

Lee, Jonathan

Working Paper

Purchase, pirate, publicize: The effect of file sharing on album sales

Queen's Economics Department Working Paper, No. 1354

Provided in Cooperation with:

Queen's University, Department of Economics (QED)

Suggested Citation: Lee, Jonathan (2016) : Purchase, pirate, publicize: The effect of file sharing on album sales, Queen's Economics Department Working Paper, No. 1354, Queen's University, Department of Economics, Kingston (Ontario)

This Version is available at:

<https://hdl.handle.net/10419/149080>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



Queen's Economics Department Working Paper No. 1354

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

Jonathan Lee
Queen's University

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

1-2016

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

Jonathan Lee*

January 2016

Abstract

This paper quantifies the relationship between private-network file sharing activity and music sales in the BitTorrent era. Using a panel dataset of 2,251 albums' U.S. sales and file sharing downloads on a private network during 2008, I estimate the effect of file sharing on album sales. Exogenous shocks to file sharing capacity address the simultaneity problem. In theory, piracy could crowd out legitimate sales by building file sharing capacity, but could also increase sales through word-of-mouth. I find evidence that additional file sharing decreases physical sales but increases digital sales for top-tier artists, though the effects are modest. I also find that file sharing may help mid-tier artists and substantially harms bottom-tier artists, suggesting that file sharing enables consumers to better discern quality among lesser-known artists.

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. I am especially indebted to the John Deutch 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 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 music sales and its file sharing counterpart. Drawing from data on US album sales and on activity within a private file sharing network, I follow 2,251 albums over 27 weeks in 2008 and estimate the effect of an exogenous change in file sharing on album sales. For the baseline specification, the file sharing elasticity of sales is -0.13 in aggregate, -0.17 for physical albums, and $+0.21$ for digital albums. I interpret these results as evidence that piracy crowds out legitimate sales but that this displacement is outweighed by a word-of-mouth effect in the market for digital albums: since pirates are more likely to share social links (and thus discuss music) with digital consumers than with physical consumers, the word-of-mouth effect matters more for digital sales. The effects of increased file sharing activity differ across artist popularity as well: top-tier artists' physical sales are slightly decreased and digital ones slightly increased, mid-tier artists' physical sales are unaffected and digital ones slightly helped, and bottom-tier artists' sales are considerably reduced across formats.

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 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 download-

ing, 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 indicators 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. 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 has not been established.

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 existence of a “sampling” effect, wherein file sharing allows users to try before they buy, and the authors conclude 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 results as evidence of a word-of-mouth effect which is similar to the sampling effect, but incorporates social network structure and thus provides a theoretical motivation for the discrepancy in results for physical and digital sales.

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 (*e.g.* 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.¹

¹Blackburn (2006) interprets his findings as evidence of “substitution” and “penetration” effects, the latter being analogous to the word-of-mouth effect proposed in this paper.

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 or inherent characteristics of an album, and is unique in that quality. The findings of the paper thus sheds 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 2 broadly outlines a theoretical model of its interaction with legitimate markets. Section 3 describes the data, which are used in Section 4 to estimate the sales-piracy relationship. Section 5 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 networks, 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.*, mu-

sic) 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 tend 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 lowers the cost of discovery of new goods: in legal markets, it is harder 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 outline a model of the market for music to address these different possibilities. Using the model, I motivate the idea that the net effect of file sharing activity on music sales comprises the displacement and word-of-mouth effects, and that the word-of-mouth effect should be stronger for digital sales.

2 The Sharing-Sales Relationship

Consider the market for a single album of music of unknown quality q^* . There is a large body of potential consumers, connected through a social network, who have some prior belief q_0 about the value of the album.² The market exists in discrete periods in which a person can choose to consume the album through a (*P*)hysical purchase at a retailer, a (*D*)igital purchase from *e.g.* iTunes, or a file (*S*)haring interaction. Alternatively, she can choose to (*N*)ot consume the album at all. Each of these actions give her value V_P , V_D , V_S , or V_N , which jointly embody price, risk, aesthetic value of a physical copy, convenience of a digital copy, etc, independent of album quality q^* . Note that while q^* takes a single value, the consumption valuations V_i vary across agents. Letting these values represent lifetime payoffs, the expected value of each consumption action is $U_i = q_0 + V_i$ for $i \in \{P, D, S\}$, and the value of non-consumption is $U_N = V_N$. At the beginning of each period t , the person chooses an action using the following decision rule: consume if $\max\{U_P, U_D, U_S\} \equiv U(q) \geq U_N$, where $U(q)$ is the value from consuming the album of expected quality q by their preferred method.

Upon consuming, the agent learns the album's true quality q^* and credibly relays this information to some share of her neighbors on the social network, after which she exits the game. In the following period, the newly-informed neighbors repeat the above consumption decision if they have not already consumed, but use q^* instead of q_0 . Depending on the distributions of the V_i , some agents will now choose consumption where they would have otherwise

²One can also account for consumers' risk-aversion, which would drive a sampling effect, by interpreting q_0 as a belief on the album's certainty-equivalent quality.

chosen no consumption. As the signal q^* propagates across the network over time, more agents learn q^* and have the opportunity to change their decision. The process halts when none of a consumer's neighbors that she informs wishes to switch based on the new information. In short, one consumer's experience can elicit a chain of consumption from agents that would otherwise not have consumed. I refer to this network propagation as the word-of-mouth effect. Note that since agents with a $q_0 > q^*$ either consume at time zero or never consume even after learning q^* , this word-of-mouth process will by definition increase consumption.

At the same time, additional consumption through file sharing increases the capacity of file sharing activity, since the new consumer may share the album to future consumers anywhere in the market. I model this change as a simultaneous increase in V_S for all agents in the game. One additional album consumed through file sharing at time t can then elicit new file sharing consumption in $t + 1$ from V_S -marginal agents. Given time, these marginal agents may have eventually been moved to purchase legally from the word-of-mouth effect, but now they are moved to file share since V_S has increased. This crowd-out will increase file sharing consumption, but will necessarily displace legal sales that would otherwise have occurred, akin to the usual substitution effect.

