Private Filesharing Markets and Album Sales

Jonathan Lee*

April 1, 2009

Abstract

Filesharing is a growing and powerful phenomenon that affects many legitimate markets, especially that of recorded music. Many economists have tried to estimate filesharing's effect on music sales with varying conclusions and degrees of success. We find that all of these studies have become obsolete and that it is necessary to approach this question again using new datasets. Using album-level sales and filesharing data, we investigate how private filesharing networks function as markets and what effect they have on legitimate music sales. We find evidence of relative price substitution, quality substitution, consumption-smoothing, and demand-side price inelasticity in private filesharing markets. We find that exogenous downloads in a particular filesharing network are good leading indicators of additional legitimate music sales.

The market for music is constantly and rapidly evolving. Genres emerge, artists come and go, but one thing that had not changed until recently was the method of distribution. The vinyl record, cassette tape and CD were synonymous with music ownership. Today, that trend is changing. Consumers have digital music players in their pockets that hold thousands of songs. Even diehard audiophiles have made the transition from physical to digital.

This would all be well and good, were it not for the fact that this new distribution method is largely illicit. These digital copies of music are downloaded from the Internet, often directly from their peers with no compensation to artists or producers. Thus it is natural to wonder how this activity affects legitimate music sales: do consumers substitute away from physical media, or do they value the disc and the packaging too much? Is filesharing really seen as immoral to those involved, or is it seen as fundamentally different from stealing? Does effectively lowering the price to zero bring in new consumers that might then purchase a physical copy? Or is physical media a dying breed?

Economists have attempted to answer this question before, with varying degrees of success. Liebowitz (2005) provides a thorough, though not unbiased,

*I thank the following for their special contributions: Amanda Chuang and Grant Ramsey for their invaluable help in data formatting; John Siegfried, Tong Li, John Conley, and John Weymark for their advice and suggestions; the administrators of TorrentSite.net for their generosity; and Victor Helmsing for stimulating my interest in the subject in the first place.

review of the existing literature on filesharing. As he reports, most papers find a negative effect on the record industry due to filesharing. These include studies by Rob & Waldfogel (2006), Zentner (2006), Hong (2004), and Peitz & Waelbroeck (2004). Using different sections of data and different methodologies, all of these studies come to the conclusion that filesharing has been detrimental to album sales.

One heavily-publicized study, Oberholzer & Strumpf (2004, 2007), however, finds a relationship between album sales and downloads that is "statistically indistinguishable from zero," and proposes a sampling effect as a possible rationale for this result. What makes this study more than an outlier is the fact that the authors use actual filesharing data collected from a popular network online at the time. The previously mentioned studies instead rely on surveys or agglomerated data. Naturally, this contradictory study has incited controversy, and Liebowitz (2007) has dedicated an entire paper to critiquing Oberholzer's & Strumpf's work.

However interesting these studies are, they examine an outdated market. Filesharing is not what it once was. Methods are more reliable, download speeds are quicker, and networks are larger than ever before. Instead of downloading individual songs, filesharers now download entire discographies. All previous analyses known to the author were performed with now-obsolete data.

This paper seeks to investigate the nature of the filesharing-sales relationship under new filesharing technology, using datasets from 2008. These datasets are analogous to those used by Oberholzer & Strumpf, as they record filesharing data and album-level sales data directly. This provides a more concrete analysis than what is provided by an agglomerated dataset.

1 Filesharing: A Primer

1.1 A Brief History

Ever since the days of the mix tape, music sharing in one form or another has been commonplace. With the advent of the compact disc and digitized music in the mid-1980s, music sharing activity began to accelerate, given the contemporaneous adoption of personal computing. The real filesharing movement as we know it, however, began in 1999 with the arrival of Napster, one of the first matchmaking services for filesharing.

Napster used a central indexing server that would list all available files at any given time. A user would log in and send a list of files he had available to the server, which listed these files centrally. If a user wanted a particular file, he would search for it and would receive a list of users with files matching his query. Selecting one, Napster would connect the two users, and they would transfer files directly between themselves.

This method was not perfect, however. Search queries matched filenames, which did not necessarily correspond to what the file actually was. "Spoofing", the act of listing bad files with good names, was commonplace, and there was no guarantee that after your download completed you would have the file for which you were searching. Additionally, since the file transfer was directly between users, if the uploader (sender) disconnected, the downloader (receiver) was left with an incomplete and unreadable file. The downloader was forced to start the process over. Finally, given dial-up connection speeds and storage restrictions of the era, it was difficult to impossible to compile an entire album of the same quality. The majority of files available were radio singles and popular tracks, not B-sides or less well-known album tracks, and rarely would an uploader stay online long enough for a downloader to receive an entire album.

