Stochastic dynamics of music album lifecycle: An analysis of the new market landscape

Sudip Bhattacharjee , Ram D. Gopal   Kaveepan Lertwachara   James R. Marsden

Abstract

The rapid emergence of file sharing networks has enabled easier information dissemination and product access to potential consumers. At the same time, copyright protection technologies for securing digital products have been compromised repeatedly. To analyze the ensuing impacts on the market landscape for music products (a digital good), we develop a stochastic model of distribution of music album longevity on the Billboard Chart. We find that since the advent of file sharing networks and other market forces (such as legal changes in copyright laws, introduction of digital rights management systems and legitimate online music download offerings), the lifecycle of music albums has shortened with lowered probabilities of survival for each week. While the probability of survival past the first week is markedly lower, the future portends well for albums that do survive on the charts beyond the first week. This is consistent with the rapid diffusion of information on music albums in the changed market landscape. Integrating this insight with user activity on online computer networks, we estimate the continued success of albums on the charts. This analysis helps to create a more dynamic decision process on resource allocation to promote and market music products. Using the robust stochastic model parameters as a benchmark, we estimate a logistic regression model which helps us make quality decisions in an uncertain environment through early monitoring of the success of music albums.

1. Introduction

Maintaining security in the digital world continues to grow in complexity. Firms must protect operating hardware and sensitive data against increasingly innovative threats. With the emergence of digital goods comes a new security front where firms face the reproduction and rapid distribution of the digital goods themselves. Certainly firms already have had to protect many of their goods from "knockoffs", but protecting digital goods represents a new level of challenge since the cost of copying and distributing such goods is virtually zero and can occur extensively within very short periods of time. The music industry has been the "poster industry" for facing such threats. The industry's goods are digital by nature. Further, the appearance of peer-to-peer (P2P) networks offered the means to copy (download) the goods and distribute (share) them rapidly.

Through its industry association, the Recording Industry Association of America (RIAA), the music industry has continued to pursue mostly legal and technological strategies to eradicate the security threat of illegal copying and distribution. Recent studies by Bhattacharjee et al. (2006a, b) provide evidence that, while individuals have tended to reduce their own sharing activity in response to RIAA legal threats and actions, significant piracy opportunity remains. While individual firms may take steps to secure their digital goods, such constraints have two major drawbacks. First, such measures tend to impede the

consumer's use of the digital good since they can restrict portability or require additional steps (e.g., security actions) that reduce consumer utility (Halderman, 2002). Second, the measures have proven less than "foolproof" and rather easily beaten (Reuters, 2002; Felten, 2003). Sony BMG's recent use of a rootkit with the XCP technology (Bergstein, 2005; Reuters, 2005) provides a prominent illustration of how an attempted technological security constraint can backfire (Bradley, 2005):

Part of Sony's anti-pirating strategy is that some of its music will play only with media software included on the CD. When a user inserts the CD, he or she is asked to consent to an "end-user licensing agreement" for a Digital Rights Management application. If the user agrees, the rootkit automatically installs and hides (or "cloaks") a suite of DRM software.

Unfortunately, the rootkit application created a possible secret backdoor for hackers which led Sony to "hastily" post a patch. However, the tool to remove the XCP application itself created new vulnerabilities (Russinovich, 2005). The tale continues as California quickly filed suit under both unfair and deceptive trace acts and consumer protection acts, Texas filed suit for including "spyware" in its media player, and the Electronic Frontier Foundation filed suit seeking class-action status over its copy-protection software (Smith, 2005). A posting (by concord (198387), 11/10/05, #13996982) in slashdot.org's bulletin board offers the following perspective on the Sony anti-piracy actions:

Now for the first time it is actually safer to download and listen to pirated music then (sic) it is to purchase and use compact disks and dvds. Piracy will become a matter of self-preservation.

In addition, security professionals have consistently noted that all CD and DVD encryption techniques that have been tried by the entertainment industry have been broken by savvy consumers (Patrizio, 1999; Schneier, 2000; Craver et al., 2001; Associated Press, 2003; Clarke, 2005). Given wide dissemination of the encrypted music product among users (factors that make breaking encryption easier), it is not unusual to observe such copy protection technologies being defeated by smart users (Bergstein, 2005; Felten, 2005).