The direction and magnitude of the net effect of file sharing on legitimate album purchases thus depends on the relative sizes of the word-of-mouth and crowd-out effects. Since consumers can purchase either physically or digitally, one might expect the two effects to manifest in different ways depending on which market is considered. Invoking the idea of homophily in social network formation, I assume that agents are more likely to be linked to neighbors whose V_i are similar. In particular, homophily would imply that agents who prefer digital consumption on a computer or portable music player are more likely to discuss their music tastes with friends who also prefer digital consumption. In the context of the model, the social neighbors of a high- V_S agent would most likely have a high V_S and V_D , which implies that the word-of-mouth effect should have a larger effect on digital sales than on physical sales.

Of course, the propagation of these effects depend on an initial seed of consumption at the beginning of the game. 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 3.

3 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

3.1 Album Sales Data

Data on album sales are provided by Nielsen SoundScan, who compile sales figures for US retail 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 number of physical and digital copies sold, the record label, 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 share of sales than others. Digital sales exhibit different life–cycle and secular patterns from their physical counterparts; I discuss these differences below.

Figure 1 shows the average album’s sales as a percentage of its first week’s sales over its life. A typical album reaches peak sales in the two weeks following release, and sales decline steadily afterwards. This trend is especially true for physical album sales, but not for digital sales. As Figure 1c shows,

³Ideally, the econometrician would observe individual agents in the midst of a purchase/pirate decision, but it is unlikely such data could be obtained without introducing self–reporting problems.

⁴I do not have data on pricing for albums, nor a coherent way to “price” downloads, so I work in quantities. As the model is populated by consumers with unit demand for a given album, estimating a quantity relationship should be appropriate.

the first week’s digital sales are high, experience a lull, are high again around ten weeks after release, and decline rapidly afterward.

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, which I attribute to the holiday retail season. Again, physical sales track this trend closely, but digital sales exhibit somewhat less seasonality, and are also more prominent during summer months, as can be seen in Figure 2c.

Since the data capture sales across many artists (about 1,400), 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 highest-ranking album and use that as my measure of artist popularity. I divide artists into quartiles according to that measure.⁵ Table 2 presents these tiers and their composition. Tiers vary in size, with higher tiers having more albums.

3.2 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 4.8 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. Each week, only about a quarter of available albums are downloaded, but this is unsurprising since old albums will often remain available long after their popularity has waned. Figure 3 depicts the average life-cycle of an album in the downloads database. 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 two weeks of its being posted on the site, and then only moderately downloaded thereafter.

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

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

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

Figure 4 shows downloads for an average album across the sample period. File sharing activity is fairly constant, with a few exceptions: large spikes occur around weeks 16 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 albums on freeleech 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 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 albums listed far exceeded the necessary amount (about 3,000), this freeleech period was stretched into about 60 hours, from September 19th through 22nd. 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 six hours. The celebration ran from October 31st to November 2nd. 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.

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 week 16. These freeleeches impacted users’ ability to download

and share files contemporaneously, but they also increased future capacity by slackening users' sharing constraints. I quantify the constraint's effects presently.

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 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.⁸ 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.⁹ I analyze the effects of these measures on file sharing activity, alongside freeleeches, in Section 4.

3.3 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. Summary statistics of the matched albums are reported in Table 5.

The match is not exhaustive. Less than 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. More than three-quarters of albums from the sales data were matched to albums in

⁸Unfortunately, the user data are anonymized, so I cannot measure individual wealth effects.

⁹This 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.

the file sharing data. I am confident that the unmatched albums are truly unmatchable due to the merging procedure used, which I outline presently.

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 titles with nonstandard characters (*e.g.* Sigur Rós or Beyoncé). For these exceptions, downloads may be slightly under-counted due to different type-setting 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, so I manually searched for a match in the file sharing data. Sometimes 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.

Given the 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.

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. The digital share of sales is very slightly higher. Figure 6 shows that an album’s downloads and sales track similar patterns, starting high and decaying gradually as they age. 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 4, I use these data to estimate effect of an exogenous change in file sharing activity on legitimate album sales.

4 Estimation

Drawing on insights from Section 2, 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:

$$\ln(s_{it}) = \alpha \ln(s_{it-1}) + \delta \ln(\hat{d}_{it}) + \beta a_{it} + \mathbf{g}_i \boldsymbol{\gamma} + w_t + u_i + \epsilon_{it} \quad (1)$$

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 , a_{it} is the age of album i in week t , the w_t are time fixed effects, the u_i are album fixed effects, and the errors $\epsilon_{it} \sim iid(0, \sigma_\epsilon^2)$.¹⁰

As alluded to in Section 2, 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 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. Other coefficients 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.¹¹ But since file sharing also affects later sales through $s_{i,t-1}$ and α , δ does not capture the full economic effect.

¹⁰Because of the dynamic nature of these effects, restricting the model to include only one lag of sales and only contemporaneous downloads may be too severe. I consider longer–term relationships in Appendix A.

¹¹The log–transformation of sales and downloads is appropriate because the capacity and word–of–mouth effects change the behavior of a share of the population, not a set number of consumers. Hence these effects should manifest as a percentage change in sales due to a percentage change in downloads.

The relevant measure is the *cumulative marginal effect* (CME), defined as

$$\begin{aligned} \text{CME} &= \frac{\partial \ln(s_{it})}{\partial \ln(d_{it})} + \frac{\partial \ln(s_{it+1})}{\partial \ln(d_{it})} + \frac{\partial \ln(s_{it+2})}{\partial \ln(d_{it})} + \dots \\ &= \delta(1 + \alpha + \alpha^2 + \dots) = \delta \sum_{t=0}^{\infty} \alpha^t \end{aligned} \tag{2}$$

The CME measures the percentage change in an album’s lifetime sales due to a one-time percentage change in 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 and CME measure the effect of a private tracker’s file sharing activity on aggregate album sales. That said, file sharing activity here will still initiate 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.¹²

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.

