MANAGEMENT SCIENCE
Vol. 53, No. 9, September 2007, pp. 1359–1374
ISSN 0025-1909 | EISSN 1526-5501 | 07 | 5309 | 1359

informs®

DOI 10.1287/mnsc.1070.0699
© 2007 INFORMS

The Effect of Digital Sharing Technologies on Music Markets: A Survival Analysis of Albums on Ranking Charts

Sudip Bhattacharjee, Ram D. Gopal
Department of Operations and Information Management, School of Business, University of Connecticut, Storrs, Connecticut 06269 {sudip@business.uconn.edu, rgopal@business.uconn.edu}

Kaveepan Lertwachara
Department of Management, California Polytechnic State University, San Luis Obispo, California 93407, klertwac@calpoly.edu

James R. Marsden
Department of Operations and Information Management, School of Business, University of Connecticut, Storrs, Connecticut 06269, jim.marsden@business.uconn.edu

Rahul Telang
H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, rtelang@andrew.cmu.edu

Recent technological and market forces have profoundly impacted the music industry. Emphasizing threats from peer-to-peer (P2P) technologies, the industry continues to seek sanctions against individuals who offer a significant number of songs for others to copy. Combining data on the performance of music albums on the Billboard charts with file sharing data from a popular network, we assess the impact of recent developments related to the music industry on survival of music albums on the charts and evaluate the specific impact of P2P sharing on an album's survival on the charts. In the post-P2P era, we find significantly reduced chart survival except for those albums that debut high on the charts. In addition, superstars and female artists continue to exhibit enhanced survival. Finally, we observe a narrowing of the advantage held by major labels. The second phase of our study isolates the impact of file sharing on album survival. We find that, although sharing does not hurt the survival of top-ranked albums, it does have a negative impact on low-ranked albums. These results point to increased risk from rapid information sharing for all but the "cream of the crop."

Key words: peer-to-peer; digitized music; online file sharing; survival
History: Accepted by Barrie R. Nault, information systems; received June 20, 2005. This paper was with the authors 7 months for 4 revisions. Published online in *Articles in Advance* July 20, 2007.

1. Introduction

The entertainment industry, in particular the music business, has been profoundly impacted by recent technological advances. Music-related technologies such as audio-compression technologies and applications (MP3 players in 1998), peer-to-peer (P2P) file-sharing networks like Napster (in 1999), and online music stores (in 2000) were introduced in a relatively short span of time and gained rapid popularity. Consumers of music adapted rapidly to the new environment. In fact, music titles, names of musicians, and music-related technologies (e.g., MP3) have consistently been among the top ten searched items in major Internet search engines since at least the year 2000 (Google, Inc.).

The music industry and its industry association, the Recording Industry Association of America (RIAA), have repeatedly claimed that emerging technologies, especially P2P networks, have negatively impacted their business. RIAA reports that music shipments, both in terms of units shipped and dollar value, have suddenly and sharply declined since 2000 (RIAA 2003). RIAA attributes these dramatic changes directly to the free sharing of music on online P2P systems. This assertion has garnered wide attention and has been the subject of numerous debates (Liebowitz 2004; King 2000a, b; Mathews and Peers 2000; Peers and Gomes 2000; Evangelista 2000). Alexander (2002) viewed P2P technologies as leading to free riders and undermining market efficiencies in the music industry with users obtaining music freely in lieu of legally purchasing the music.

Claiming that the impact of online music sharing on the music business has been devastating, RIAA has

aggressively pursued stronger copyright enforcement and regulations (Harmon 2003). RIAA's initial legal strategy was aimed at Napster—RIAA succeeded in shutting down the network largely due to potential liability around Napster's centralized file search technology. The so-called sons of Napster quickly emerged to fill the vacuum, attempting to escape legal wrath by deploying further decentralized structures. In response, RIAA has since altered its legal strategy by seeking sanctions against individuals "who offer a significant number of songs for others to copy" (Zeidler 2003, Bhattacharjee et al. 2006c). But there is an opposing view arguing that P2P systems significantly enhance the ability of users to sample and experience songs. Digital technologies have undoubtedly made information sharing and sampling easier[1] (Bakos et al. 1999, Barua et al. 2001, Brynjolfsson and Smith 2000, Bhattacharjee et al. 2006a) and less costly (Cunningham et al. 2004, Gopal et al. 2004) for individuals. Consumers' increased exposure to music, made possible by P2P systems, also has potential benefits to the music industry. An expert report in the Napster case alludes to the possibility that such online sharing technologies provide sampling mechanisms that may subsequently lead to sales (Fader 2000). The report also argues that the decline in the music industry is due to factors other than P2P-enabled music sharing. Concomitant with the introduction and popularity of P2P systems, the music industry has seen increasing competition for consumer time and resources from nonmusic activities such as video games, DVDs, and online chat rooms (Mathews and Peers 2000, Mathews 2000, Boston 2000) and a downturn in the macroeconomic conditions (e.g., drop in gross domestic product growth rates and employment figures since 2000 through the end of our study period in late 2003).

Empirical evaluation of the impacts of sharing on the success of music products has yielded conflicting results and sparked continued controversy (Liebowitz 2006). Self-reporting bias, sample selection, simultaneity problems, and lack of suitable data to draw the reliable conclusions may all have contributed to contradictory findings. Recent work (Oberholzer and Strumpf 2007) relates downloading activity on two P2P servers with sales of music albums. The authors' data set spans the final 17 weeks of 2002 and was obtained from OpenNap, a relatively small P2P network with a centralized structure as in Napster. Oberholzer and Strumpf (2007, p. 1) found that the effect of downloads on sales is "statistically indistinguishable from zero." However, other studies argue that P2P sharing hurts the music industry (Liebowitz 2006).

[1] Online fan clubs exist for numerous popular performers.

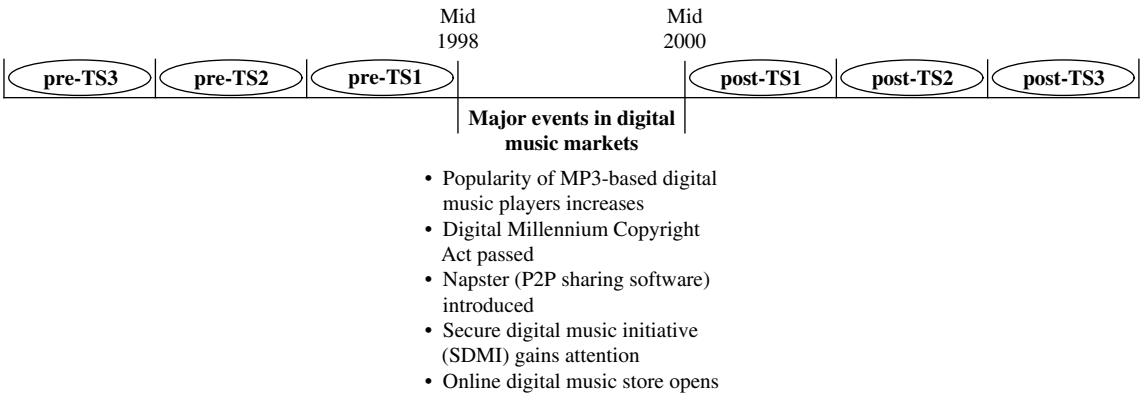
The objectives of our study are twofold: (1) assess the impact of recent market and technological developments related to the music industry on survival of music albums on the top 100 charts, and (2) evaluate the specific impact of P2P sharing on album chart survival. We use data on music albums on the top 100 weekly charts together with daily file-sharing activity for these albums on WinMx, one of the most popular file-sharing P2P networks (Pastore 2001; Graham 2005a, b).

Since 1913, Billboard magazine has provided chart information based on sales of music recordings (Gopal et al. 2004). The chart information for the weekly Top 100 albums is based on "a national sample of retail-store sales reports collected, compiled, and provided by Neilsen Soundscan" (Billboard). Appearance and continued presence on the chart has important economic implications and influence on awareness, perceptions, and profits of an album (Bradlow and Fader 2001). Having an album appear on the charts is an important goal of most popular music artists and their record labels (Strobl and Tucker 2000). Our focus is on the survival of albums as measured by the number of weeks an album appears on the top 100 chart before final drop-off. This survival period on the chart captures the "popular life" of an album and has been the object of analysis in a number of studies related to music (Strobl and Tucker 2000, Bradlow and Fader 2001).

Figure 1 illustrates the time frame of analysis for the initial phase of our study. The two-year span, mid-1998 to mid-2000, represents a watershed period in the music industry during which a number of significant events unfolded, including (i) introduction and rapid popularity of MP3 music format, (ii) passage of the Digital Millennium Copyright Act, (iii) introduction and rapid rise in the usage of Napster and P2P networks, (iv) surge in the popularity of DVDs, online chat rooms, and games; and (v) start of a downturn in the overall economy.