However, the biggest disadvantage to the Napster system revealed itself in 2000-2001, when Napster was found liable for contributory infringement of various artists' copyright. This ruling was due in no small part to the centralized configuration of Napster's network. A new generation of networks spawned, but these decentralized the file index process so that users' computers performed the function of the central server. These servers did not have the legal liability issues that plagued Napster, but the problems of incomplete, corrupt, and slow downloads were still present.

1.2 BitTorrent

Decentralized networks were ubiquitous during the early and mid-2000s, and some are still around today. However, they are quickly giving way to a new file-sharing protocol: BitTorrent. BitTorrent solves the problems of older networks in novel ways, but first we need some terminology. A *seeder* is a user who is currently uploading data, a *leecher* is a user who is currently downloading, a *torrent* file acts as a pointer to the actual file being distributed, a *snatch* is a download of a torrent file, a *tracker* is a network that keeps track of what users have which files and is where users download torrents, and a *client* conglomerates files from different seeders into a full file. The typical sharing process is as follows: a user searches the tracker for a specific file and snatches a matching torrent. The user's client reads the torrent and finds seeders, choosing based on availability and connection speed. It then begins downloading parts of the file from different seeders and builds the complete file from these parts. Once the leecher has a complete copy of the file, he becomes a seeder and the process is repeated.

This method has several advantages. First, all the seeders are in a "swarm"; if one disconnects there are still several to choose from. If somehow all of the seeders disconnect, the download simply pauses until another signs on, and does not have to be restarted. Second, since all of the users have the same file, the risk of downloading corrupt or low-quality files is greatly diminished (users tend to self-police in terms of quality). Third, torrent files can actually point to a family of files, allowing the download of complete albums by the download of one file. Finally, this method purportedly jumps through a few legal loopholes. Since the trackers are not actually hosting copyrighted material, only torrents, they are not liable. Ostensibly, no one user is uploading a complete copy of a file, so it is unclear if this counts as full-on piracy.

The communal nature of BitTorrent lends itself to the creation of insular networks. Indeed, many of the most reliable networks are members-only, and require an invitation to join. Failure to contribute to the community (i.e., seeding often) can result in banning, and the userbase does a great job of supporting quality files and removing poor ones. These private networks effectively act as a black market for shared files, and their dynamics are discussed in Section 1.3.

1.3 Private BitTorrent Network Dynamics[1]

The interaction between legitimate, legal music markets and filesharing networks is essentially one of substitution. In order to investigate these interactions, we first need to understand the filesharing market as a standalone unit.

Markets can be classified by the good traded and the medium of exchange (or lack thereof): a father pays dollars for a basketball, a goat farmer trades livestock for assistance in building a barn, etc. In private filesharing networks, the good traded is access to a copy of a file, and the medium of exchange is file size in MB.[2] Measures of wealth can be defined in a few different ways. A user's *ratio* is defined as his amount uploaded divided by amount downloaded, his *buffer* is amount uploaded minus amount downloaded, and his *target ratio buffer (TRB)* is his ratio minus some target ratio (determined by the user's tastes). These measures of wealth differ in applicability and implementation, but all are meant to measure how active a given user can be in a market.

Filesharing networks are markets for public goods, since consumption of a downloaded file is nonexclusive. As such, these networks are often plagued by free-riding problems. Private filesharing networks usually address this problem by requiring users to maintain a minimum ratio. Thus any downloading must be matched by a certain amount of uploading, keeping supply active. From this, it seems that a high ratio would indicate a wealthy user, one that can be relatively more active in a market.

This is not necessarily the case. A user with 100MB uploaded and 10MB downloaded cannot be as active as a user with 1000MB uploaded and 500MB downloaded, even though the first's ratio is five times larger than the second's. Thus a more reliable measure of wealth here is his buffer, which explicitly states how much a user can download before having a ratio of one.

This, however, is still a naïve measure of wealth. As shown by oorza[1], many users do not strive towards a ratio of one. Some will strive to exactly match the network's minimum ratio, which is often less than one. Others, as a matter of pride or to save for the future, will strive for a higher ratio. This number is the user's *target ratio*, and when subtracted from his ratio becomes a decent measure of wealth, but it does not explicitly measure a user's potential economic activity.

The most concrete measure we can formulate to predict a user's activity on a network is his *target-adjusted buffer (TAB)*, defined as

[1] Adapted and extended from "On Rationomics", a forum post by a user named oorza.