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 behave differently for different albums at different times. Further, any given album could be available for downloading in different formats, each of which could differ in their freeleech status, and a copy of each format could be shared by a number of users. 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

¹²Hammond (2014) provides some observational evidence that albums first appear on private file sharing networks, but are made available on public networks quickly thereafter.

provide more information by counting the number of freeleech hours, available formats, or shared copies. These measures will obviously exhibit a high degree of correlation, so not all measures can be employed simultaneously.

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.

The freeleech measures from Table 6, as well as the wealth measures (ratio, buffer, and minimum buffer), are highly correlated: the first five measures in Table 6 all have correlations of 0.95 or higher among themselves, the last four measures have correlations of at least 0.99 among themselves, and the two buffer measures have a correlation of 0.99. To avoid collinearity in the first-stage regressions, I only use one variable from each group: *flavgcopyh*, *flsumh*, *ratio*, and *bufmin*.¹³ To quantify the relationship between these measures and downloading activity, I regress $\ln(d_{it})$ on *flavgcopyh*, *flsumh*, *ratio*, *bufmin*, and album-level fixed effects. I obtain estimates for the full matched sample, as well as for each artist tier in isolation. Results are reported in Table 7.

The freeleech measures exhibit very strong correlation with downloads, and the magnitude of the effect is similar across all artist tiers. The ratio and buffer measures, however, vary in their effects by tier. I cannot find

¹³Results are similar under different choices of instruments; see Appendix A.

evidence that a positive wealth shock affects downloading of albums, with the exception of tier 4 albums.¹⁴ I retain both measures as instruments, however, since the coefficients may be jointly significant and at worst would just add noise.¹⁵ Regardless of sample, the measures explain about a quarter of the variation in an album’s downloading activity.

For each measure, I test the hypothesis that its effect is equal across all tiers using a Chow test. Results are reported in Table 8. The tests provide evidence that the effects differ across artist tiers. I take this heterogeneity into account in my second-stage estimation in Section 4.3.

4.3 Results

To estimate (1), I employ the two-step system GMM variety of the Arellano–Bond dynamic panel estimation technique.¹⁶ In all specifications, I include album fixed effects and weeks since the album’s release, as well as genre and time dummies. I report results in Table 9 for three specifications: one each for total, physical, and digital sales. I present estimates of the coefficients in equation (1) as well as the CME for each specification.¹⁷ I find that a transient (*i.e.*, one-period) percentage increase in file sharing results in a 0.13% decrease in overall sales for a given album, with a 0.17% decrease in physical sales and a 0.21% increase in digital sales. These results support the hypothesis proposed in Section 2 that the word-of-mouth effect is stronger for digital sales than for physical sales. In fact, the results imply that file sharing will actually increase digital sales on net. Still, the decrease in physical sales is larger, yielding a negative effect overall.

As shown in Section 4.2, exogenous changes in the determinants of file sharing differ in effect depending on an artist’s 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 10. File sharing seems to harm physical sales but help digital sales of tier 1 artists, while hurting tier 4 artists regardless of format. It is hard

¹⁴A regression using the full downloading sample produces significant coefficients, but I do not use them due to sample selection issues.

¹⁵Also note that the second-stage results presented below do not change if ratio is omitted.

¹⁶For a full description, see Roodman (2006).

¹⁷Standard errors on the CME estimate are computed using the delta method.

to draw inference about file sharing’s effect on mid-tier artists, but results suggest there may be a positive impact. The economic magnitudes of these effects are much larger for less popular artists. The results provide mixed evidence for idea that file sharing helps lesser-known artists through exposure. If we assume that tier 3 artists are obscure yet talented but tier 4 artists are obscure because they lack talent, then the results show that file sharing helps talented artists find a market at the expense of less-talented ones: an asymmetric information friction is ameliorated.

To further support the legitimacy of the instrument set, I calculate Hansen tests of overidentifying restrictions for the model with file sharing instruments and without in each specification. If the difference of these Hansen statistics is large, one must reject the null hypothesis that the overidentification restrictions are valid; *i.e.*, one must conclude that *flavgcopy*, *flsumh*, *ratio*, and *bufmin* are not jointly exogenous.¹⁸ I report *p*-values from that test as ΔH *p*-values in each table, and in no case can I find any evidence to doubt the exogeneity of my instrument set. Strictly speaking, this test does not establish instrument validity, but it does represent a lack of evidence against the instruments.

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

5 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 modest negative effect on legitimate music sales. This relationship varies by medium: file sharing decreases sales of physical copies but boosts sales of digital ones for top-tier artists, suggesting that the word-of-mouth effect is most relevant for the digital market. The effect varies by artist popularity as well. Top-tier artists lose sales,

¹⁸For a full treatment of this test, see Hayashi (2000).

but the loss is partially offset by an increase in digital sales and the overall effect is small. Mid-tier artists are helped slightly and bottom-tier artists are significantly hurt by file sharing, which could indicate that file sharing helps lesser-known artists only if they are actually talented. The results are robust to different instrumentation strategies 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 digital music, and less relevant for physical copies.

Piracy not only affects the market for music, it affects markets for all media that can be digitized: television, movie, 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 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 movies or television, digital distribution should be a primary focus for film studios and television networks as well.

A Specification & Robustness Checks

Alternative Instrument Sets

Since I have multiple freeleech and user wealth measures, it could be that my results depend crucially on which of those measures are included in the instrument set. This is not the case. I have recalculated the estimates in Tables 9 and 10 using two alternative instrument sets. The first simply uses all nine available freeleech measures and all three available wealth measures. The sample size is large enough that instrument proliferation is not an issue, and all results are quantitatively similar to the ones reported in Section 4.3. For the second alternative instrument set, I perform a principle component analysis on the available measures and extract the four components with the largest eigenvalues. Together, these components account for more than 99% of the variation in the freeleech and wealth measures. Again, the results in Section 4.3 are quantitatively similar. I can find no evidence against instrument validity in either case, just as in Section 4.3. Hence I am confident the instrumental variables used are valid, and results are robust to the choice of specific instrument sets. Tables of regression results are available upon request.