The first reported decline in music shipments occurred in 2001, suggesting the possibility that the influence of these events was beginning to be experienced by the music industry. The first phase of our study provides a comparative analysis of album survival before and after the mid-1988 to mid-2000 event window. As depicted in Figure 1, chart information was compiled for three time segments (TSs) before and three after the event window, depicted as pre-TS1 to pre-TS3 and post-TS1 to post-TS3, respectively. In total, over 200 weeks of chart information, spanning the years 1995–2004, was collected for this phase of the study. The following explanatory variables of album survival are analyzed to assess possible changes in impact between the pre- and post-TSs: debut rank of the album, reputation of the artist (as captured by superstar status), the record label

Figure 1 Survival Analysis Time Frame



Note. Each time segment (TS) signifies sample of 34 weeks. An oval signifies a sample between indicated times.

that promotes and distributes the album, and artist descriptors (i.e., solo female/solo male/group).

The second phase of the study attempts to identify the impacts of file sharing on chart success. Our analysis utilizes: (1) data on sharing activity on WinMx for 300+ albums over a period of 60 weeks during 2002 and 2003, (2) corresponding Billboard chart information, and (3) relevant values for other variables detailed above. Our analysis and findings relate only to those albums that appear on the charts. Over 30,000 albums are released each year, but only a small proportion of these appear on the charts. However, this small set of successful albums provides the lion's share of the profits for the record companies (Seabrook 2003).

Our analysis uses sharing that occurs after an album has made an appearance on the charts. We ask the research question: Does the level of sharing influence survival time on the charts? We investigate the impact of sharing in the debut week and also the maximum level of sharing in each of the four-week periods (see details in §4). Much of the initial sales of an album are to the so called "committed fan base" (Strobl and Tucker 2000). This core set of consumers are early adopters who have often completed their purchase by the time the album has appeared on the chart. Consequently, the number of weeks an album remains on the chart tends to reflect its receptiveness by the nonhard-core consumers. An impediment in investigating the impact of sharing on album survival is the issue of endogeneity (or omitted variable bias), in that albums that are shared more may also survive longer. Finding an appropriate and strong instrument to address endogeneity is a key requirement in empirical work in this domain, and our paper makes a significant methodological contribution in that regard.

Our expanded analysis offers significant new insights tied to our inclusion of P2P sharing, major/minor label release, and gender of the artist. We find that, overall, sharing has no statistically significant effect on survival. However, a closer analysis reveals that the effect of sharing appears to differ across certain categories. Successful albums (albums that debut high on the chart), are not significantly impacted by sharing. However, online sharing has a low but statistically significant negative effect on survival for less successful (lower debut rank) albums.

Four recording labels (Sony-BMG, Universal, EMI, and Warner Brothers) dominate the music industry and are often referred to as the "major labels." We find that since the occurrence of the significant events outlined above (in the mid-1998 to mid-2000 time frame), the effect of debut rank on chart success has risen whereas the effect of being released by a major label has fallen. In addition, solo female artists perform better than either solo male artists or groups across the periods.

Section 2 discusses related literature that aids in the development of our empirical methodology. Alternative model forms are presented in §3. We detail the proportional hazard (PH), accelerated failure time (AFT), and ordinary least squares (OLS) model approaches, illustrating their interrelationships. The details of the data collection are presented in §4. Section 5 centers on model estimation. We demonstrate that, for the first phase of our analysis, the estimates of the alternative model forms (PH, AFT, and OLS) are virtually identical. As we address potential omitted variable bias and spurious implication issues using an instrumental variable approach, the second phase of our analysis uses the OLS approach. Section 6 is devoted to a discussion of key findings, their implications, and suggested future research directions.

2. Related Literature

Although research on the post-P2P music world is just emerging, there exists a rich body of earlier work in economics, marketing, and information systems related to the markets for music and the music

industry. Music is an experience good whose true value is revealed only after its consumption (Nelson 1970), a product whose evaluation is based primarily on personal experience and individual consumer tastes, rather than specific, objectively measurable product attributes (Dhar and Wertenbroch 2000, Moe and Fader 2001). Music is often alluded to as a fashion-oriented product, where customer tastes and preferences can change rapidly and can be influenced by other consumers who have purchased it. Thus, sampling and experiencing music prior to purchase, along with cues on how well a music item is perceived by other individuals, can be important components in consumer purchase decisions. But sampling music items can require significant time and effort, given the large body of available recorded music (Bhattacharjee et al. 2006b). The four major music labels alone release about 30,000 albums annually (RIAA, Goodley 2003). Only a tiny fraction of the albums released are profitable and achieve the success indicated by appearing in the top 100 charts (Seabrook 2003). In fact, of the albums released in 2002, the vast majority (over 25,000) sold less than 1,000 copies each (Seabrook 2003). The fact that music is fashion-oriented adds a degree of complexity for music labels seeking to assess the likely success of a product (Bradlow and Fader 2001). Additionally, the introduction rate of new music albums and overall album sales vary across the year. Industry figures show that a large number of albums are released during the Christmas holiday period, suggesting that the success of music albums might also be impacted by their time of release (Montgomery and Moe 2000).

Prior research has examined various factors that can influence the success of music albums, including the phenomenon of superstardom in the music industry and its correlation with album success (Rosen 1981, Hamlen 1991, MacDonald 1988, Towse 1992, Chung and Cox 1994, Ravid 1999, Crain and Tollison 2002). Adler (1985) suggested that the superstar effect results from consumer desire to minimize search and sampling costs by choosing the most popular artist. The search for information is costly. Consumers must weigh their additional search costs for unknown artists or items of music with their existing knowledge of a "popular" artist. MacDonald (1988) suggests that, in a statistical sense, consumers correlate past performance with future outcomes and try to minimize the variability in their expectations of individual performances.

Four major labels account for about 70% of the world music market and 85% of U.S. market (International Federation of Phonographic Industry 2005, Bemuso 2006, Knab 2001, Spellman 2006). The majors exert significant control in recording, distributing, and promoting of music albums and possess the financial resources to gain access to large customer bases. There are thousands of minor labels which together account for less than 30% of world-market share. These labels, hampered by the lack of resources to reach wider audiences, tend to operate in niche segments (Spellman 2006). The albums released by the major labels are promoted more, have wider audience exposure, and, consequently, tend to last longer on the charts (Strobl and Tucker 2000).

Previous research suggests that one of the most important characteristics in guaranteeing survival on the charts is the initial debut rank (Strobl and Tucker 2000). This relationship may be due to the *bandwagon effect* in the demand for music (Towse 1992, Strobl and Tucker 2000). This effect arises from the process of *acquiring tastes* in which preferences for a good increase because others have purchased it (Leibenstein 1970, Bell 2002). The initial debut rank reflects an album's acceptance by early adopters, which can create further demand from remaining consumers (Yamada and Kato 2002).

All these factors—superstar effect, major label promotion effect, and debut-rank influence—reflect consumers' unwillingness to incur additional search and sampling costs to identify unknown music of potentially high value (Adler 1985, Rosen 1981, Leibenstein 1970). P2P technologies have significantly lowered consumer costs to sample and experience music, to acquire and enhance their knowledge on artists, and to interact with other individuals. Walls (2005a, p. 178) suggests that the demand processes for popular general entertainment products are "characterized by recursive feedback" (see also Krider and Weinberg 1998). Word-of-mouth, now spread electronically, can significantly impact the consumption decisions of potential customers. Further, Chevalier and Mazylin (2006) find that in the case of books, one-star reviews have a larger impact on book sales than five-star reviews. That is, less well-received books are hurt more significantly by information sharing. Gopal et al. (2006) suggest that sharing technologies enable consumers to be more discerning on their purchases from music products by superstars. The authors predict that sharing technologies will lead to a dilution in the superstar effect and the emergence of more new artists on the charts, because sharing will enable purchase behavior to be driven more by the value attached to the album and less by the reputation of the artist. The focus of Gopal et al. (2006) is on the impact of superstardom *prechart* appearance of an album, whereas our focus is on continued success "post-chart" appearance of an album.

Several recent papers have suggested the use of specialized skewed distributions to model the success of entertainment products, most notably motion pictures (see Krider and Weinberg 1998; De Vany and Walls 1999, 2004; Walls and Rusco 2004; Walls 2005a, b). There is a key difference between related

work on motion-picture returns and our work. The former have typically analyzed all products (movie releases) released over a specified time period, using data sets that include numerous poor performing (in terms of revenue) movies. What we model is quite different. We have a prefilter in that we consider only albums that succeed in appearing on the Billboard top 100 chart. In a given year, only a few hundred albums make it to the Billboard charts. Given that over 30,000 albums debut each year, our analysis does not include the heavy failure rate inherent in much of the prior work.