[2] MB = megabytes, or 1,048,576 bytes. This is the measure of file size that will be used throughout the paper.

$$uploaded - (target\ ratio \cdot downloaded)$$

This reports how much data a user can download (buy) and still not feel inadequate or deprived of resources. Of course, this measure is difficult to pin down; every user's target ratio is different. However, as seen below in Section 3, we can define a measure of typical wealth in a network and show that it is correlated with activity on the network.

Now that we have identified TAB as a measure of wealth for filesharing networks, we walk through a typical exchange of goods on a network. A user finds a torrent to snatch, and as he leeches, his TAB decreases with his downloaded amount. Simultaneously, the participating seeders' TABs are increasing with the amount they upload. The leecher sacrifices some of his currency to obtain a file, and the seeders stay online and active in order to reap more TAB.

Other events can influence how much TAB a user has. We identify three: freeleech periods, the bounty system, and ratio requirement changes.

Freeleech periods are times, designated by the administrator of the tracker, during which leeching does not affect a user's downloaded amount but does affect uploaded amounts. These periods typically last from a few hours to a week. Thus during a freeleech, users' TABs can only rise, and everyone is made wealthier. Afterwards, users spend their excess TAB, and activity increases. Freeleech periods can be network-wide, restricted to a set of files, or be partial freeleeches (e.g., download costs are cut in half).

The bounty system is a method used by some trackers to stimulate growth. Users identify a file they'd like to have but which is not currently on the network. Along with this, users post a "bounty" (in MB) that signals how much they value this currently unavailable file. Suppose this file is an album. If another user has a hard copy of this album, he can upload it and reap the benefits. These benefits include a transfer of some or all the bounty currently posted for the file. This transfer of wealth rewards new content, keeping the network current and fresh. Additionally, if these transfers are taxed (the uploader only receives, say, 90% of the total bounty), administrators can remove currency. This could be useful if users are too wealthy; they have accumulated too much TAB and feel no need to seed files, leading to an empty network.

Network ratio requirements state how low a user's ratio can be before the user is banned and are a large determinant of a user's target ratio. At the very least, the value provides a lower bound on every user's target. Changing requirements therefore changes network activity, especially if target ratios are a positional good and the minimum allowed ratio is seen as something of a last place. If the minimum ratio increases from .5 to .8, a user's target ratio may rise from 3.0 to 3.5 simply to stay ahead of the curve. This lowers the user's wealth.

Freeleech periods and other such shocks (there are many) change users' TABs and therefore influence their propensity to download any given file. It is presumed that these shocks do not directly affect uploads and downloads except through the income effect of a change in TAB, an assumption used later in the analysis.

2 Motivations

There are plenty of questions to be asked here. We will attempt to answer the following:

First, how do simple economic ideas like supply and demand, substitution effects, and income effects affect downloading on private torrent sites? File size can vary greatly between different torrents for a particular album, even after controlling for quality. Do users care? In the same line, do differences in sound quality for a given album induce users to download the higher-quality version? Finally, are downloads more prevalent when users have a lot of TAB to spare, or do they "hoard" ratio so that they can consumption-smooth? We answer these questions in Section 4.1.

Second, how do downloads of popular albums on private networks affect sales in legitimate markets? Do downloads displace sales or bring new consumers into the market? Are there substitution effects or sampling effects as proposed by Oberholzer & Strumpf? Does the original downloader tell his friends and create new consumers? If there is an effect on album sales from download, how long does it take to happen? Is it instantaneous or is there a lag between hearing about new music and actually purchasing it? We answer these questions in Section 4.2.

Of course, to answer these questions empirically, we need robust datasets. We have managed to obtain data, and they are described in Section 3.

3 Data

Access to data on filesharing phenomena is difficult to obtain, especially for private networks. Privacy concerns are a huge issue with members of these networks, and potentially incriminating information is well-guarded. Album sales data are restricted as well. The industry standard, the Nielsen SoundScan dataset, is proprietary and is very expensive to obtain. Other data sources are less reliable, considerably smaller, or only report rankings and not actual levels of sales. However, we have been able to overcome these hurdles and have access to formidable datasets with which to test hypotheses.

3.1 Filesharing Data

The dataset obtained for filesharing data is collected from a tracker we will call TorrentSite.net.[3] As of this writing, TorrentSite.net has over 90,700 enabled users, 89% of which have been active in the past month. More than 453,000 torrents are hosted for 250,000 albums and 119,000 artists, and there have been

[3] Per agreement with the site administrators, the site's name has been changed for privacy reasons. However, the dataset itself is available upon request.

14.2 million snatches since the site's inception (October 29, 2007). The site administrators have been kind enough to provide weekly data on all hosted torrents and anonymous user statistics, compiled automatically at 12:01 AM each Friday from June 20, 2008 to December 19, 2008.