Alternative Lag Structures

A substantive objection to the proposed model is that equation (1) does not fully account for the dynamic effects inherent in the sales–piracy relationship. The problem is twofold: first, album sales may not decay geometrically, but instead through a higher–order autoregressive process. Second, there is no reason to believe that the direct effect of downloads only acts once and only contemporaneously; perhaps the effect of previous weeks’ file sharing affects current album sales, apart from the autoregressive sales process. The word–of–mouth effect may not fully propagate through the social network in a single period, and the crowd–out effect may take time to fully manifest as new consumers begin to share their files over the following weeks.

If one is willing to assume that the effect of file sharing on sales decays at the same rate as sales themselves (*i.e.*, that the effect of increasingly lagged file sharing is δ , $\alpha\delta$, $\alpha^2\delta$, etc.), then one can justify equation (1) as the result of a Koyck transformation of a more sophisticated model with longer–run effects of file sharing. The reported estimates of CME in Tables 9 and 10

are consistent with this interpretation. If one is not willing to make such an assumption, however, then a more complex model should be considered.

To investigate the importance of alternative lag structures, I generalize the model to include P lags of album sales, contemporaneous downloads, and Q lags of downloads. I expand my instrument set to include Q lags of the instruments as well. This yields an autoregressive distributed lag process of order (P, Q) :

$$\ln(s_{it}) = \sum_{p=1}^P \alpha_p \ln(s_{i,t-p}) + \sum_{q=0}^Q \delta_q \ln(\hat{d}_{i,t-q}) + \beta a_{it} + \gamma g_i + w_t + u_i + \epsilon_{it} \quad (3)$$

Selecting the “right” values of P and Q is not trivial: if a candidate (P, Q) is unsatisfactory, should P or Q be decreased first? I used the following general-to-specific selection procedure, starting with $(P, Q) = (15, 15)$:

1. Evaluate the model for the current P and Q .
2. If the p -value for either α_P or δ_Q is larger than some predefined threshold \bar{p} , then determine which p -value is larger, reduce that lag parameter (P or Q) by one, and return to step one.
3. If the p -value for neither α_P nor δ_Q is larger than \bar{p} , the current (P, Q) is selected.

This algorithm whittles down the model by eliminating the least-significant final lag until the final lag for both sales and downloads is significant at level \bar{p} , and a lower \bar{p} results in a smaller model. Table 11 lists the model that is selected by this algorithm for various \bar{p} and sales variables.

In the interest of space, I present results for significance threshold $\bar{p} = 0.01$ in Table 12.¹⁹ All coefficients are significant, and the δ_q are of alternating sign. This is not unexpected: downloads exhibit a secular trend, are thus somewhat collinear, and will therefore oscillate around some midpoint through a kind of mean-reversion. One should not conclude, then, that an exogenous increase in file sharing will increase current sales, decrease them next week, increase them a week later, etc: the coefficients cannot be interpreted in isolation. As before, the appropriate measure is the cumulative marginal effect (CME) of an increase in file sharing in period t on sales in

¹⁹Full results for all models up to (15,15) are available by request.

period t and all future periods. Each specification's CME is reported in Table 12 as well. The cumulative effects are qualitatively similar to those in Table 9, but the magnitude is larger for total and physical sales. Overall, the simplicity of model (1) may somewhat underestimate the magnitude of the effects, but the qualitative impact is unchanged by allowing for a more sophisticated lag process.

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.
- 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.
- 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.
- Hayashi, Fumio (2000), “Econometrics. 2000.” *Princeton University Press. Section*, 1, 60–69.
- 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/econ/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.” *CEISifo 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 Waldfoegel (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*.
- Waldfoegel, Joel (2010), “Music file sharing and sales displacement in the itunes era.” *Information economics and policy*, 22, 306–314.
- Waldfoegel, Joel (2011a), “Bye, bye, miss American pie? the supply of new recorded music since Napster.” Technical report, National Bureau of Economic Research.
- Waldfoegel, 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

Statistic	Mean	Percentiles		
		25th	50th	75th
Sales per album	4,552.30	1,201.65	1,667	2,770.83
Sales per album per week	656.84	8.09	50.85	330.07
Sales per week (millions)	4.68	3.72	3.92	5.00
Digital share per album per week	18.7%	2.2%	9.4%	24.2%
Age in weeks since release	149.43	6	30	177
Sales (millions)		131.09		
Digital share		11%		
Albums		2,875		
Weeks		28		

Table 1: Summary Statistics for Sales

Tier	Best Rank	Artists	Share	Albums	Share
1	1–82	358	25%	840	37%
2	83–271	356	25%	555	25%
3	272–576	360	25%	460	20%
4	577–999	358	25%	396	18%

Table 2: Division of Albums into Sales Tiers

Statistic	Mean	Percentiles		
		25th	50th	75th
DLs per album (all)	19.27	0	4	13
DLs per album (active)	26.14	3	7	19
Weekly DLs	188,864.89	113,278	139,602	172,541
Weekly DLs per album (all)	0.91	0.66	0.68	0.74
Weekly DLs per album (active)	3.63	2.75	2.87	3.12
Weekly active albums	24%	22%	24%	26%
File size of album in MB	175.72	61.98	96.80	223.95
Weeks in database, final week	23.26	9	19	36
New files per week	5,250.69	3785	4007	4482
New albums per week	4,726.54	3206	3607	4776
Downloads	5,099,352			
Albums	264,672			
Share of albums active	73.7%			
Weeks	27			

