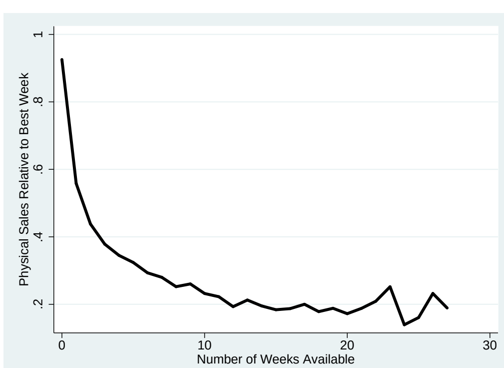
Figures

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

(a) Total Sales



(b) Physical Sales



(c) Digital Sales

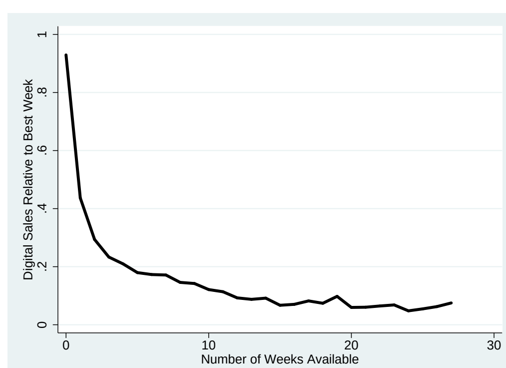
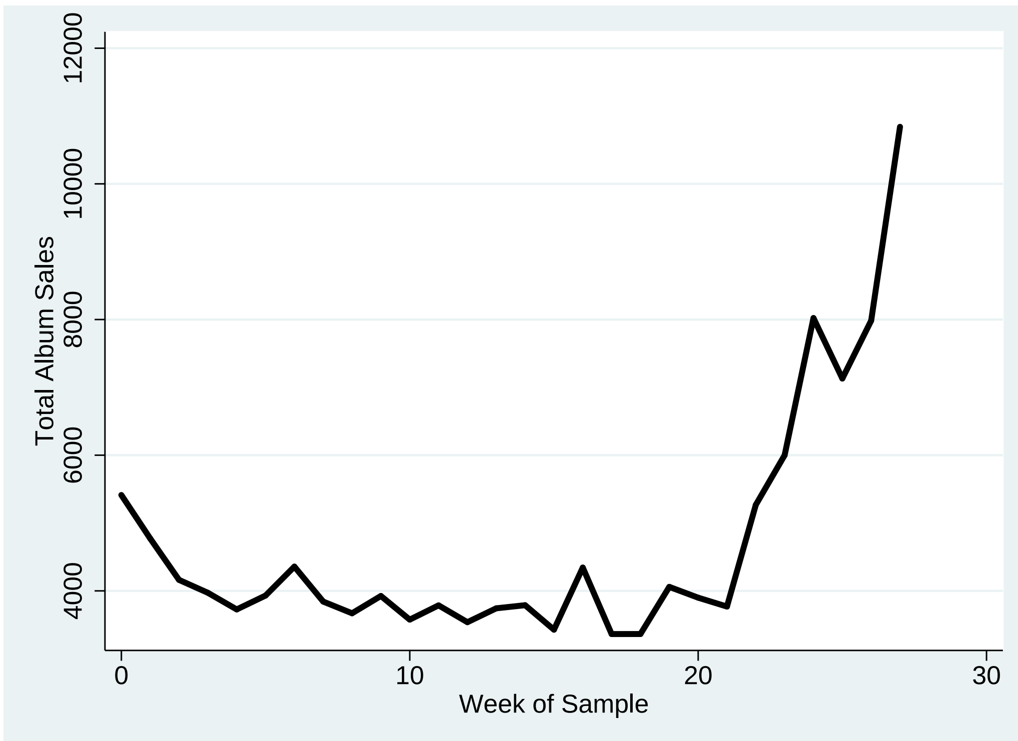
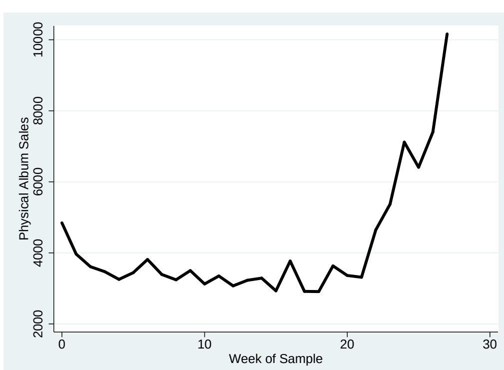