3. Model of Album Survival

The *survival* we model is the length of time or duration that an album remains on the charts before dropping off. This survival process is a stochastic process (where the time index is one week) that governs whether an album exits the charts (see Kiefer 1988 for a detailed discussion of duration models). Survival models differ from hazard models where the focus is on understanding the relationship between the event ("death" or exiting the Billboard chart) occurring at a time t and values of a variety of explanatory variables. One popular form of hazard model is the following PH model for a point in time, t:

$$h(t) = h_0(t)\exp(X_1\beta_1^{\mathrm{PH}} + X_2\beta_2^{\mathrm{PH}} + \cdots + X_p\beta_p^{\mathrm{PH}}), \quad (1)$$

where the X_i's are a set of explanatory variables, which shift the hazard function proportionally, β_i^{PH}'s are the parameters to be estimated, and $h_0(t)$ is called the "baseline hazard…the value when all X_i are equal to zero" (see Bradburn et al. 2003, p. 432). In the Cox specification of (1), no assumption is incorporated about the distribution of $h(t)$. In a fully parametric regression model of (1), $h(t)$ is assumed to follow a specific distribution, often the Weibull. As our interest is in modeling chart survival time, we consider an AFT model. Following Bradburn et al. (2003), we write the model as

$$S = S_0(T\phi) = S_0(T(\exp(\gamma_1 + \gamma_2 X_2 + \cdots + \gamma_p X_p))), \quad (2)$$

where S is the duration of survival and

$$\phi = \exp(\gamma_1 X_1 + \gamma_2 X_2 + \cdots + \gamma_p X_p)$$

is termed the *acceleration factor*. When all the X_i's equal zero, the model collapses to $S_0(T)$, which is referred to as the baseline survivor function. For estimation, the AFT model in (2) is commonly put into log linear form with an additive residual term (ε), that is, $\ln(S) = \gamma_0 + X_i\gamma_i + \varepsilon$, where γ_0 is the baseline survivor value or intercept term. This is similar to linear regression models (OLS) except that the error terms follow different distributions. Bradburn et al. (2003, p. 434) report the following important result:

> When the survival times follow a Weibull distribution, it can be shown that the AFT and PH models are the same. However, the AFT family of models differs crucially from the PH model types in terms of interpretation of effect sizes as time ratios opposed to hazard ratios.

The Cox formulation of the PH model cannot be transformed to an AFT specification as the hazard is nonparametric (Kalbfleisch and Prentice 2002, p. 44).

Thus the issue really focuses on whether there are significant differences in the error term structure. In other words, is our analysis satisfactorily characterized by normally distributed error terms or other nonnormal and possibly skewed error distributions? If the former holds, then OLS becomes an attractive candidate for our work because, as explained in §3.3 below, we need an instrumental variable to evaluate the specific impact of P2P sharing on an album's survival on the chart. When using instrumental variables, the two-stage least squares (2SLS) method is particularly robust and widely used. AFT or hazard models are not particularly suitable for such analysis (Belzil 1995).

Additional concerns also arise with the use of OLS for survival analysis. First, left or right data censoring issues (i.e., inability to identify birth or death times of some data entities) often occur in survival analysis. However censoring does not occur in our data as we track each album from its debut (birth) till its final drop-off (death) from the charts. With no censoring issues, OLS regression sing logarithmic transformation of the dependent variable yields results that closely approximate those from hazard models. Second, the use of OLS is suspect if there are time-varying covariates. In our case, there are no time-varying covariates, because, for a given album, our covariates (e.g., debut rank, gender) do not change over the duration of survival.

Up until the introduction of the instrumental variable, we provide estimation results for all three specifications and show they are (1) virtually the same for both the nonparametric Cox and the Weibull PH models, and (2) virtually the same for the Weibull AFT (which is equivalent to Weibull PH) and OLS. In summary, we use the OLS specification because (i) left or right data censoring issues do not arise in our data, as we track each album from its debut (birth) till its final drop-off (death) from the charts; (ii) for any given album, there are no time-varying covariates; and most importantly (iii) we employ an instrumental variable approach to estimate the impact of sharing on album survival. Formal tests for normality of residuals lend

additional support for the use of OLS (see the online appendix provided in the e-companion).[2]

3.1. Album Survival

In the first phase, we focus on possible shifts in album survival following the major events related to the music industry. The initial model to be estimated is presented as an OLS formulation:

$$\ln(survival_i) = X_i\beta^{\text{OLS}} + debut\ post\text{-TS}_i\delta^{\text{OLS}} + \mu_i \quad (3)$$

where $survival_i$ denotes the total number of weeks an album i appears on the Billboard top 100 charts. X_i is a vector of album specific control variables: debut rank, superstar status, distributing label (major/minor), debut month, and gender (artist type). *Debut post-TS*$_i$ is an indicator that signifies an album's debut period (see Figure 1). This variable is set to 1 if the album debuted in the post period (2000–2002) and 0 otherwise. The estimate of δ is of significant interest, as it indicates how survival has changed from the pre-TS period to the post-TS period. However, the change in survival may not be linear and may be moderated by album characteristics. For example, top-ranked albums (numerically lower ranks) may be more affected across pre- and post-TS periods. Similarly, minor (or major) record labels may have benefited more (or less) after the popularity of file-sharing networks. To be able to consider such possibilities, we interact album-specific characteristics with *debut post*-TS$_i$ and estimate the following model:

$$\ln(survival_i) = X_i\beta^{\text{OLSI}} + debut\ post\text{-TS}_i\delta^{\text{OLSI}} + (X_i \times debut\ post\text{-TS}_i)\xi^{\text{OLSI}} + \mu_i^{\text{OLSI}}, \quad (4)$$

where ξ^{OLSI} is the vector of parameters to be estimated, along with β^{OLSI} and δ^{OLSI}.

We estimate Equation (3) with both Weibull and Cox PH specifications and show that the estimates are quite similar (see the e-companion). Weibull and Cox PH are estimated controlling for unobserved heterogeneity. In particular, in continuous time PH models, not controlling for heterogeneity may produce incorrect estimates. To incorporate unobserved heterogeneity, we modify (1) such that

$$h(t) = h_0(t_i)\exp(X_i\beta^{\text{PHUH}} + v^{\text{PHUH}}), \quad (5)$$

where v^{PHUH} has a gamma distribution with mean 0 and variance ς^2, which can be estimated. We then estimate (3) again with Weibull AFT and OLS specifications and show in §5 that they are virtually identical. A similar approach is used to estimate Equation (4).

[2] An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

3.2. Impact of Sharing on Survival

In the second phase of the analysis, we examine the impact of file sharing on an album's survival. As discussed later in §4, we observe the number of files being shared for each album in time segment post-TS3. We use this information to understand how the intensity of file sharing can affect an album's survival. The OLS formulation is

$$\ln(survival_i) = X_i\beta^{\text{OLSS}} + \ln(shares_i)\lambda^{\text{OLSS}} + \mu_i^{\text{OLSS}}, \quad (6)$$

where, as before, X_i is a vector of album-specific control variables, and $shares_i$ denotes the number of files being shared for a given album. As we observe high variance and skewness in the sharing levels across albums in our data set, we use a logarithmic transformation for shares. The estimate of λ^{OLSS} is of key interest as it indicates the impact of sharing levels on an album's continued survival. As in §3.1, Equation (6) is estimated with PH, AFT, and OLS specifications. As before, AFT and OLS estimates are included in §5, and PH estimates are presented in the e-companion.

3.3. Omitted Variable Bias: Analysis Using Instrument

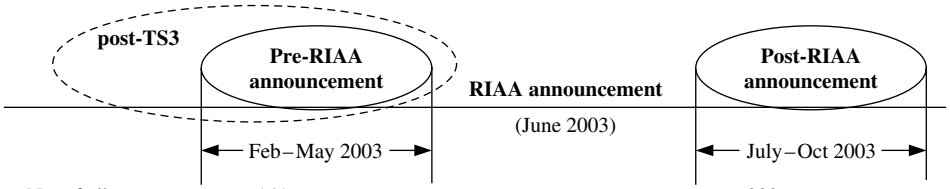
A direct estimation such as in Equation (6) may not be appropriate as sharing may be closely correlated with unobservable (or not directly measurable) album characteristics (perhaps "popularity" of a particular artist). Record labels often promote certain albums through radio airplay to enhance popularity and signal potential hit songs. Such actions to enhance popularity of selected albums may influence both album survival on the charts and sharing on P2P networks. Thus "popularity" may be an important omitted variable driving both sharing and survival. It is also possible that an album's position on the chart could affect its sharing. Although debut rank should control for some of this, such a correlation would bias the estimate for λ^{OLSS}, as $shares_i$ would be correlated with the error term μ_i^{OLSS}, thus violating the assumptions of the general linear model. One strategy is to find an instrument which is correlated with sharing but not with survival. We would then estimate

$$\ln(shares_i) = Z_i\alpha^{\text{INS}} + X_i\beta^{\text{INS}} + v_i^{\text{INS}}, \quad (7)$$

where Z_i is a vector of instruments uncorrelated with μ_i^{OLSS}. A general strategy is to substitute the predicted values of sharing into the first stage (Equation (6) above) and reestimate the first stage, which yields unbiased estimators.

On June 25, 2003 RIAA announced that it would start legal actions against individuals sharing files on P2P networks—an announcement extensively disseminated through various print and broadcast media the following day. Unless RIAA was mistaken, this event

Figure 2 Time Frame for Sharing Analysis with Instrument



No. of albums:	141	229
Mean sharing level:	345.1	61.9
Max sharing level:	3,671	973
Std. dev. of sharing level:	575	138.7
Mean survival time:	7.17 weeks	8.34 weeks

Note. An oval signifies a sample between indicated times.

should have had a direct impact on users sharing files on the network. But, because the event would likely be uncorrelated with the error term, this event can be used as an instrument shifting the intensity of sharing. Thus Z_i is 1 for data after June 2003 and 0 otherwise. We analyze our data using 2SLS and report the estimates.