Thus, a firm considering possible actions to protect its digital product may find little return in costly technological and legal anti-piracy measures. But can the firms identify and respond to the changing market they face? The post-1998–1999 period is characterized by consumers who increasingly search and consume music products in digital formats. Here we focus attention on what significant changes have occurred in the landscape of music products and their market lifecycle since the introduction of significant new technology, including P2P networks and other market forces (including online music stores, higher penetration of broadband into homes, digital rights management (SDMI initiative) and evolving copyright laws (DMCA 1998, Sonny Bono Copyright Term Extension Act 1998). We first develop an analytical model of music album lifecycle to provide a robust foundation to develop effective decision making tools for a music company to better manage its music products in the market place. The model demonstrates how the pattern of album lifecycle has undergone a shift in the years following the introduction of new technologies and other market forces. Following the analytic model, we utilize actual user activities on an online P2P network to fine-tune the decision-making process. We show how incorporating actual user behavior from interactive P2P computer networks helps a decision maker to better predict and respond to market success of a digital good in a dynamic environment. A firm's ability to act with these decision tools, which combine product lifecycle analytics with analysis of consumer actions on online computer networks, would provide greater market value protection for the firm's digital products than would technological and legal anti-piracy measures alone.

2. The landscape—rankings and survival longevity

In a number of domains—including music, movies, books, university sports and academics—rankings are the yardsticks to measure success. Appearance and longevity of survival on ranking charts are important for market success and job security. Rankings have limited slots (e.g., top 10, top 25 or top 100) and are reported on a periodic basis (ranging from weekly for music charts or in-season sports to annually for business school rankings).

High rankings and longevity on ranking charts would seem to have inherent links to the concept of "superstars," a phenomenon studied by Rosen (1981). Following Rosen's initial work, Adler (1985) suggested the existence of the superstar phenomenon in artistic industries where only a relatively small number of music artists and their products garner enormous success. Adler argued that consumers minimize the cost of search by simply choosing artists who are already popular among other consumers. Adler's "concentration of success" phenomenon has been empirically studied by several authors, many of whom found accurate estimation results using a straight-forward stochastic process suggested by Simon (1955) and Yule (1924). Examples cover quite a range and include Albert's (1998) analysis of motion pictures, Cox and Chung's (1991) study of research output in academics, Simon's (1955) examination of the distribution of words in prose and Levene et al.'s (2002) consideration of the growth of Internet websites. Approximately 30,000 albums are released annually by the major music labels alone (Goodley, 2003). Given that a mere handful of successful albums can significantly affect the profitability of a music label, it is critically important for the labels to have an estimation of the potential lifecycle of the albums released early in their release period. This would enable them to form informed decisions and channel limited marketing and promotional budgets towards potential winners.

But what happens if the landscape changes significantly and past business practices do not apply as well? What happens when advances occur in markets that make a consumer's search for information and product access far easier? Does ranking longevity, or lifecycle on the chart, change dramatically? In fact, in the past few years, the music industry has seen such a technological and market revolution. Easier search for information and product sampling is an integral part of buying an experience product such as music. The advent of MP3 and online file-sharing technologies now allow consumers to access and exchange millions of digitized music files over P2P networks.[1,2]

We develop a stochastic model of the distribution of album longevity on the Billboard Top 100 Chart.[3] We estimate the model annually for periods before and after the major technological and market changes, that is, the introduction of MP3, broadband and the Napster P2P online sharing technology that took place over the 1998–1999 period. What we find is that, despite declining numbers of new album releases after 1999 (Ziemann, 2002), the probability of survival on the Billboard Chart had a major shift downward. We use this survival information and develop a logistic regression model that incorporates consumer behavior on online file-sharing networks. This is used to estimate the continued success of albums on the Billboard Chart. We emphasize that the same stochastic model form yields similar useful fit results for the differing periods with, of course, different parametric estimation values. Continued refreshing of the model estimation can be utilized by firms as a benchmark to adjust their decision making on individual albums as the market continues to shift over time.