User Data

User data are collected weekly from TorrentSite.net, and each set reports the following for each user: amounts uploaded and downloaded to date, whether the user has donated monetarily, and whether the user's membership is pending, enabled, or disabled. These data are anonymous, so we cannot track individuals from week to week. We can, however, investigate trends in the user base over time. These trends are shown in Figure 1. Week 14 sees a huge increase in both the mean ratio and buffer. This is no coincidence; a site-wide, week-long freeleech began at the start of week 14. We also see a general downward trend of both measures. This suggests that after users experience a positive income shock, they may download more to return to their target ratio. Approximately 5% of the user base has made a cash donation to the site; ratios and buffers for these users differ insignificantly from the whole.

Torrent Data

Torrent data are collected concurrently with user data from TorrentSite.net, and each set reports the following for each torrent hosted: artist name, album name, genre, file format, encoding, size, number of snatches to date, and date first uploaded. The snatch data are first differenced for the analysis. Summary statistics for the pooled dataset are reported in Table 1. On average, 27% of all available torrents were snatched in a given week.

3.2 Sales Data

Sales data are collected from HitsDailyDouble.com, an online magazine that compiles sales data for the top 50 selling albums each week. These figures include traditional brick-and-mortar retailers as well as online stores such as iTunes. Figures are compiled each week, and are finalized on the Tuesday of that week.[4] Each week reports last week's position, this week's position, artist name, album name, record label, gross sales, and percent change of sales from the previous week. Summary statistics are reported in Tables 2.

Sales numbers are fairly persistent over time. Using OLS, we estimate the coefficients in $Sales_t = \beta_0 + \beta_1 \ Sales_{t-1} + \varepsilon_t$ and report the results in Table 3. From these estimates, we see that a large portion of the variation in top 50 album sales can be explained by persistence.

Seasonal trends can also be discerned. Figure 2 plots sales for the median album during the sample period. We see median sales deviating from trend in mid-November. Presumably, holiday purchases account for this deviation. Note

[4] This is so done because most new albums drop on Tuesdays.

that this graph is very similar for other positions on the chart, but chart-topping albums exhibit wild deviations and almost no discernible trend.

4 Hypothesis Testing

4.1 Price & Quantity Demanded, Income Effects

Though TAB is a measure of wealth, it is not the prevailing measure of currency or the price of a download. To measure that, we use the actual size of a download, in MB. There are various hypotheses about ceteris paribus effects that we can test. First, we should expect that smaller files will be downloaded more frequently than larger files. Second, we should expect that higher-quality files will be downloaded more frequently than poorer ones. Finally, we should expect that the richer a user is, the more he will download.

It is necessary to discuss the quality of different file formats before we proceed. The two most prevalent file formats on TorrentSite.net are MP3 and FLAC, which make up 98% (73% and 25%, respectively) of the hosted files. MP3 is a "lossy" format, while FLAC is a "lossless" format. The distinction is a trade-off between size and quality. Lossy formats use compression technology that drop off certain pieces of the file while maintaining the gist of the information contained, while lossless formats faithfully reproduce every piece of information in the file. For music, the difference is undetectable unless one is using high-quality audio equipment. In our dataset, the mean MP3 file is 94MB, while the mean FLAC file is 376MB.

We test the first two hypotheses, that larger files are downloaded less and higher-quality files are downloaded more, in Table 4. Specification (1) regresses new downloads on a dummy specifying whether a file is FLAC. This proxies for file size[5]. We can see that FLAC files are, economically speaking, not significantly less downloaded than MP3s. However, when we control for file size in specification (2), we see that size has zero effect on downloads but that FLAC files are ever-so-slightly more downloaded than MP3s.

What Table 4 shows us is that much larger files are downloaded slightly less, but that higher quality files are downloaded slightly more. However, the coefficients on this equation are small enough that their effects are negligible. Economically, large and small files and high- and low-quality files are pretty good substitutes. Put another way, demand is very inelastic when we view file size as currency.

Another testable hypothesis concerns the effect of income shocks on downloads. Do changes in TAB influence the amount of new downloads significantly? We estimate the coefficients on regressions of new downloads on the mean user buffer and on the mean user ratio [6]. The results of these estimations are shown in Table 5. We see that an increase in these measures has no discernible effect

[5]We use FLAC instead of actual size here, because much of the variation in size is only a few MB for different kinds of MP3s. We want to see the effect of a larger change.

[6]We use these measures since TAB cannot be observed directly.

on new downloads in general. We will find later that this is not the whole story, however, and that substitution effects play a large part in downloading. For now, we can say that users seem to consumption-smooth; they download the same number of albums regardless of income effects.