The need to use an instrumental variable to deal with the omitted variable issue prompts us to consider use of OLS models in the first phase of our study. Although 2SLS has been heavily analyzed and considered quite robust, we were not able to find an equivalent method in the context of hazard models. Although a hazard model is a more natural choice to estimate survival on the charts, the ability to use the well-established methodology of 2SLS to consider the omitted variable issue leads us to select OLS as appropriate for our work. In addition, the OLS estimates turn out to be nearly identical to the hazard models estimates.

We collected sharing data from October 2002 to June 2003, and from July 2003 to December 2003. The sharing statistics before (October 2002–June 2003) and after (July–December 2003) suggest that the intensity of sharing fell considerably after the event, from a mean of 345.1 to 61.9, whereas survival increased slightly from 7.17 weeks to 8.34 weeks. To avoid a temporal effect or other exogenous variables that might have an impact on survival, we chose a relatively short window of four months before and after the RIAA announcement. We include only those albums that debut between February–May 2003 and July–October 2003 (Figure 2). We also tried to control for factors like overall economic indicators by incorporating the S&P 500 market index.[3] Using the sample described above and the June 2003 event as the instrument, we estimate Equations (6) and (7) using 2SLS.

Finally, similar to album survival analysis before and after major market and other events (§3.1), we consider possibly significant interactions between *shares* and other variables in X_i. Thus we estimate the vector of parameters θ^{OLSSI} along with β^{OLSSI} and λ^{OLSSI} in the following:

$$\ln(survival_i) = X_i\beta^{\text{OLSSI}} + \ln(shares_i)\lambda^{\text{OLSSI}} + (X_i \times \ln(shares_i))\theta^{\text{OLSSI}} + \mu_i^{\text{OLSSI}}. \quad (8)$$

As before, we use $Z_i \times X_i$ as a potential instrument for the interaction term $X_i \times \ln(shares_i)$.

4. Data

4.1. Data Set 1

The first data required are the weekly rankings of albums on the Billboard top 100 charts. In year 2003 and in earlier years, album sales accounted for a dominant majority of the total sales (RIAA 2003) with RIAA reporting that in 2003 digital downloads (online sales) were just 1.3% of revenue and "singles" sales just 2.4%.

For each TS (see Figure 1), the data relate to albums that debut during 34 consecutive weeks of observation. Exact start dates for each year, shown in Table 1, indicate that our data collection covers the traditional holiday sales period, when new releases and sales volume are the highest, as well as the more tranquil first and second quarters.

Table 1 Billboard Top 100 Data Collection

Time segment	Start date
Pre-TS3	27 October 1995
Pre-TS2	25 October 1996
Pre-TS1	24 October 1997
Post-TS1	27 October 2000
Post-TS2	26 October 2001
Post-TS3	25 October 2002

[3] For example, it may be that economic outlook is substantially different over these periods, thus affecting buyers' purchasing behaviors systematically. We tested with various dummy variables indicating the month of album debut. All lead to insignificant results.

Table 2 Mean Statistics for Key Variables

	(Mid-1998 to mid-2000)					
Variables	pre-TS3 ($n = 218$)	pre-TS2 ($n = 224$)	pre-TS1 ($n = 234$)	post-TS1 ($n = 248$)	post-TS2 ($n = 261$)	post-TS3 ($n = 307$)
Survival	14.2 weeks	14.6 weeks	15.3 weeks	11.3 weeks	9.5 weeks	9.6 weeks
Debut rank	49.9	49.15	49	42.9	39.5	34.5
Albums released	30,200	30,200	33,700	35,516	31,734	33,443
Superstar (%)	31.6	28.50	27.8	26.6	23.3	15.6
Minor label (%)	13.7	16	13.2	22.9	25.6	24.7
Solo male (%)	29.8	33	31.6	29.7	34.8	34.5
Solo female (%)	11.5	9.40	12.3	12.5	15.3	14
Group (%)	58.7	57.60	55.9	57.6	49.8	51.5

We operationalize the survival model explanatory variables (X_i's) as follows:

Survival. number of weeks an album appears on the Billboard top 100 charts. On occasion, an album may drop off for some weeks and reappear again on the chart. Each album is continuously tracked till its final drop-off. As detailed earlier, our data does not suffer from left or right data censoring issues, as we track each album from its chart debut (birth) until its final drop-off (death) from the charts, which may occur well beyond the 34 weeks of each time segment;

Debut rank. the rank at which an album debuts on the Billboard top 100 chart. Numerically higher-ranked albums are less popular;

Debut post-TS. this is an indicator variable, which is 0 for albums that debut in pre-TS and 1 for post-TS;

Albums released. number of albums released during each year of the study period. This is used as a control variable as more albums released in a given year may signify increased competition amongst albums and reduce survival;

Superstar. a binary variable denoting the reputation of the artist. If a given album's artist has previously appeared on the Billboard top 100 charts for at least 100 weeks (on or after January 1, 1991) prior to the current album's debut, then the variable is set to 1, otherwise 0;

Minor label. a binary variable that is set to 0 if the distributing label for a given album is one of (Universal Music, EMI, Warner, SONY-BMG). A value of 1 denotes independent and smaller music labels;

Solo male. a binary variable that denotes if an album's artist is a solo male (e.g., Eric Clapton);

Group. a binary variable that denotes if an album's artist is a group (e.g., U2, The Bangles);

Solo female. a base control variable that denotes if an album's artist is a solo female (e.g., Britney Spears); artist is *solo female* if *solo male* = 0 and *group* = 0.

Holiday_month debut. To control for the holiday effect (or "Christmas effect"), we include an indicator variable for December, which is 1 if album debuted in that month and 0 otherwise.

Table 2 presents descriptive statistics for our first data set. The average survival decreased between the two periods, from about 14 to 10 weeks. Conversely, average debut rank improved from 49 to less than 40 on average. Together, these results indicate that, on average, albums tend to debut at better positions but drop more steeply in the post-TS period.[4] The number of albums released was roughly the same, with slightly higher numbers in two of the three post-TS years. The number of superstars appearing on the chart decreased marginally for the post-TS period. The percentage of male and female solo artists registered a small increase at the expense of groups. Finally, the number of albums from minor labels appearing on the charts increased substantially for the post-TS period.

4.2. Data Set 2

Our second data set relates to album-level sharing activity captured from WinMX for the 34-week period corresponding to the time segment post-TS3. We collected additional data from July–December 2003 for our analysis using the instrumental variable to assess the impact of sharing on album survival. In each of three reported years (2001, 2002, and 2005), the top file-sharing application had slightly over two million unique users (see Pastore 2001; Graham 2005a, b) with the second most popular having 1.3 to 1.5 million users. In 2001, Morpheus held the top spot but was overtaken by KaZaA in 2002. During our data collection period, WinMx was second behind KaZaA with a user base of over 1.5 million (Pastore 2001; Graham 2005a, b). By 2005, WinMX had overtaken KaZaA and was reported to have 2.1 million users. We used WinMx and not KaZaA because the latter places a

[4] This may indicate that album sales are concentrated upfront in this period, but lack of publicly available sales data precludes us from investigating this phenomenon. There is also a physical limit to the size of upfront sales in consecutive weeks, which is primarily constrained by logistics, distribution, and retailer shelf space. Retail distribution is the major sales channel, accounting for more than 98% of sales.

fixed limit on the number of files returned in any given search result. Using KaZaA could thus result in significant understatement of the level of sharing activity due to this hard upper limit imposed by the KaZaA search option.

The WinMX data was collected daily. Each day, we began with the list of albums that appeared on the Billboard top 100 chart since October 25, 2002 until the current week. The list of albums was randomly sorted to determine the order in which the search was conducted each day. The daily results were averaged to produce weekly information on sharing for each album. Although we have data on the sharing activity for every week after an album makes its first appearance on the chart, our analysis focuses on sharing levels during the debut week as sharing activity levels in the first few weeks were highly correlated (e.g., a correlation coefficient of 0.93 between sharing levels in the debut week and week after). We did use two alternative measures of sharing level, one relating to average sharing level observed in the debut week and one relating to the maximum sharing level observed in the first four weeks:

Shares_debut. average number of copies of an album available on the network during the debut week;[5] and,

Shares_max. maximum available copies of a file over a four-week period or until the album drops off the charts (whichever is less).

As reported in §5, we find that both measures yield consistent results. The mean number of copies available for sharing in our sample was approximately 802, with a minimum of 1 and a maximum of 6,620. Our analysis is at the aggregate level. That is, observed aggregate P2P sharing is an explanatory variable for album survival, where album survival is based on total aggregate sales. Further, we are not measuring the impact of downloading on an album's survival. Rather, we use "shares" as an indication of an album's availability on the network. We use availability because this corresponds to the modus operandi of RIAA, which has targeted legal action against file sharing rather than file downloading.