Figure 2: Average Album Sales per Week

(a) Total Sales



(b) Physical Sales



(c) Digital Sales

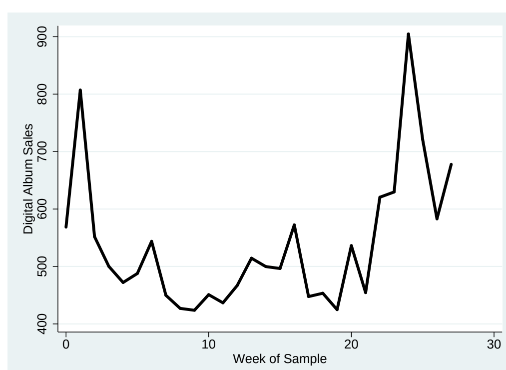


Figure 3: The Average Downloading Life–cycle of an Album

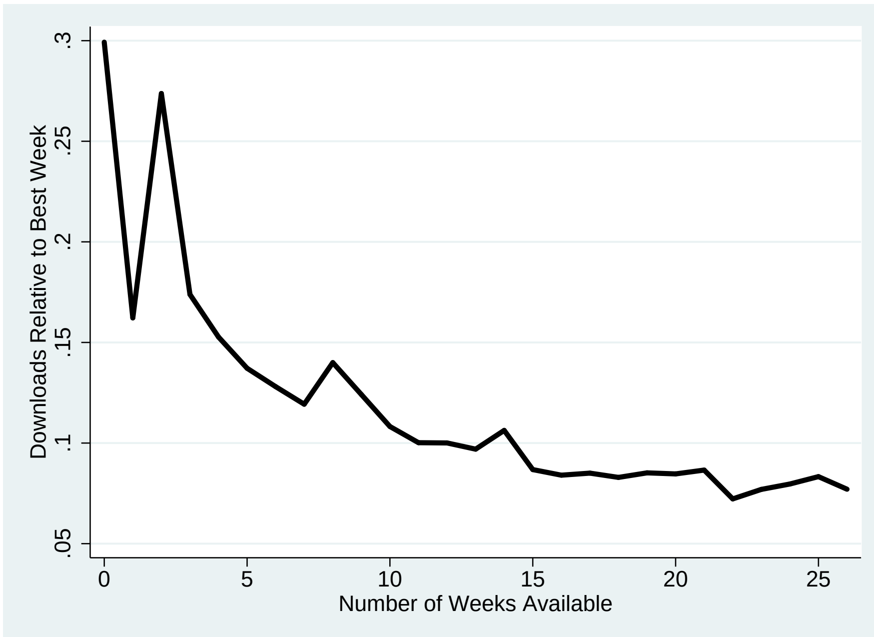


Figure 4: Average Downloads by Week, All Albums