4.2 Downloading's Effect on Sales

We seek to determine the partial effects of private-network downloads on the top 50 most popular albums' sales. We write

$$y_{a,t} = \beta_0 + \beta_1 y_{a,t-1} + \beta_2 d_{a,t} + \varepsilon_{a,t} \tag{1}$$

where $y_{a,t}$ is sales of album a in week t and $d_{a,t}$ is downloads of album a in week t.

Normally it would suffice to estimate (1) using OLS; however, there are endogeneity concerns. It is very likely that there are unobserved variables that influence both sales and downloads, such as artist popularity. Therefore we use instrumental variable techniques to estimate (1) and identify an instrumental variable for $d_{a,t}$. The natural choices are the measures of user wealth as described above, such as TAB. More practical choices include the mean user ratio and buffer as described in Figure 1. These are presumably exogenous in (1). We refer to these wealth measures as w_t and estimate the following instrumental equation

$$d_{a,t} = \pi_0 + \pi_1 w_t + \varepsilon_{a,t} \tag{2}$$

Results are reported in Table 6, with columns corresponding to different w_t. Since we can reject $H_0 : \pi_1 = 0$ at the 1% significance level for both measures, w_t can serve as an exogenous instrument for (1). We therefore estimate the coefficients in (1) using a two-stage least squares (2SLS) methodology, with w_t as the exogenous instrument. The results are reported in Table 7. We see that all coefficients are significant at the 1% level, with the exception of $\hat{\beta}_2$ when the mean user buffer is used as an instrument. However, this is likely due to the higher standard errors involved with IV techniques. At any rate, $\hat{\beta}_2$ is significant when the mean user ratio is used as an instrument. The 95% confidence interval for its point estimate is [14.17, 30.72].

These results are based on a zero-week lag; torrent data begin compiling on a Friday at 12:01 AM, sales data begin compiling on a Monday at 12:01 AM, torrent data finish the following Friday, and sales data finish the following Monday. Data in these two sampling periods are termed to be in the same period for the analysis. However, one could suppose that different lags would be more or less significant. Perhaps filesharing's effects are more pronounced after one or two weeks. To this end, equation (1) was re-estimated eight times for different lags, from -1 to 7. The results are reported in Figure 3. We see that the zero-week lag has the largest coefficient, though other lags are also significant. Much more data would be needed to see a full picture; presumably the estimates move around the x-axis until they settle there and become insignificant. There is a temptation to fit these estimates to a curve and integrate; the result would be misleading. There is yet again an endogeneity concern. The value of $\hat{\beta}_2$ with a three-week lag, for example, depends in the equation on downloads in week 0 and sales in week 3. However, sales in week 3 are also correlated with downloads in week 3. We would need to control for this effect in such an integration.

5 Results and Conclusions

We have investigated private filesharing network dynamics, both as a standalone market and in conjunction with legitimate markets for music. We summarize our findings below.

Private Network Market Dynamics

In the data, we have seen evidence of many elementary economic ideas functioning in private filesharing networks, including demand inelasticity, quality substitution, consumption-smoothing, and relative price substitution.

From Table 4, we conclude that considerably more expensive (large) files are statistically downloaded less, but that the economic effect is insignificant (specification 1). This shows us that users exhibit fairly inelastic demand for shared files, and this hints at consumption-smoothing by the same userbase.

From Table 4, we also conclude that higher-quality files (FLAC) are statistically downloaded more after controlling for file size, but that the economic effect is fairly insignificant (specification 2). This suggests that users substitute fairly freely between lower- and higher-quality files, but note that our conclusions here would likely be different if we could observe the quality of audio equipment that users own.

From Table 5, we see that an increase in average user wealth has no discernible effect on new downloads in net. This is evidence of consumption-smoothing, at least in aggregate: regardless of income and income elasticity, users seem to download the same number of files. However, as discussed below, consumption patterns *do* change with changes in average user wealth.

From Tables 5 and 6, we can learn a few more things about user downloading dynamics. There is an apparent contradiction between these tables. They estimate the same equations, but the first estimates a statistically nonexistent effect and the second estimates a highly negative effect. The key to understanding this difference is the different samples being used. Table 5's sample uses the entire universe of activity on TorrentSite.net, while Table 6 uses only the top 50 selling albums from each week in the sample. Simple substitution is occuring here, between more and less popular albums. When average wealth is high, there is more activity, and it becomes easier to obtain less popular albums due to a higher number of seeders and leechers. When average wealth is low, these albums are harder to obtain. However, in either case, popular albums are fairly easy to obtain. Thus when average wealth rises, the *relative* price of less popular albums to more popular albums falls, and users substitute into downloads of less popular albums. When average wealth falls again, it is harder to find less popular releases, but more popular ones are still available. Relative price rises, and users substitute into downloads of more popular albums.