Table 3: Summary Statistics for Downloads

Week	All	Matched
14	0.73%	0.50%
15	9.76%	3.30%
16	100%	100%
21	0.89%	2.49%
22	1.53%	3.11%
All Others	0%	0%

Table 4: Share of Albums on Freeleech by Week

Statistic	Mean	Percentiles		
		25th	50th	75th
Weekly downloads per album	18.94	1.54	5.26	15.82
Weekly sales per album	3,578.43	157.25	712.19	2091.54
Downloads per week	27,920	17,091.5	21,187	30,224
Sales per week (millions)	4.29	3.44	3.62	4.17
Digital share per album per week	20.9%	5.4%	12.5%	27.6%
Number of downloads	781,774			
Number of sales (millions)	120.23			
Digital share	12%			
Number of albums	2,251			
Match rate, file sharing data	0.9%			
Match rate, sales data	78.3%			

Table 5: Summary Statistics for Matched Albums

Variable	Description
<i>fl</i>	Freeleech dummy
<i>flpct</i>	Percent of formats on freeleech
<i>flpctcopy</i>	Percent of shared copies on freeleech
<i>flavgh</i>	Average format-hours on freeleech
<i>flavgcopyh</i>	Average copy-hours on freeleech
<i>flsum</i>	Number of formats on freeleech
<i>flsumcopy</i>	Number of shared copies on freeleech
<i>flsumh</i>	Number of format-hours on freeleech
<i>flsumcopyh</i>	Number of copy-hours on freeleech

Table 6: Various Freeleech Measures

	All	Tier1	Tier2	Tier3	Tier4
<i>flavgcopyh</i>	0.0230*** (0.0000)	0.0230*** (0.0000)	0.0249*** (0.0000)	0.0208*** (0.0000)	0.0204*** (0.0000)
<i>flsumh</i>	0.0015*** (0.0000)	0.0016*** (0.0000)	0.0012*** (0.0000)	0.0019*** (0.0000)	0.0020*** (0.0000)
<i>ratio</i>	0.0547 (0.2898)	0.1114 (0.2404)	0.1286 (0.1896)	0.0117 (0.9174)	-0.1076 (0.3006)
<i>bufmin</i>	0.0025 (0.4800)	-0.0058 (0.3874)	-0.0031 (0.6549)	0.0046 (0.5156)	0.0240*** (0.0021)
Observations	37,333	13,600	9,449	7,649	6,635
Albums	2,198	823	546	446	383
Within R^2	.26	.26	.28	.24	.27
Joint wealth p -value	.14	.5	.38	.67	.0073

p -values in parentheses

Standard error estimates are cluster-robust.

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

Table 7: Estimation of First-Stage Relationship

Variable	<i>flavgcopyh</i>	<i>flsumh</i>	<i>ratio</i>	<i>bufmin</i>
Chow p -value	0.01	0.01	0.32	0.02

Table 8: Chow tests; H_0 : effect is equal across all sales tiers

	Total		Physical		Digital	
	Coef.	CME	Coef.	CME	Coef.	CME
Downloads	-0.03** (0.02)	-0.13** (0.04)	-0.04*** (0.00)	-0.17*** (0.01)	0.07** (0.01)	0.21*** (0.01)
Lagged sales	0.76*** (0.00)		0.78*** (0.00)		0.65*** (0.00)	
Age	-0.00 (0.27)		-0.00 (0.26)		0.00 (0.44)	
Observations	16,819		16,806		15,556	
Albums	1,588		1,581		1,463	
ΔH p -value	0.9796		0.9130		0.9823	

p -values in parentheses

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

Standard errors calculated using Windmeijer correction.

All specifications include album, week, and genre controls.

Table 9: Estimation of (1).

	Total		Physical		Digital	
	Coef.	CME	Coef.	CME	Coef.	CME
Tier 1 DLs	-0.04** (0.04)	-0.14* (0.06)	-0.05*** (0.00)	-0.21*** (0.01)	0.07** (0.03)	0.20** (0.01)
Tier 2 DLs	-0.02 (0.39)	-0.08 (0.40)	-0.01 (0.70)	-0.04 (0.70)	0.05 (0.29)	0.14 (0.28)
Tier 3 DLs	0.07* (0.08)	0.28* (0.07)	0.05 (0.18)	0.24 (0.18)	0.11* (0.06)	0.30* (0.06)
Tier 4 DLs	-0.26*** (0.00)	-1.01*** (0.00)	-0.26*** (0.00)	-1.15*** (0.00)	-0.22* (0.10)	-0.59 (0.10)
Lagged sales	0.74*** (0.00)		0.77*** (0.00)		0.63*** (0.00)	
Age	0.00 (0.92)		-0.00 (0.79)		0.00 (0.10)	
Observations	16,819		16,806		15,556	
Albums	1,588		1,581		1,463	
ΔH p -value	0.9415		0.9939		0.7123	

p -values in parentheses

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

Standard errors calculated using Windmeijer correction.

All specifications include album, week, and genre controls.

Table 10: Estimation of (1) with Artist Tiers.

\bar{p}	All Sales	Physical Sales	Digital Sales
0.10	(10,9)	(8,10)	(11,7)
0.05	(10,9)	(8,10)	(10,7)
0.01	(1,3)	(1,3)	(1,2)