The use of "availability" of a file also does not suffer from potential bias associated with "download" data. First, availability of a file on a user's computer indicates that the user has archived the file and is offering it for sharing. On the other hand, using downloading activity would include files sampled but discarded. Second, search results for the number of available copies of a file returns information from a large number of nodes on the network. On the other hand, collecting downloading information requires monitoring "super nodes" through which control information is routed.[6] Finally, we suggest that higher availability (more copies) of a music item available on a network increases the ease and opportunity of finding and downloading.

[5] Various other formulations of *shares* were considered, including the proportion of tracks from an album that are available and the number of unique users sharing a particular album. All formulations produced similar and consistent results.

Table 3 Album Survival Estimation Results: OLS and AFT Models (Without Interaction Terms)

Parameter	(1) OLS	(2) Weibull AFT
Constant	0.45 (0.1)	8.86** (2.0)
Debut rank	−0.02** (24.0)	−0.02** (35.0)
Debut post-TS	−0.54** (8.3)	−0.28** (5.6)
Albums released	0.27 (0.47)	−0.60 (1.4)
Superstar	0.30** (4.8)	0.44** (8.7)
Minor label	−0.26** (3.8)	−0.16** (3.04)
Solo male	−0.36** (4.2)	−0.31** (4.6)
Group	−0.42** (5.1)	−0.43** (6.7)
Holiday_month debut	0.21** (2.9)	0.18** (2.8)
Frailty variance^		3.52** (14.6)
σ (Weibull shape parameter#)		3.62** (21.3)
Adjusted R^2	0.348	LL+ = −2,014

^*Frailty variance* is the estimated variance of the gamma distribution. Recall that we assume a gamma distribution for unobserved heterogeneity. The mean of the gamma distribution is not identified (it is fixed at 1) but the variance (sigma) is identified. A large variance suggests the existence of heterogeneity.

#*Weibull* is a two-parameter distribution with a shape and scale parameter. The shape parameter determines whether the hazard is increasing or decreasing. The scale parameter is simply subsumed in constant term of the regression and not identified.

+*Hazard models* (or accelerated failure models) are estimated using log likelihood (LL) functions and LL indicates the fit of the model, with lower absolute values indicating a better fit.

$^*p < 0.05$, $^{**}p < 0.01$; t-statistics in parentheses; $n = 1{,}484$.

5. Results

We now present the estimation results for each of the two phases of our analysis.

5.1. Phase 1—Analysis of Album Survival

Table 3 presents the Weibull AFT and OLS estimation results for the main effects models (Equation (3)) of the first part of our analysis. The corresponding PH estimates are detailed in the e-companion. Comparing Columns (1) and (2) in Table 3, we find the estimates are quite similar. The only minor difference is in the sign of the statistically insignificant coefficient on *albums released*. Though the values may be

[6] Several nodes are connected to a super node, which monitors the activity of the connected nodes. Hence it is possible that the downloading information may be biased by the types of users connected to the monitored super node. Availability information, as collected and used in this paper as "shares," usually is gathered by contacting several super nodes for the information if it is not available with the nearest super node, which reduces bias.

slightly different, all other estimates are consistent in sign and statistical significance. As the results in Table 3 suggest that OLS estimates are virtually identical with Weibull AFT estimates, the following discussion only focuses on OLS estimates. Low values of correlations between the variables suggest that collinearity is not a concern in our estimation (see details in the e-companion).

In the model without interactions (Table 3), coefficients on all variables, except *albums released*, were significant (0.01 level). *Superstar* and *holiday_month debut* enhance album survival, but the other variables display a deleterious impact.[7] Survival in the post-TS period is estimated to have declined by approximately 42%.[8] This significant shift in the survival pattern is consistent with our summary data in Table 2, where the mean survival time shows a sharp decrease. Albums that debut at higher numerical chart rank (hence less popular) tend to survive for a shorter period. In particular, a unit change in rank is estimated to reduce survival time by approximately 1.98%. An album debuting at rank 25 (out of 100) is estimated to fall off the charts 38.1% sooner than one debuting at rank 1 on the charts, whereas one debuting at 50 has a corresponding estimated drop in survival time of 62.5%. These estimates suggest the continued existence of the bandwagon effect in the music business (Towse 1992, Strobl and Tucker 2000).

The estimation results also highlight the importance of an artist's superstar status for chart success, with an album by a superstar estimated to survive 35% longer on the charts, ceteris paribus. Further, albums promoted by minor labels tend to have a survival duration 23% less than those promoted by major labels. Neither albums by solo male artists nor albums by groups survive as long as female artists. Albums that are released in December are estimated to survive 23% more weeks than albums released at other times, reflecting the holiday effect (Montgomery and Moe 2000). Overall, the regression model is highly significant with F-value significant at 1% and an adjusted R^2 of approximately 35%.

Table 4 presents the comparative results of OLS and Weibull AFT with interaction effects. Similar to Table 3, we find that all parameters of interest in Columns (1) and (2) of Table 4 are consistent in sign and level of significance, suggesting the OLS estimates are consistent with the Weibull AFT model. Focusing on OLS estimates, we find two statistically significant interaction effects, *debut post*-TS with *debut rank* and *minor label*, but the main effect coefficient of *debut post*-TS was statistically insignificant (with a negative sign). However, the main effect needs to be interpreted differently when an interaction term is present. In this situation, the main effect measures the impact of *debut post*-TS for the album debuting at top rank (or more precisely rank 0). This is in contrast to the results without interaction terms (Equation (3)), where the impact of *debut post*-TS is measured at the mean value of *debut rank*.

The interaction *debut rank* × *debut post*-TS suggests that the survival of top-ranked albums has not suffered in the post-TS period. However, in the post-TS period, the survival climate is increasingly hazardous for lower debut ranked (higher numerical debut rank) albums. Although the survival time for albums has decreased in the post period, this decrease is sharper for less popular albums (numerically higher debut rank). This is graphically illustrated in Figure 3, where predicted survival is plotted with respect to debut rank keeping other variables at their mean values, for both pre- and post-TS periods. The figure highlights the increasingly hazardous environment as an album debuts at higher (numerical) ranks.

The interaction *minor label* × *debut post*-TS suggests that minor labels have benefited considerably in

Table 4 Album Survival Estimation Results: Model with Interaction Effects

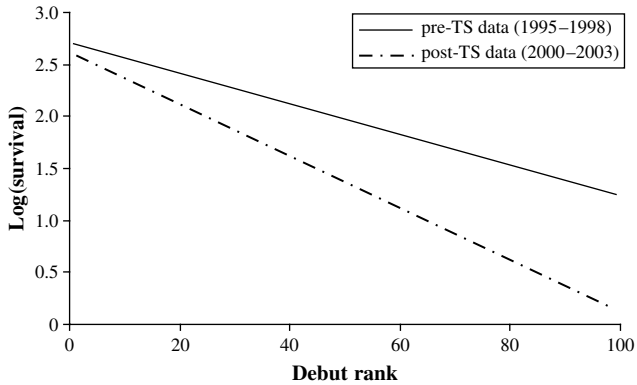
Parameter	(1) OLS with interaction effects		(2) Weibull AFT with interaction effects	
Constant	−0.62	(0.1)	8.24*	(1.8)
Debut rank	−0.014**	(12.4)	−0.02**	(21.8)
Debut post-TS	−0.20	(1.2)	−0.12	(0.9)
Albums released	0.35	(0.7)	−0.53	(1.2)
Superstar	0.30**	(3.4)	0.49**	(6.7)
Minor label	−0.43**	(3.9)	−0.30**	(3.3)
Solo male	−0.41**	(3.1)	−0.30**	(2.7)
Group	−0.45**	(3.6)	−0.43**	(4.1)
Holiday_month debut	0.19**	(2.6)	0.19**	(2.9)
Debut rank × *debut post*-TS	−0.01**	(6.6)	−0.005**	(2.9)
Minor label × *debut post*-TS	0.28*	(2.0)	0.19*	(1.9)
Superstar × *debut post*-TS	0.02	(0.2)	−0.09	(0.9)
Solo male × *debut post*-TS	0.11	(0.7)	0.02	(0.1)
Group × *debut post*-TS	0.09	(0.5)	0.007	(0.05)
Frailty variance			3.42**	(10.0)
σ (Weibull shape parameter)			3.56**	(12.4)
R^2	0.366		LL = −2,008	
Adjusted R^2	0.360			

*$p < 0.05$, **$p < 0.01$; t-statistics in parentheses; $n = 1{,}484$.

[7] The music labels may be engaged in activities to control the timing of album releases. The results with respect to the *holiday_month debut* variable must be interpreted with caution due to potential endogeneity. Reestimating the model without this variable suggests that it is orthogonal to other variables, as the estimates are similar with and without this variable. We had used *holiday_month debut* only as a control variable: inclusion of this variable improves the overall fit of the model only marginally.

[8] This result follows since the dependent variable is in logarithmic form while the explanatory variable is not. Comparing the pre- and post-TS periods yields a difference of $1 - e^{-0.54}$, which equates to a 42% decline.