Private Networks and Album Sales

From Table 7, we see that the correlation of exogenous filesharing and album sales on TorrentSite.net is statistically significant and positive. This seems counterintuitive; real sales and virtual downloads should be substitutes and thus one would expect a negative correlation. However, this does not seem to be the case. According to the value of $\hat{\beta}_2$, an additional download on TorrentSite.net corresponds to about 22 new sales in the marketplace. Obviously, one exogenously-motivated downloader would not have otherwise purchased 22 copies. Further, it is fallacious to assume that these regressions show that one new download directly causes 22 new sales in the marketplace. We have reined in endogeneity, but the instrument we have used is specific to TorrentSite.net. Changes in user ratios and buffers here will not directly influence downloading on another tracker. To establish a direct cause-and-effect relationship, we would need a unverse of filesharing data or a better instrument that affects all filesharing.

However, the regression is not spurious; real economic conclusions can still be drawn. Consider a user with a newly-released album who may upload to a public or private tracker. Because he need not upload in order to use the public tracker (ratio requirements are not an issue with public trackers), the incentives are likely higher for the user to upload to the private tracker. Thus new albums are much more likely to show up on private trackers like TorrentSite.net before other trackers. Now, those who download from private trackers have incentives to upload to other trackers, possibly public. In this way, new albums become available first on private trackers and then on public ones. Thus it is plausible that exogenous changes in downloading on private trackers have an effect on all downloading and thus all sales. Table 7 gives empirical evidence for this hypothesis. At the very least, we see that exogenous downloads on TorrentSite.net are a leading indicator of additional sales in the real world.

Now that a possible relationship between the dependent and independent variables has been established, we can interpret the coefficient. The point estimate, $\hat{\beta}_2 = 22.45$, is statistically significant and positive. Naïvely, we should expect that $\hat{\beta}_2 < 0$, in that downloading and album sales should be substitutes. This does not seem to be the case. Instead, a few processes are probably going on here. First, we likely see the results of a sampling effect *vis-a-vis* Oberholzer & Strumpf (2007). Second, downloads may represent word-of-mouth publicity; a user might tell his friends that some artist is fantastic and that they should also purchase the album. Third, and most importantly, a dissemination of shared files from TorrentSite.net to other trackers is occuring. This expands the effects of the first two processes so that they occur with millions of users instead of a few hundred thousand.

Policy Considerations

While the coefficient in question, $\hat{\beta}_2$, is statistically significant, its economic significance is another story. For the albums studied, 22 new sales represent .0005% and .001% of the mean and median album's respective sales in a given week (cf. Figure 2). These are infinitesimal proportions, and as such have no real economic impact. Thus we conclude that new illicit downloading on a private network has no discernable economic effect on legitimate music markets. This raises the question of whether or not we should allocate resources to forcibly shutting down private filesharing networks, arresting their administrators, and prosecuting their users.[7] If their activity has no real effect, and in fact serves largely as a method of advertising, fighting against these networks seems to be an inefficient use of resources.

Whatever the conclusion we draw, it is important to distinguish the effect filesharing has on album sales and on revenues to artists. The typical recording contract stipulates royalty rates to artists on album sales, but it is often the case that recoupables — recording costs and the like — cancel out the received royalties, especially for small- and medium-name artists. It is common knowledge (and somewhat documentable[8]) that artists receive the bulk of their revenue from concert tickets and merchandise sales. Filesharing undoubtedly allows users to discover new artists and music at a fraction of the cost of buying CDs. Word-of-mouth advertising brings in more concertgoers, and bands have a better chance of being successful, especially if they're talented. As a personal anecdote, in the past year the author has been to more than five concerts, at which he has purchased merchandise, for artists he would have otherwise never heard of or purchased from if he had not fileshared. Thus when making policy considerations, we must ask whether to concern ourselves with an industry using a seemingly-outdated distribution model, or to support a new distribution model that more efficiently selects artists and performers based on consumer preference. It is fundamentally a question of efficiency.

[7] Harris 2007. OiNK.cd was a predecessor to and model for TorrentSite.net.

[8] http://www.rapcoalition.org/label_exec_$$_breakdown.htm

Figure 1: Mean ratios and buffers over the sample period.