3. The stochastic model of survival

Rankings and longevity on the charts is a key indicator of a music album's success and is closely followed by music labels and music industry analysts each week. Since 1913, *Billboard* magazine has provided weekly summary chart information based on sales of music recordings (Bradlow and Fader, 2001). The Billboard Top 200 Chart is based on "… a national sample of retail store sales reports collected, compiled and provided by Nielsen Soundscan" (from Billboard website). We use the freely available list of the weekly top 100 albums in our analysis. Based on empirical observation, we assume that once an album drops off the Billboard Top 100, the album does not re-appear on the chart.[4] Thus each week, some albums drop off the ranking chart and an equal number of albums appear for the first

[1]MP3 is a commonly used audio compression technology.
[2]Recently, legal threats from RIAA may be changing the landscape a bit (e.g., see Bhattacharjee et al. (2006a)).
[3]The top 100 albums per week are available free at /chart_display. jsp?f The + Billboard + 200&pageNumber Top + 1 10&g Albums.
[4]Our empirical data show that the probability of an album re appearing on the chart is minimal.

time. At the end of a hypothetical "first" week of the chart, we would have 100 albums that have appeared on the chart for exactly 1 week.

Let p_i denote the probability that an album will remain on the chart for 1 more week after having been on the chart for exactly i weeks. In week 2, the expected number of albums that would remain on the chart is $100p_1$. The expected number of albums that drop out of the chart after the first week is $100(1-p_1)$ which is also the same number of albums expected to enter the chart for the second week since there must be a total of 100 active albums on the Billboard Top 100 chart in any given week. That is, we model a stochastic process with one absorptive state that might be termed "falling off the chart." Table 1 details the stochastic process for the first three periods.

More formally, let $C_{k,i}$ indicate the number of albums that appear on the kth week's Billboard Chart and have appeared for i weeks ($i = 1, \ldots, k$). $C_{12,5}$ would be the number of albums on the 12th week's chart that had appeared for 5 weeks (charts 8–12). Let $D_{k-1,w}$ be the number of albums that appeared on week $k-1$'s chart, do not appear on week k's chart, and which were on the charts for w weeks. $D_{21,4}$ would be the number of albums that met the following criteria: appeared on chart 21, did not appear on chart 22, and appeared on the chart for 4 weeks (from weeks 18 to 21).

The following summarize the stochastic process (for convenience, we ignore expected value signs and use a general "n" rather than the 100 total for our Billboard Chart):

$$\sum_{i=1}^{k} C_{k,i} = n$$

(in each week, k, there must be $n = 100$ albums on the chart so summing across various weeks on the chart, from 1 week to k weeks, must yield 100 albums).

$$TD_{k-1,w} = \sum_{j=1}^{k-1} D_{j,w}$$

(at the end of the $k-1$ chart, this is the total number of albums that were on the chart exactly w weeks before falling off).

$$\sum_{w=1}^{k-1} D_{k-1,w} = C_{k,1}$$

(the number of new albums coming onto the kth chart must be equal to the number of albums that fell off the charts after week $k-1$; we sum across those that were on for one week, two weeks, up to $k-1$ weeks to get the total number that dropped off after the k-1st chart).

$$C_{k,i} = p_{i-1} C_{k-1,i-1}$$

(*Note*: expected value operators are not shown) (expected number of albums that appear on chart k and have then survived for i weeks. Value is obtained by multiplying the eligible albums (those which have been on the chart $i-1$ weeks) times the

Table 1
Illustration of stochastic process

	Expected number of albums that have been and are currently on the chart for:				Expected number of albums that had dropped out of the chart after:			
	1 week	2 weeks	3 weeks	4 weeks	1 week	2 weeks	3 weeks	4 weeks
Week1	100	0	0	0	0	0	0	0
Week 2	$100(1-p_1)$	$100p_1$	0	0	$100(1-p_1)$	0	0	0
Week 3	$100p_1(1-p_2)+100(1-p_1)^2$	$100(1-p_1)p_1$	$100p_1p_2$	0	$100(1-p_1)+100(1-p_1)^2$	$100p_1(1-p_2)$	0	0

probability of remaining on the chart for an ith week given album was on the chart $i-1$ weeks).

$$D_{k-1,w} = (1 - p_w)C_{k-1,w}$$

(expected number of albums that drop off after week $k-1$ having been on the charts for w weeks is obtained by multiplying the probability of dropping off given that the album has been on the charts w weeks times the number of eligible albums, those that have been on the charts w weeks in the $k-1$ chart).