Figure 3 Impact of Market and Other Factors on Album Survival



the post period with increased survival time. However, the main effect of *minor label* is still negative and significant, suggesting that major labels continue to survive longer but the difference narrowed. This is consistent with a variety of anecdotal evidence (Spellman 2006, Green 2004) suggesting that minor labels have adapted better to technological and market changes, and have, in fact, utilized file-sharing networks and other nontraditional methods to popularize their albums. Finally, there are no significant interaction effects with *superstar*, *solo male*, or *group*.

5.2. Phase 2—Analysis of Sharing on Survival

The previous analysis indicated that album survival has suffered in the post period—a period characterized by the presence of P2P-sharing networks. We now investigate whether this drop in survival might be attributable to the intensity of sharing.

Tables 5 and 6 present our estimation results for our two alternative measures of sharing, *shares_debut* (Table 5) and *shares_max* (Table 6). The estimates in Columns (1) and (2) of Table 5 are directly comparable and virtually the same, indicating little difference between AFT and OLS estimates. The R^2 values in Tables 5 and 6 are virtually identical and the coefficient estimates are also quite close, but there is one notable difference. The coefficient estimate tied to *shares_debut* (Table 5) is not statistically significant but the coefficient estimate tied to *shares_max* (Table 6) is significant. Although this suggests that sharing helps survival, these estimates could be spurious due to potential omitted variable bias. We do remark on this difference, but only in passing as we now address the omitted variable issues using the instrumental variable in the next section.

Table 5 Overall Impact of Sharing on Survival with ln(*shares_debut*): OLS and AFT Models

Parameter	(1) OLS		(2) Weibull AFT	
Constant	2.54**	(14.6)	2.37**	(13.4)
Debut rank	−0.03**	(17.0)	−0.03**	(17.8)
ln(*shares_debut*)	0.015	(0.8)	0.009	(0.5)
Superstar	0.24*	(1.9)	0.32**	(2.8)
Minor label	0.10	(0.9)	0.08	(0.9)
Solo male	−0.04	(0.3)	−0.06	(0.5)
Group	−0.19	(1.4)	−0.20	(1.5)
Holiday_month debut	0.55**	(3.7)	0.63**	(4.4)
Frailty variance			2.01**	(9.3)
σ (Weibull shape parameter)			3.28**	(9.6)
R^2	0.58			
Adjusted R^2	0.57		LL = −330	

$^*p < 0.05$, $^{**}p < 0.01$; t-statistics in parentheses; $n = 299$.

Table 6 Overall Impact of Sharing on Survival with ln(*shares_max*): OLS and AFT Models

Parameter	(1) OLS		(2) Weibull AFT	
Constant	2.37**	(13.4)	2.18**	(11.5)
Debut rank	−0.03**	(16.7)	−0.03**	(16.8)
ln(*shares_max*)	0.036*	(2.01)	0.03*	(1.8)
Superstar	0.25*	(2.02)	0.34**	(3.0)
Minor label	0.11	(1.1)	0.10	(1.1)
Solo male	−0.03	(0.2)	−0.05	(0.4)
Group	−0.17	(1.3)	−0.17	(1.3)
Holiday_month debut	0.53**	(3.6)	0.62**	(4.4)
Frailty variance			2.01**	(9.3)
σ (Weibull shape parameter)			3.28**	(9.6)
R^2	0.58		LL = −329	
Adjusted R^2	0.57			

$^*p < 0.05$, $^{**}p < 0.01$; t-statistics in parentheses; $n = 299$.

5.3. Omitted Variable Bias in Model Estimation: Results Using Instrument

We use an instrumental variable in our 2SLS squares estimation of Models (6) and (7). In the first stage we estimate (7), and in the second stage we estimate (6) with the predicted values from (7). We estimate this model with the four-month prior and post samples around the RIAA announcement event described in §3 (February–May 2003 and July–October 2003). Independent of the event, it is possible that the July–October 2003 album sample was inherently different from the February–May 2003 sample. Table 7 provides the average survival times for albums that debut between February–May and July–October for the similar period in the previous two years. The results of the t-tests suggest that overall survival across these two periods is quite similar. Hence we use the corresponding periods in 2003 for analysis.

Table 7 Average Survival Times (Number of Weeks)

Year	February–May	July–October	t-test of difference between means
2001	10.81	10.36	$p > 0.71$
2002	8.01	8.60	$p > 0.64$

Table 8 Average Survival Rates by Month of Album Chart Debut

Year	February	March	April	May	June
2001	11.5	9.21	9.94	11.80	10
2002	10.4	8.88	10.30	7.20	9.43
2003	6.7	7.96	7.90	6.33	8.30

A requisite condition for the choice of the instrument is that it is uncorrelated with the error term. Based on an extensive search on news sources (LexisNexis, Google, and Yahoo), it appears that the timing of the RIAA announcement was driven primarily by legal considerations, and not by album survival statistics on the chart (Gibson 2003, Stern 2003, Musgrove 2003, Zeidler 2003). Further, it is instructive to verify that survival times for albums in June were not unusual, in relation to the months immediately preceding it. The mean survival times for the months of February–June in the years 2001–2003 are shown in Table 8. We find nothing markedly different in the pattern of June across time in comparison to other months. For each of the months, including June, the survival rates have generally fallen over time. Thus the evidence we have offers no indication that June 2003 was strategically chosen for the announcement time based on survival patterns. These observations together point to the lack of correlation between the instrument and the error term.

Table 9 reports the estimation results with the instrument. We confine our discussion to the results with average sharing (*shares_debut*) as they are consistent (both in sign and significance) with those using maximum sharing (*shares_max*). In the first stage regression, the coefficient of the instrument, *RIAA announcement indicator,* is highly significant and negative. The estimated sharing decrease linked to the RIAA announcement (threat to sue file sharers) is approximately 80% (computed as $1 - e^{-1.61}$). *Debut rank* is also highly significant and negative, indicating that less popular albums (which debut at higher numerical rank) have significantly less sharing opportunities available. The first stage results also indicate that albums from superstars and those released by groups are shared less. The fit of the first stage model is approximately 38%. The second stage analysis indicates that, overall, sharing does not significantly affect survival (the sign is negative, but insignificant). This is in contrast to the results without the use of the instrument. The F-value of 53.0 for the first stage regression in our analysis indicates that the RIAA announcement is a strong instrument. The fit of the second stage model is 48%.

Recall that the overall survival estimation (§5.1) shows that survival of less popular albums has declined significantly in the post-TS period, but there was no significant change in the survival of more popular albums. Given that we find that overall file sharing does not affect album survival, we now consider whether such a differential decline in survival might be attributable to file sharing. To operationalize this, we estimate Model (5) by interacting ln(*shares*) with *debut rank*. With two endogenous variables, ln(*shares*) and ln(*shares*) × *debut rank*, we run two first stage regressions: ln(*shares*) with *RIAA announcement indicator* as instrument, followed by ln(*shares*) × *debut rank* with *RIAA announcement indicator* × *debut rank* as instrument. As before, our results are consistent for *shares_debut* and *shares_max*. Thus Table 10 provides just the results for *shared_debut*. The instrument appears only in the first stage of the estimation.

A key outcome in both first stage regressions (Table 10) is that the coefficient on the respective instrument is highly significant, suggesting that sharing has decreased after the RIAA announcement in June 2003 (a result that was the focus of the microlevel

Table 9 Overall Impact of Sharing on Survival Using Instrument

	Model estimates with ln(*shares_debut*)				Model estimates with ln(*shares_max*)			
Parameter	First stage		Second stage		First stage		Second stage	
Constant	6.00**	(19.5)	2.7**	(8.2)	6.86**	(20.7)	2.79**	(6.5)
Debut rank	−0.029**	(8.9)	−0.027**	(11.0)	−0.032**	(9.01)	−0.028**	(10.3)
ln(*shares_debut*)			−0.054	(0.9)				
ln(*shares_max*)							−0.056	(0.89)
Superstar	−1.12*	(2.8)	0.12	(0.9)	−1.37**	(5.2)	0.11	(0.7)
Minor label	−0.11	(0.5)	0.06	(0.6)	−0.29	(1.1)	0.05	(0.5)
Solo male	−0.33	(1.05)	−0.03	(0.02)	−0.41	(1.2)	−0.04	(0.25)
Group	−0.79**	(2.6)	−0.28	(1.9)	−1.07**	(3.3)	−0.30	(1.9)
RIAA announcement indicator (*instrument*)	−1.61**	(7.4)			−1.52**	(7.0)		
R^2	0.38		0.48		0.40		0.48	
Adjusted R^2	0.37		0.47		0.39		0.47	

$^{*}p < 0.05$, $^{**}p < 0.01$; t-statistics in parentheses; $n = 345$.