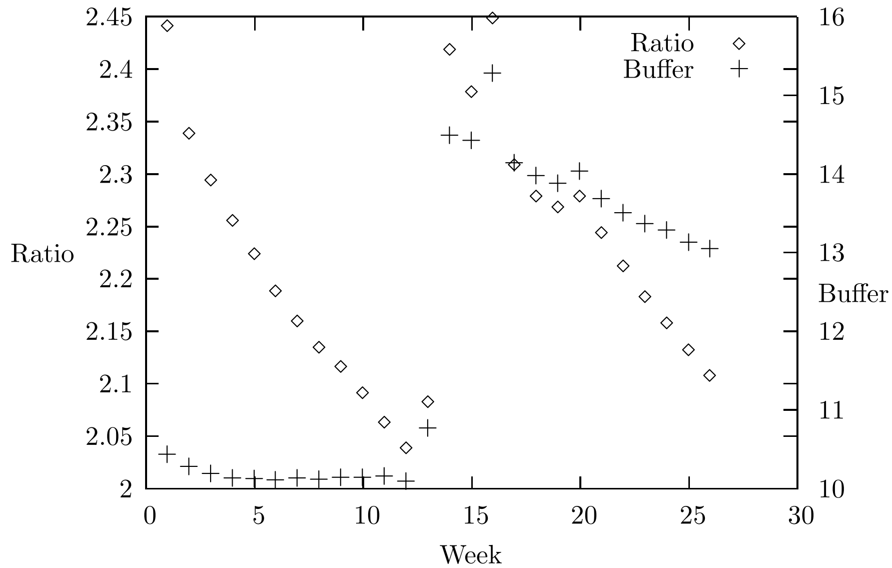


Table 1: Summary Statistics for Torrent Snatch Data

Measure	Total	MP3
Mean per Torrent	1.84	2.67
Median per Torrent	0	0
Weeks	27	
Obs.	4,068,254	2,987,432
Percent	100%	73%

Table 2: Summary Statistics for Album Sales Data

Measure	Total
Mean Weeks in Top 50†	7.75
Median Weeks in Top 50†	4
Mean Weekly Sales	40,392.9
Median Weekly Sales	21,524
Weeks	27
Obs.	1,350

Estimates are biased downward, since some albums at the end of the sample will appear on future charts.

Table 3: Regression of $Sales_t$ on $Sales_{t-1}$

y	Cons.	$y_{a,t-1}$
Est.	14,801	.438
σ	794	.009
R^2	.682	
Obs.	1072	

Figure 2: Median Sales per Week

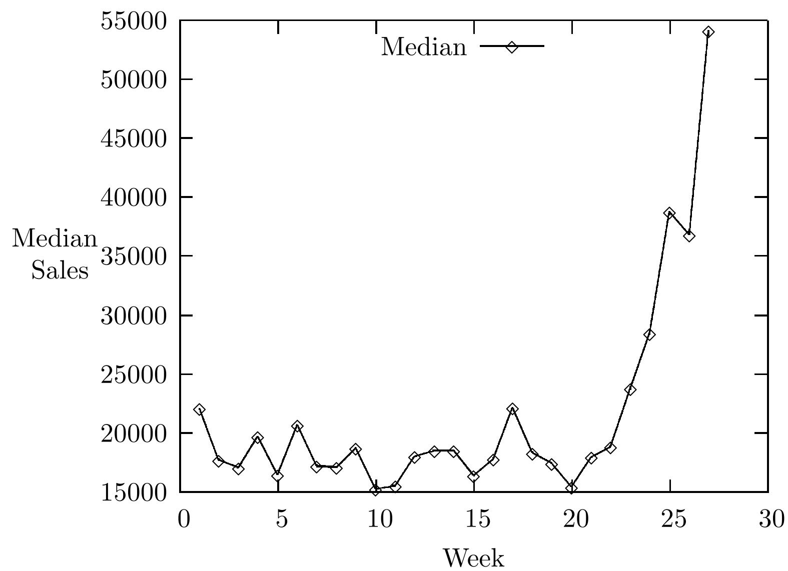


Table 4: Regressions of New Downloads on FLAC Dummy and MP3 Size

	(1)		(2)		
	Cons.	FLAC Dummy	Cons.	FLAC Dummy	Size (MB)
Est.	-.110	-.260	-1.388	.664	6.56×10^{-6}†
σ	(.021)	(.042)	(.017)	(.035)	(5.32×10^{-5})
F-stat	38.59		220.59		
Obs.	4,346,963		4,043,597		

All estimates are significant at the 1% level, except †, which has a t-stat of 0.12.