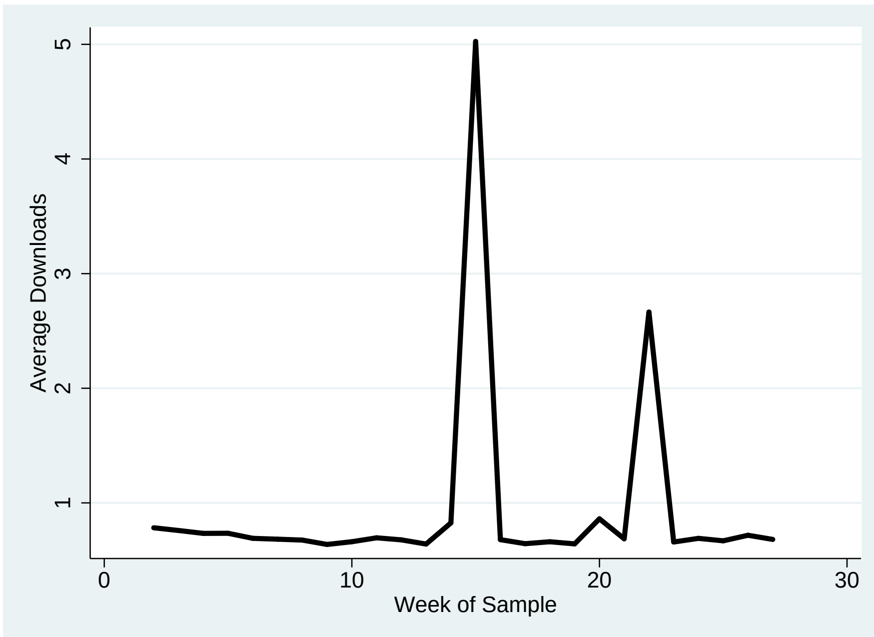


Figure 5: Average Ratio and Median Buffer

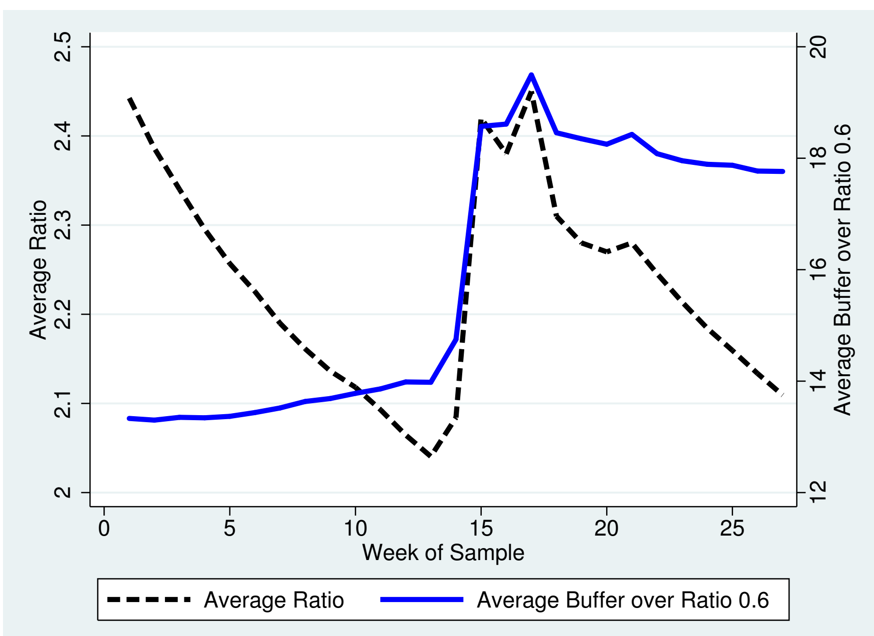


Figure 6: The Life–Cycle of a Matched Album

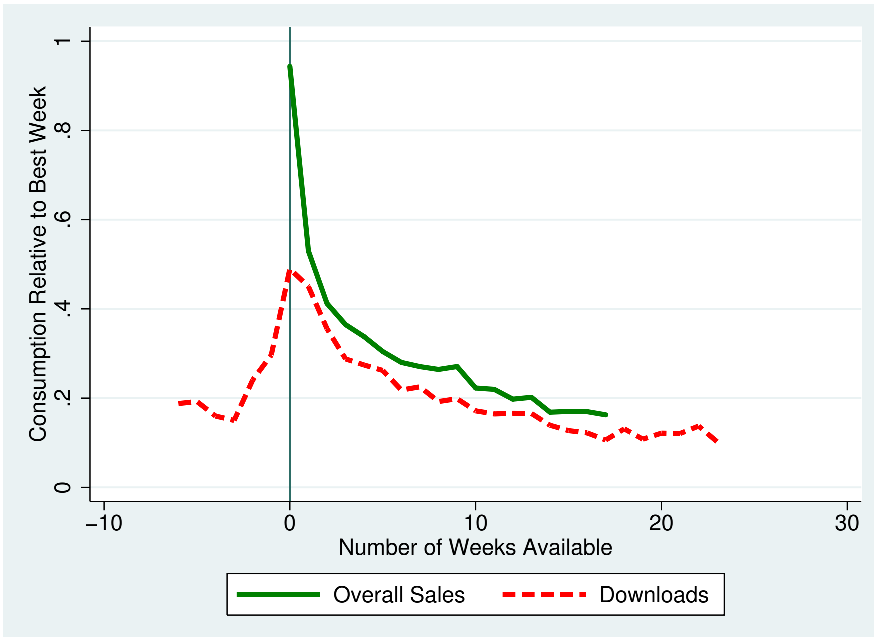


Figure 7: Average Consumption of a Matched Album