Let T_k be the total number of music albums that had appeared on the chart at the end of week k. The steady state ($TD_{k,w}/T_k$) for this stochastic model can be shown to be (see Appendix A):

$$\lim_{k \to \infty \leftarrow} \frac{TD_{k,w}}{T_k} = (1 - p_w)\prod_{j=1}^{w-1} p_j.$$

Note that when $p_i = p$ for all i, the steady state is that of the geometric distribution (see similar distributions, e.g., Chung and Cox, 1994, going back to Simon, 1955 and Yule, 1924).

For $k > i$, by expansion, we have $C_{k,i} = p_{i-1}p_{i-2}, \dots,$ $p_2 p_1 C_{k-i,1}$. At one extreme, it is possible that all p_i's are equal, that is, that the probability of an album remaining on the Billboard Chart is independent of the number of weeks the album has already been on the chart. At the other extreme, all p_i values could be different. From empirical observations, we choose a step function for the p_i values as explained below. It is consistently the case (see below) that the largest "falling off the chart" occurs for albums that have been on the chart just one week. In addition, there appears to be at least one clear "shift" point. After albums have been on the chart for some number of weeks, the probability of remaining on the chart shifts upward. As an example, for three shift points (four p's), our model would utilize p_i values as follows:

$$p_1 < p_2 = p_3 = \leftarrow \cdot = p_a < p_{a+1} = p_{a+2}$$
$$= \leftarrow \cdot = p_b < p_{b+1} = p_{b+2} = \leftarrow \cdot = p_{k-1}.$$

4. Data and stochastic model estimation

Our Billboard Chart data includes all weekly data over the periods 1995–1997 and 2000–2002, the pre- and post-change periods in the markets. As explained above, we investigate whether the market landscape has shifted and, if so, what the implications are for music firms. We note that the data observations are not a random sample and, in reality, are the entire populations for the two periods studied. We view them as all realizations from a stochastic process for the selected periods.

Preliminary evaluation of the data and discussions with individuals knowledgeable about the industry suggest that the album "chart drop-off process" is quite rapid. During the years studied, while one album did in fact remain on the chart for 151 weeks, the vast majority of albums had much shorter chart life spans. Table 2 summarizes the number and percentage of albums that debuted in a given year and the number of weeks they remained on the chart before departing.

Since the majority of albums dropped off the chart within the first three months, we decided to focus on modeling and estimating a stochastic process of that length.[5] As outlined in the previous section, the family of stochastic processes we are utilizing includes an array of shift points from 1 to 13. That is, one case would be where the probability of falling off the chart remains the same no matter the number of weeks the album has appeared on the chart. The other extreme would be 13 shift points where the probability is different for each of the 13 possible weeks an album could have remained on the chart.

We used a brute force solution process beginning with one p value. Table 3 summarizes the outcomes for the single p stochastic process.

We used the χ^2 goodness of fit test for the null hypothesis that the stochastic model is appropriate for the observed process. The α^* values (normally indicated as p values, but we use α^* here to avoid any confusion) indicate that level of significance at which we would begin rejecting the null hypothesis. That is, we only reject the null hypothesis for a level of significance greater than α^*. Thus, at a 0.05 level of significance we would reject the proposed stochastic model

---

[5]This also avoided the inclusion of periods with expected frequencies less than 5, a consideration when we analyze the appropriateness of our stochastic model.

Table 2
Album dropoff behavior on charts

	Year of debut					
	1995	1996	1997	2000	2001	2002
Number of albums debuting on the Billboard 100 during the year	323	339	361	341	366	383
Total number of albums dropping off after one week	43 (13.3%)	41 (12.1%)	55 (15.2%)	86 (25.2%)	91 (24.9%)	91 (23.8%)
Total number of albums dropping off after weeks 1 4	122 (37.8%)	119 (35.1%)	120 (33.2%)	190 (55.7%)	189 (51.6%)	197 (51.4%)
Total number of albums dropping off after weeks 1 8	162 (50.2%)	169 (49.9%)	189 (52.4%)	252 (73.9%)	262 (71.6%)	282 (73.6%)
Total number of albums dropping off after weeks 1 13 (3 months)	205 (63.5%)	222 (65.5%)	234 (64.8%)	284 (83.3%)	310 (84.7%)	331 (86.4%)
Total number of albums dropping off after weeks 1 20	247 (76.5%)	267 (78.7%)	282 (78.1%)	290 (85.0%)	324 (88.5%)	344 (89.8%)