Table 11: Selected Models for Various Significance Thresholds

Lag	All		Physical		Digital	
	Sales	DLs	Sales	DLs	Sales	DLs
0	—	0.03** (0.05)	—	0.04** (0.02)	—	0.09*** (0.00)
1	0.88*** (0.00)	-0.07*** (0.00)	0.88*** (0.00)	-0.06*** (0.00)	0.71*** (0.00)	-0.08*** (0.00)
2	—	0.03** (0.02)	—	0.03** (0.02)	—	0.06*** (0.00)
3	—	-0.05*** (0.00)	—	-0.05*** (0.00)	—	—
CME	-0.39** (0.02)		-0.32** (0.03)		0.22** (0.02)	
Observations	12,075		12,074		12,488	
Albums	1,235		1,234		1,251	
ΔH p -value	0.78		0.54		0.67	

p -values in parentheses

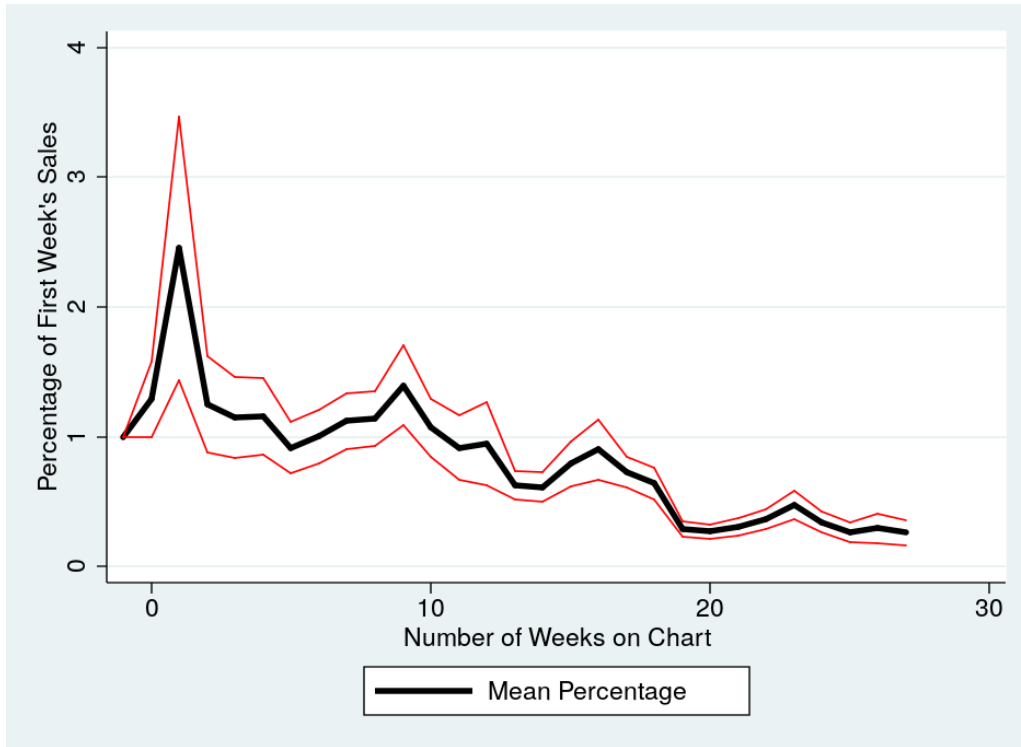
* $p < .1$, ** $p < .05$, *** $p < .01$

Standard errors calculated using Windmeijer correction.

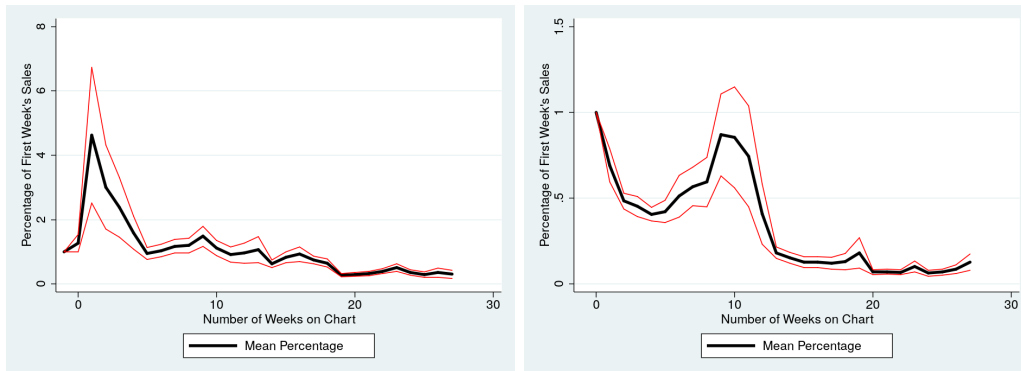
All specifications include album, week, and genre controls.

Table 12: Estimation of (3) for Selected ARDL(P,Q) Models.