Table 5: Regression of New Downloads on Mean User Buffer, Ratio

	(1)		(2)	
	Cons.	Buffer	Cons.	Ratio
Est.	-911.50	66.02	1932.19	939.15
σ	(767.73)	(64.30)	(2334.14)	(1060.90)
t-stat	-1.19	1.03	0.83	-0.89
F-Stat	1.05		0.78	
Obs.	2,502,149		2,502,149	

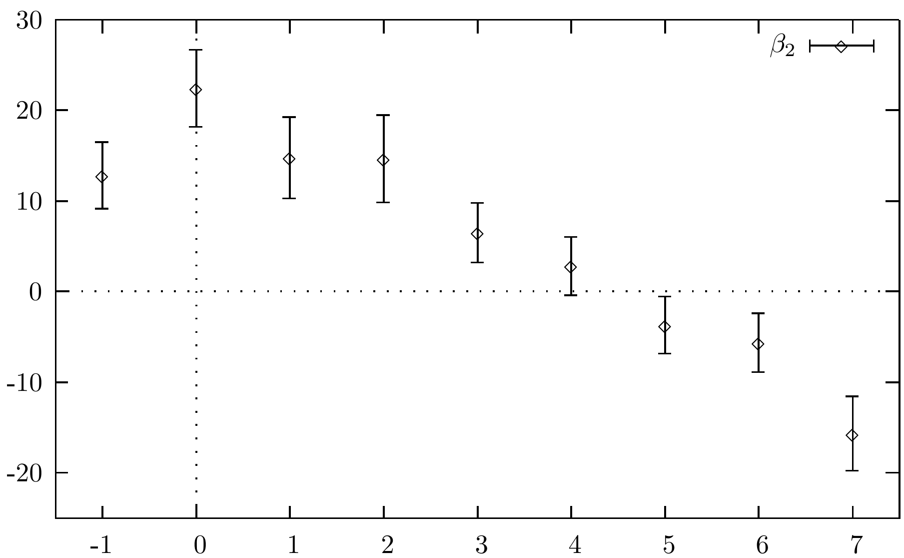
Table 6: Estimation of Equation (2)

w_t	Buffer		Ratio	
	Cons.	w_t	Cons.	w_t
Est.	647.61	-54.63	3379.59	-1514.55
σ	(137.86)	(10.60)	(395.63)	(174.22)
t-stat	4.70	-5.15	8.54	-8.69
R^2	.026		.070	
Obs.	1,011		1,011	

Table 7: Estimation of Equation (1)

Instrument	Buffer			Ratio		
	Cons.	$y_{a,t-1}$	$d_{a,t}$	Cons.	$y_{a,t-1}$	$d_{a,t}$
Est.	14,526	.419	5.557	15,458	.419	22.445
σ	(761)	(.008)	(5.724)	(872)	(.010)	(4.224)
z-stat	19.10	50.89	.97	17.73	41.92	5.31
R^2		.718			.585	
Obs.		1011			1011	

Figure 3: Estimates of β_2 in Equation (1) with x-Week Lags



Bars above and below each estimate show a one-standard deviation from its value.

References

[1] Liebowitz, Stan. *Economists Examine File-Sharing and Music Sales.* Excerpt from Industrial Organization and the Digital Economy, edited by Illing, Gerhard and Peitz, Martin. MIT Press. 7 July 2006.

[2] Waldfogel, Joel and Rob, Rafael. *Piracy on the High C's: Music Downloading, Sales Displacement, and Social Welfare in a Sample of College Students.* The Journal of Law & Economics, pp. 29-62, April 2006.

[3] Zentner, Alejandro. *Measuring the Effect of File Sharing on Music Purchases.* The Journal of Law & Economics, pp. 63-90, April 2006.

[4] Harris, Chris. *Music File-Sharing Site OiNK Shut Down Following Criminal Investigation.* MTV News, 23 Oct 2007. Availible at http://www.mtv.com/news/articles/1572554/20071023/id_0.jhtml.

[5] Hong, Seung-Hyun. *The Effect of Napster on Recorded Music Sales: Evidence from the Consumer Expenditure Survey.* SIEPR Policy paper No. 03-018. January 2004.

[6] Peitz, Martin and Waelbroeck, Patrick. *The Effect of Internet Piracy on Music Sales.* Review of Economic Research on Copyright Issues, 2004, vol. 1(2), pp. 71-79.

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

[8] Liebowitz, Stan. *How Reliable is the Oberholzer-Gee and Strumpf Paper on File-Sharing?.* September 2007. Available at SSRN: http://ssrn.com/abstract=1014399.

[9] Goldberg, Danny. *The Ballad of the Mid-Level Artist.* Reprinted from Inside.com. Available at http://www.rapcoalition.org/label_exec_$$_breakdown.htm.

[10] oorza. *On Rationomics: An Analysis into the Workings of BitTorrent Tracker's Economy and Its Composition.* TorrentSite.net User Forums, The Library. 13 May 2008. Available upon request.