Table 3
Single p stochastic process estimates

	1995	1996	1997	2000	2001	2002
$\hat{p}$	0.83	0.85	0.84	0.76	0.79	0.80
Computed χ^2	23.47	16.62	29.29	13.89	28.14	32.93
α^*	0.0240	0.1646	0.0036	0.3077	0.0053	0.0010
Degree of freedom	12	12	12	12	12	12

Table 4
Multiple p stochastic process estimates

	1995	1996	1997	2000	2001	2002
$\hat{p}_1$	0.86	0.89	0.86	0.72	0.74	0.76
$\hat{p}_2$	0.92	0.90	0.91	0.82	0.85	0.85
Weeks: $\hat{p}_1$	2	3	2	1	1	1
Computed χ^2	11.16	7.99	17.60	7.52	8.16	14.46
α^*	0.4301	0.7138	0.0913	0.7556	0.6993	0.2087
Degree of freedom	11	11	11	11	11	11

for years 1995, 1997, 2001 and 2002. We would accept the null hypothesis (the proposed model) for years 1996 and 2000.

We then repeated the brute force solution process for a model with one shift point (two p's). The results are summarized in Table 4. This time, using the χ^2 goodness of fit test and a 0.05 level of significance, we would accept the null hypothesis of model appropriateness for all years. The α^* levels ranged were quite high: 0.4301 (1995), 0.7138 (1996), 0.0913 (1997), 0.7556 (2000), 0.6993 (2001) and 0.2087 (2002). Repeating the process for two shift points (three p's), we found little improvement and thus focus on p_1, p_2 stochastic model with parameter estimates as indicated in Table 4 and illustrated in Fig. 1.

Consider the model specification differences in the periods before (1995–1997) and after (2000–2002). We note the following:

(i) the shift period occurs earlier (the $\hat{p}_1$ estimate has only a one week duration in each of the 2000–2002 years compared to two or three week duration in each of the 1995–1997 years); and,

(ii) in every year during the 2000–2002 period, the $\hat{p}_1$ and $\hat{p}_2$ values are less than the corresponding $\hat{p}_1$ and $\hat{p}_2$ values for each year in the 1995–1997 period.

The 2000–2002 values of $\hat{p}_1$ are 0.72, 0.74 and 0.76, respectively, compared to $\hat{p}_1$ values for 1995–1997 of 0.86, 0.89 and 0.88. The 2000–2002 values of $\hat{p}_2$ are 0.82, 0.85 and 0.85, respectively, compared to $\hat{p}_2$ values for 1995–1997 of 0.92, 0.90 and 0.91. These outcomes suggest quite different parameters for our stochastic model before and after the 1998–1999 market shift. The probability of remaining on the chart after one week fell by an average of 0.1 (0.84 before and 0.74 after). Further, the probability of remaining on the chart (after the process shift) was, on average, 0.07 lower (0.91 before and 0.84 after). Table 5 provides the estimated probabilities of remaining on the chart for a set number of weeks for each of the years.