Figures



(a) Total Sales



(b) Physical Sales

(c) Digital Sales

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

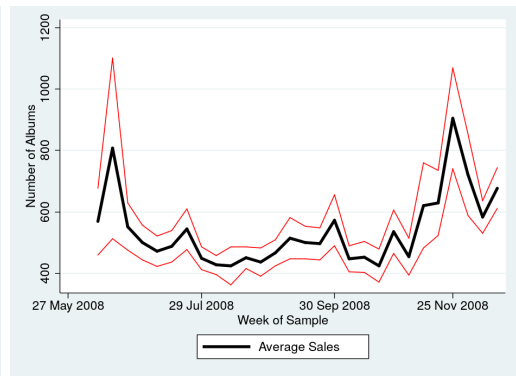
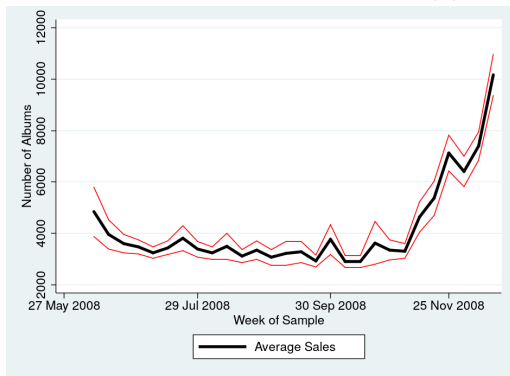
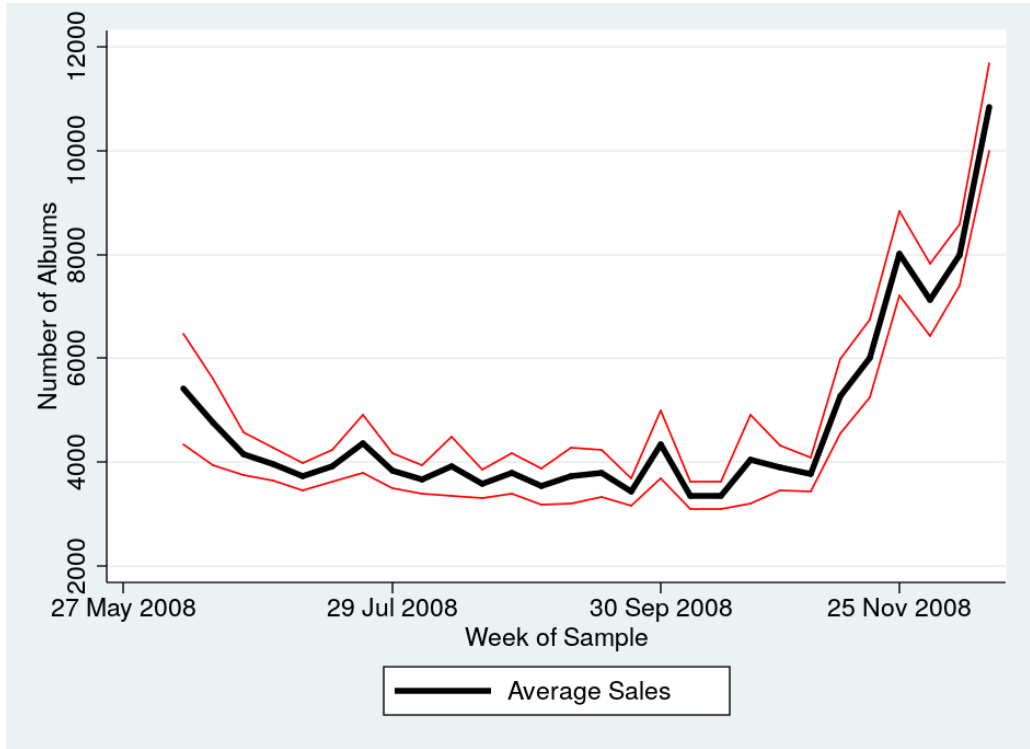