Table 10 Impact of Sharing on Survival: Interaction Effects

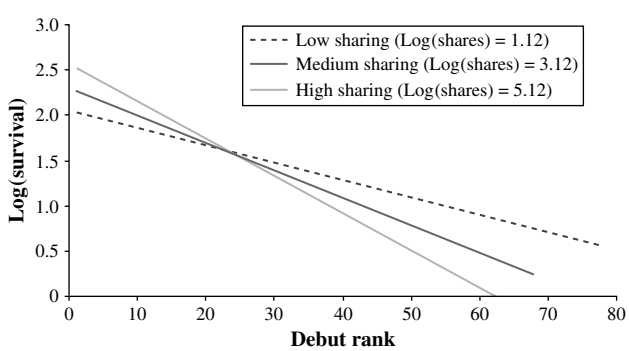
	First stage					
Parameter	ln(*shares_debut*)		ln(*shares_debut*) × *debut rank*		Second stage ln(*survival*)	
Constant	6.4**	(18.6)	47.0**	(2.7)	2.1**	(5.4)
Debut rank	−0.04**	(7.7)	2.14**	(8.1)	−0.012*	(1.8)
ln(*shares_debut*)					0.12	(1.5)
ln(*shares_debut*) × *debut rank*					−0.006**	(2.3)
Superstar	−1.16**	(4.6)	−49.2**	(4.0)	0.008	(0.1)
Minor label	−0.11	(0.5)	−5.25	(0.5)	0.05	(0.4)
Solo male	−0.34	(1.1)	1.65	(0.1)	0.03	(0.2)
Group	−0.76**	(2.5)	−11.5	(0.8)	−0.21	(1.3)
RIAA announcement indicator (*instrument*)	−2.23**	(7.1)	−16.3	(1.1)		
RIAA announcement indicator (*instrument*) × *debut rank*	0.018**	(2.7)	−0.75**	(2.3)		
R^2	0.40		0.30		0.40	
Adjusted R^2	0.38		0.29		0.39	

Note. Results are consistent with average sharing and maximum sharing. We report figures for average sharing.
$^*p < 0.05$, $^{**}p < 0.01$; t-statistics in parentheses; $n = 345$.

analysis reported in Bhattacharjee et al. 2006c). The adjusted R^2's of the first stage regressions are 37% and 39%, respectively. In the second stage regression, we find that the main effect of sharing, although estimated to be positive, is not significant. This result again suggests that the survival time of top albums are not adversely affected by sharing. The interaction term ln(*shares*) × *debut rank* is negative and significant. That is, less popular albums suffer more from increased sharing whereas top albums experience no significant deleterious impact on survival. For illustration purposes only, Figure 4 depicts the relationship between survival time and debut rank for three arbitrarily chosen levels of sharing—ln(*shares*) equal to 1.12 (low), 3.12 (medium; set equal to the actual average of ln(*shares*)), and 5.12 (high).

Figure 4 suggests that for albums that debut at a rank worse than about 20, sharing hurts survival (negative interaction term). The effect increases progressively as debut rank worsens. Moreover this effect appears more pronounced as sharing increases. Thus, a higher sharing of albums that are not very popular is initially linked to an adverse impact on the survivability of those albums. On the other hand, albums that debut near the top ranks do not appear to be adversely affected by sharing. Table 11 reports the mean survival times of albums before and after the RIAA announcement. Survival of albums debuting at 20 or better has not dramatically altered since the RIAA announcement (which reduced sharing); however, survival of albums debuting below 20 has shown an increase from 2.92 to 4.7 weeks when sharing was decreased.

Figure 4 Impact of Sharing on Survival



6. Discussions and Conclusion

We analyzed the survival of albums as measured by the number of weeks an album appeared on the Billboard 100 charts. During the two-year span, mid-1998 to mid-2000, several events with potentially important repercussions for the industry occurred, including introduction and rapid popularity of the MP3 music format and usage of Napster, passing of the Digital Millennium Copyright Act, surging popularity of DVDs, online chat rooms and games, and beginning of a downturn in the overall economy.

The first phase of our study was a comparative analysis of album survival before and after this event window (mid-1998 to mid-2000). We included the

Table 11 Mean Survival Time

	Debut rank ≤ 20	Debut rank > 20
February–May 2003	13.11 weeks	2.92 weeks
July–October 2003	13.65 weeks	4.70 weeks

following explanatory variables of album survival: debut rank of the album, reputation of the artist (as captured by superstar status), major or minor label promoting and distributing the album, artist descriptors (solo female/solo male/group), and holiday month debut. This phase considered the cumulative effect of technology and other factors on chart survival. Our second phase used actual sharing data to isolate the impacts of file sharing on chart success.

Phase 1 yielded several key results. We found that debut rank had a highly significant negative impact on album survival, an impact that was even more pronounced in the post-TS period for albums debuting lower on the charts (i.e., less popular albums). Average survival time on the charts decreased by 42% in the post-TS period. However closer inspection revealed that albums that debut at the top of charts did not suffer significantly shorter survival times. Rather, less popular albums had dramatically reduced survival times in the post-TS period. The superstar effect appeared to be alive and well, with albums by such performers surviving approximately 35% longer even after controlling for other variables. This superstar advantage remained unchanged in the post-TS period. Across both pre- and post-TS periods, albums promoted by major labels tended to survive longer than those promoted by minor labels. However, our results indicated that minor label albums have experienced a significant beneficial shift in the post-TS period and are surviving longer than before. If anecdotal evidence (Spellman 2006, Green 2004) is correct in suggesting that minor labels have utilized file-sharing networks to popularize their albums, then the majors have an added incentive to fight file sharing. Finally, albums by female artists continue to have a survival advantage over those by solo male artists and groups.

In Phase 2 we used the June 2003 RIAA announcement as an instrumental variable in our analysis of the impact of sharing activity. Although we found no significant impact of sharing for top debut ranked albums, we did find that the level of sharing had a significantly negative impact for lower debut ranked albums. The results suggest that a music label offering a less popular album might be hurt by negative word-of-mouth that can quickly chill demand. A similar negative information-sharing effect was previously shown for online book sales (Chevalier and Mazylin 2006).

In the new music market landscape, what is it that firms can do to enhance success? Our results emphasize that superstars and debut rank are important markers of success. Thus firms that do the best job in enlisting superstars and successfully utilizing prerelease marketing to impact debut rank are well-positioned to succeed in the new marketplace. Finally, there is evidence that minor labels are closing the gap with the major labels. The innovative approaches adopted by the minor labels might provide strategies for major labels to emulate. One approach that minor labels have been adept at is embracing the use of technologies to brand and reach out to potential customers. In this vein, it has been suggested that sharing through online networks might have beneficial sampling and word-of-mouth effects. Our Phase 2 results suggest otherwise, especially for albums debuting lower on the charts.

Our analysis has some limitations because we consider sharing that occurs after the album has made an appearance on the charts. File sharing may take place before chart appearance, and such sharing could influence "if, when, and where" an album appears on the charts. One can consider *prechart-appearance sharing* as an omitted variable in our model. We use an instrumental variable approach and incorporating *prechart-appearance sharing* would not bias our findings. However, music firms (labels) may be able to impact prerelease and/or prechart-appearance sharing and thus impact chart debut rank. The firms themselves may even offer prerelease sampling opportunities. We have begun work to track and evaluate the impact of prerelease sharing.

7. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Acknowledgments

The authors gratefully acknowledge the Treibick Family Endowed Chair, the Treibick Electronic Commerce Initiative, the XEROX CITI Endowment Fund, the GE Endowed Professor Fund, The Center for Internet Data and Research Intelligence Services (CIDRIS), the Connecticut Information Technology Institute (CITI), and the Gladstein Endowed MIS Research Lab for support that made this work possible. The last author acknowledges the generous financial support of National Science Foundation (NSF) through the CAREER Award CNS-0546009. The authors thank the department editor, associate editor, and two reviewers for many constructive suggestions.

References

Adler, M. 1985. Stardom and talent. *Amer. Econom. Rev.* **75**(1) 208–212.

Alexander, P. J. 2002. Digital distribution, free riders, and market structure: The case of the music recording industry. *Rev. Indust. Organ.* **20**(2) 151–161.

Bakos, Y., E. Brynjolfsson, D. Lichtman. 1999. Shared information goods. *J. Law Econom.* **42**(1) 117–155.

Barua, A., P. Konana, A. B. Whinston, F. Yin. 2001. Driving e-business excellence. *Sloan Management Rev.* **43**(1) 36–44.

Bell, A. M. 2002. Locally interdependent preferences in a general equilibrium environment. *J. Econom. Behav. Organ.* **47** 309–333.

Belzil, C. 1995. Unemployment insurance and unemployment over time: An analysis with event history data. *Rev. Econom. Statist.* **77**(1) 113–126.

Bemuso.com. 2006. Labels and publishers: All about record labels and music publishers. Retrieved May 26, 2006, http://www.bemuso.com/musicbiz/labelsandpublishers.html.

Bhattacharjee, S., R. D. Gopal, K. Lertwachara, J. R. Marsden. 2006a. Whatever happened to payola? An empirical analysis of online music sharing. *Decision Support Systems* **42**(2) 104–120.

Bhattacharjee, S., R. D. Gopal, K. Lertwachara, J. R. Marsden. 2006b. Retailer strategies in the presence of online music sharing. *J. Management Inform. Systems* **23**(1) 129–159.

Bhattacharjee, S., R. D. Gopal, K. Lertwachara, J. R. Marsden. 2006c. Impact of legal threats on individual behavior: An analysis of music industry actions and online music sharing. *J. Law Econom.* **49**(1) 91–114.