Figure 2: Average Album Sales per Week

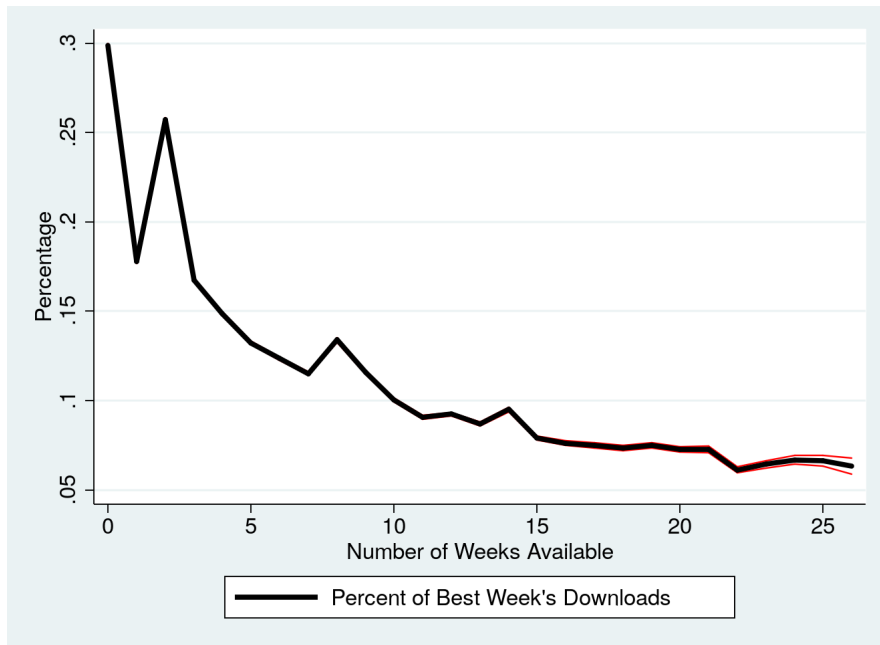


Figure 3: The Average Downloading Lifecycle of an Album

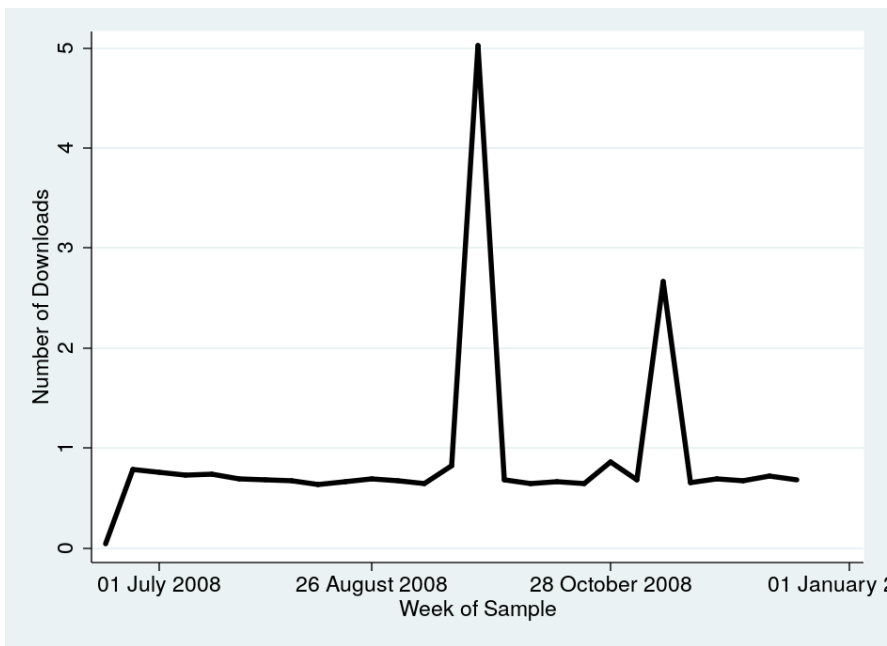


Figure 4: Average Downloads by Week of Observation

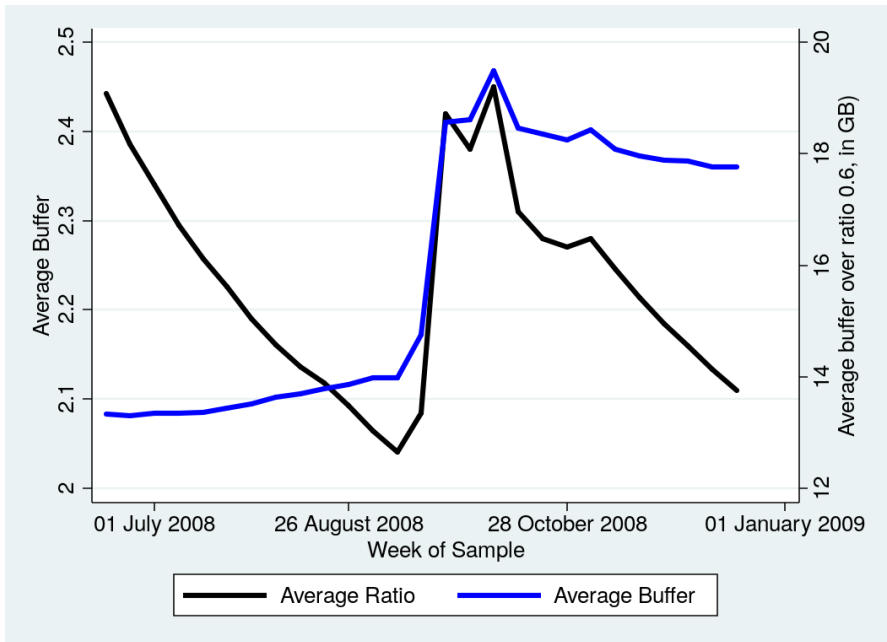


Figure 5: Average Ratio and Median Buffer

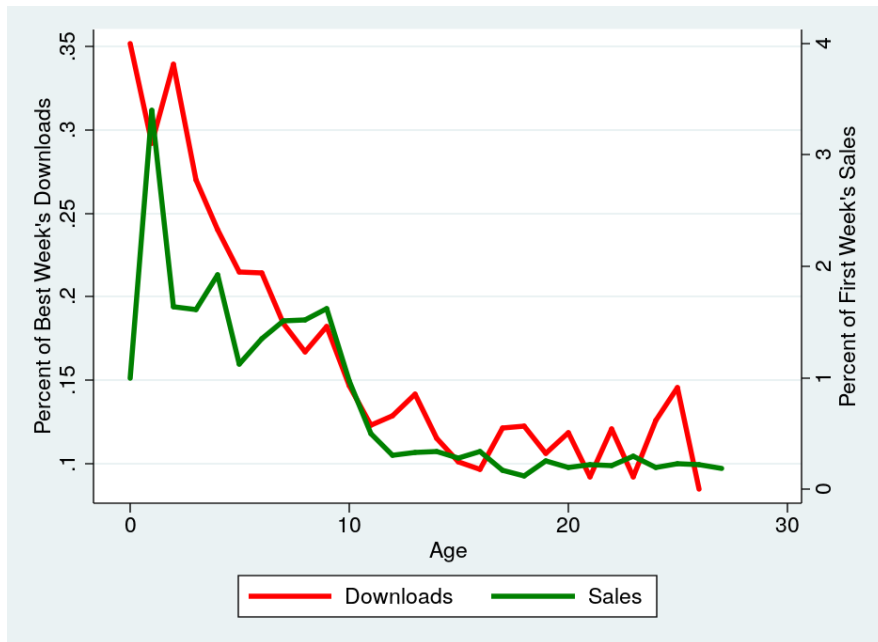


Figure 6: The Life-Cycle of a Matched Album

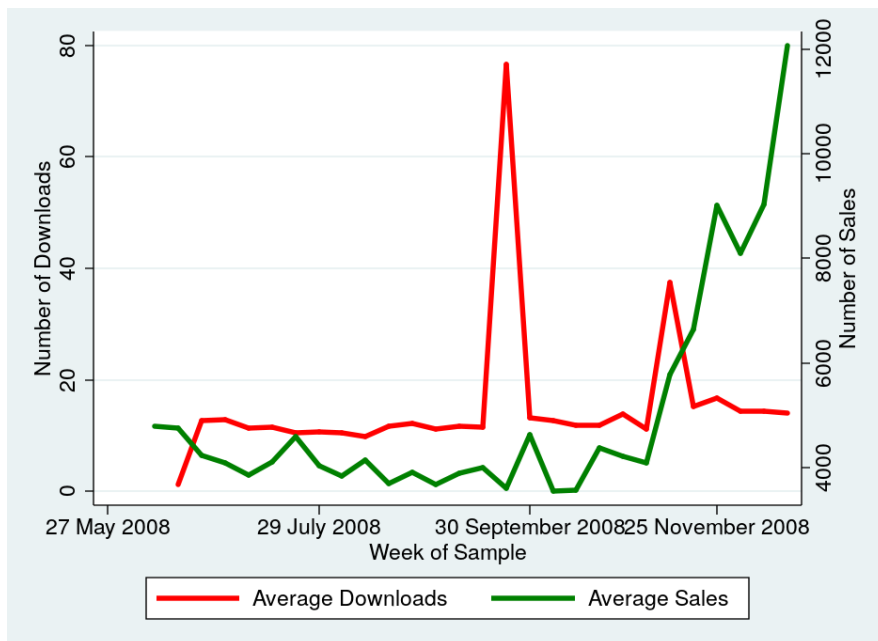


Figure 7: Average Consumption of a Matched Album