Billboard.com. 2006. Billboard methodology. Retrieved May 26, 2006, http://www.billboard.com/bbcom/about_us/bbmethodology.jsp.

Boston, W. 2000. Bertelsmann is betting that users, content rule with Napster deal. *Wall Street J.* (November 2).

Bradburn, M. J., T. G. Clark, S. B. Love, D. G. Altman. 2003. Survival analysis part II: Multivariate data analysis—an introduction to concepts and methods. *British J. Cancer* **89** 431–436.

Bradlow, E. T., P. S. Fader. 2001. A Bayesian lifetime model for the "Hot 100" billboard songs. *J. Amer. Statist. Assoc.* **96** 368–381.

Brynjolfsson, E., M. Smith. 2000. Frictionless commerce? A comparison of Internet and conventional retailers. *Management Sci.* **46**(4) 563–585.

Chevalier, J. A., D. Mazylin. 2006. The effect of word of mouth on sales: Online book reviews. *J. Marketing Res.* **43**(3) 345–354.

Chung, K. H., A. K. Cox. 1994. A stochastic model of superstardom: An application of the Yule distribution. *Rev. Econom. Statist.* **76**(4) 771–775.

Crain, W. M., R. D. Tollison. 2002. Consumer choice and the popular music industry: A test of the superstar theory. *Empirica* **29**(1) 1–9.

Cunningham, B. M., P. J. Alexander, N. Adilov. 2004. Peer-to-peer file sharing communities. *Inform. Econom. Policy* **16**(2) 197–213.

De Vany, A., W. D. Walls. 1999. Uncertainty in the movie industry: Does star power reduce the terror of the box office? *J. Cultural Econom.* **23** 285–318.

De Vany, A., W. D. Walls. 2004. Motion picture profit, the stable paretian hypothesis and the curse of the superstar. *J. Econom. Dynam. Control* **28**(6) 1035–1057.

Dhar, R., K. Wertenbroch. 2000. Consumer choice between hedonic and utilitarian goods. *J. Marketing Res.* **37** 60–71.

Evangelista, B. 2000. CD Soars after net release. *San Francisco Chronicle* (October 12).

Fader, P. S. 2000. Expert Report of Peter S. Fader, Ph.D. in Record Companies and Music Publishers vs. Napster, July 26. United States District Court, Northern District of California.

Gibson, O. 2003. Gabriel's call to arms: OD2 has signed deals with all five major record labels to blaze a trail for paid-for music downloads on the net. *Guardian* (April 28) 46.

Goodley, S. 2003. Disharmony over music pirates on the Internet. *Telegraph* (January 9).

Google, Inc. 2006. Google Press Center Zeitgeist. Retrieved May 26, 2006, http://www.google.com/press/zeitgeist.html.

Gopal, R. D., S. Bhattacharjee, G. L. Sanders. 2006. Do artists benefit from online music sharing? *J. Bus.* **79**(3) 1503–1534.

Gopal, R. D., G. L. Sanders, S. Bhattacharjee, M. Agrawal, S. Wagner. 2004. A behavioral model of digital music piracy. *J. Organ. Comput. Electron. Commerce* **14**(2) 89–105.

Graham, J. 2005a. Morpheus takes a stand—again. *USA TODAY* (March 10), http://www.usatoday.com/money/industries/technology/2005-03-10-morpheus-usat_x.htm.

Graham, L. 2005b. iTunes more popular than most peer-to-peer file sharing services. *NPD Group* (June 7), http://www.npd.com/dynamic/releases/press_050607.htm.

Green, H. 2004. Kissing off the big music labels. *Businessweek* (September 6).

Hamlen, W. A., Jr. 1991. Superstardom in popular music: Empirical evidence. *Rev. Econom. Statist.* **73**(4) 729–733.

Harmon, A. 2003. Suit settled for students downloading music online. *New York Times* (May 2) A22.

International Federation of Phonographic Industry. 2005. IFPI releases definitive statistics on global market for recorded music. (August 2) London, UK, http://www.ifpi.org/site-content/publications/rin_order.html.

Kalbfleisch, J. D., R. L. Prentice. 2002. *The Statistical Analysis of Failure Time Data*. J. Wiley, Hoboken, NJ.

Kiefer, N. M. 1988. Analysis of grouped duration data. *Contemporary Math.* **80** 107–139.

King, B. 2000a. Napster: Music's friend or foe? *Wired.com News* (June 14).

King, B. 2000b. Napster's Good? Bad? Er, What? *Wired.com News* (June 15).

Knab, C. 2001. Inside record labels: Organizing things. Fourfront Media Music (August), http://www.musicbizacademy.com/knab/articles/insidelabels.htm.

Krider, R. E., C. B. Weinberg. 1998. Competitive dynamics and the introduction of new products: The motion picture timing game. *J. Marketing Res.* **35**(February) 1–15.

Leibenstein, H. 1970. Bandwagon, snob, and veblen effects in the theory of consumer demand. *Quart. J. Econom.* **64** 183–207.

Liebowitz, S. 2004. Will MP3 downloads annihilate the record industry? The evidence so far. *Adv. Stud. Entrepreneurship, Innovation, Econom. Growth* **15** 229–260.

Liebowitz, S. 2006. File sharing: Creative destruction or just plain destruction? *J. Law Econom.* **49**(1) 1–28.

MacDonald, G. M. 1988. The economics of rising stars. *Amer. Econom. Rev.* **78**(1) 155–166.

Mathews, A. W. 2000. Sampling free music over the Internet often leads to a sale—Poll adds to conflicting data as recording industry sorts out web's impact. *Wall Street J.* (June 15) A3.

Mathews, A. W., M. Peers. 2000. Teen music buying dropped last year, according to data. *Wall Street J.* (June 26) B14.

Moe, W. W., P. S. Fader. 2001. Modeling hedonic portfolio products: A joint segmentation analysis of music compact disc sales. *J. Marketing Res.* **38** 376–385.

Montgomery, A. L., W. W. Moe. 2000. Should record companies pay for radio airplay? Investigating the relationship between album sales and radio airplay. Working paper, University of Pennsylvania, Philadelphia, PA.

Musgrove, M. RIAA plans to sue music swappers; no more warnings to individuals. *Washington Post* (June 26) E01.

Nelson, P. 1970. Information and consumer behavior. *J. Political Econom.* **78**(2) 311–329.

Oberholzer, F., K. Strumpf. 2007. The effect of file sharing on record sales: An empirical analysis. *J. Political Econom.* **115**(1) 1–42.

Pastore, M. 2001. Net users finding P2P music alternatives. (October 11), http://www.clickz.com/stats/sectors/software/article.php/901921.

Peers, M., L. Gomes. 2000. Music CD sales suffer in stores near "wired" colleges, study says. *Wall Street J.* (June 13) A4.

Ravid, S. A. 1999. Information, blockbusters and stars: A study of the film industry. *J. Bus.* **72**(4) 463–492.

Recording Industry Association of America (RIAA). 2003. Yearend statistics. Retrieved May 26, 2006, http://www.riaa.org.

Rosen, S. 1981. The economics of superstars. *Amer. Econom. Rev.* **71**(5) 845–858.

Seabrook, J. 2003. The money note: Can the record business survive? *New Yorker* (July) 42–55.

Spellman, P. 2006. *Indie Power: A Business-Building Guide for Record Labels, Music Production Houses, and Merchant Musicians*, 2nd ed. MBS Business Media, Boston, MA, http://www.mbsolutions.com/books/howtostart.html.

Stern, C. 2003. Verizon identifies download suspects; firm says fight goes on to guard privacy. *Washington Post* (June 6) E05.

Strobl, E. A., C. Tucker. 2000. The dynamics of chart success in the U.K. pre-recorded popular music industry. *J. Cultural Econom.* **24** 113–134.

Towse, R. 1992. The earnings of singers: An economic analysis. R. Towse, A. Khakee, eds. *Cultural Economics*. Springer-Verlag, Heidelberg, Germany, 209–217.

Walls, W. D. 2005a. Modeling movie success when "Nobody Knows Anything": Conditional stable-distribution analysis of film returns. *J. Cultural Econom.* **29** 177–190.

Walls, W. D. 2005b. Modeling heavy tails and skewness in film returns. *Appl. Financial Econom.* **15** 1181–1188.

Walls, W. D., F. Rusco. 2004. Independent film finance, pre-sale agreements, and the distribution of film earnings. V. Ginsburgh, ed. *Economics of Art and Culture: Invited Papers at the 12th Internat. Conf. on the Association for Cultural Economics Internat.* Contributions to Economic Analysis 260. Elsevier Science, Amsterdam, The Netherlands, 19–32.

Yamada, M., H. Kato. 2002. A structural analysis of sales patterns of music CDs. *INFORMS Marketing Sci. Conf., Edmonton, Alberta, Canada.*

Zeidler, S. 2003. Streamcast says to mount peer-to-peer protest. *Yahoo Technology News* (June 26), http://in.tech.yahoo.com/030627/137/25h1i.html.