The results provided in Tables 2, 4 and 5 and Fig. 1 indicate a shift in chart lifecycles of music albums following the technological and market innovations of 1998–1999. In the 1995–1997 period, 50% of albums that appeared on the chart would be expected to last at least seven weeks and at
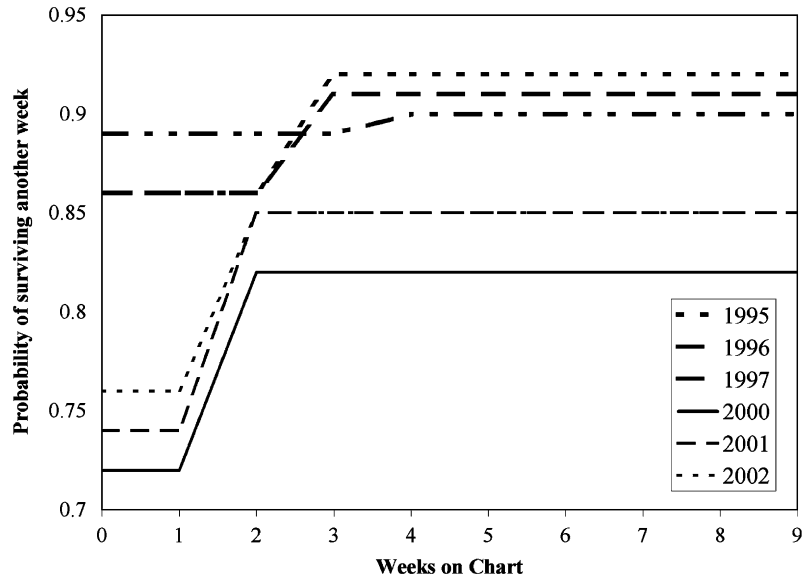
Fig. 1. Multiple p stochastic process.

Table 5
Estimated probability of survival on chart

Weeks	1995	1996	1997	2000	2001	2002
2	0.86	0.89	0.86	0.72	0.74	0.76
3	0.74	0.79	0.74	0.59	0.63	0.65
4	0.68	0.70	0.67	0.49	0.53	0.55
5	0.63	0.63	0.61	0.40	0.45	0.47
6	0.58	0.57	0.56	0.33	0.39	0.40
7	0.53	0.51	0.51	0.27	0.33	0.34
8	0.49	0.46	0.46	0.22	0.28	0.29
9	0.45	0.42	0.42	0.18	0.24	0.24
10	0.41	0.37	0.38	0.15	0.20	0.21
11	0.38	0.34	0.35	0.12	0.17	0.18
12	0.35	0.30	0.32	0.10	0.15	0.15
13	0.32	0.27	0.29	0.08	0.12	0.13

least 40% would be expected to last nine weeks. In the 2000–2002 period, less than 50% would be expected to last 5 weeks. Less than 40% would be expected to make it to the sixth week. It is important to note that although the probability of remaining on the chart has changed before and after the 1998–1999 period, the structural robustness of the model with one shift point (two p's) is maintained in both periods.

5. Decision making incorporating online user behavior

In the face of file sharing networks that enable wide-spread sharing and downloading of music in digital forms, music companies have felt pressured to take steps to simultaneously safeguard their digital products and bolster their market performance. As discussed before, their strategic decisions and actions have thus far focused on incorporating security mechanisms in the digital products themselves and on legal threats and actions against both operators of file sharing networks and individual file sharers (Bhattacharjee et al., 2006a, b). As illustrated by the Sony BMG situation earlier, embedded security measures can frustrate consumers and have significant negative impacts. Further, no security measure used by the entertainment industry so far has been foolproof. As Bhattacharjee et al. (2006a, b) detail, legal threats and actions have reduced sharing by individuals, but significant piracy opportunities remain. Further, such actions have been industry actions rather than individual firm actions.

Our focus has been on modeling lifecycle on the charts and how it has been affected by significant changes in the landscape of the music market. A key finding is that the market landscape has shifted and that lifecycle has shortened with lowered probabilities of surviving for each subsequent week on the chart. The significantly shorter shelf life of digital music calls for accelerated tactical and operational decision making on resource allocations, in particular marketing and promotional efforts that target potential winners. In the latter period (2000–2002), the likelihood of surviving another week falls below one-half by the fifth week while this does not occur until the eighth week in the earlier period (1995–1997). Hence, music companies may well opt to move promotional efforts earlier in the cycle. Interestingly, while the landscape has shifted, the underlying drivers that govern the lifecycle process appear to have remained steady. That is, even with significant "churn" in the music market and related environment between the 1995–97 and 2000–2002 periods, our underlying model form is robust and succinctly captures the lifecycle process for the entire duration. Thus, the same underlying decision models, where the parameters are constantly monitored and re-estimated, would provide a music firm with a reliable benchmark to gauge and assess

their suite of the music albums in the marketplace and make better decisions in an uncertain environment.

Our investigation reveals that it is critical to monitor and react to the early performance of music albums in the marketplace. Fig. 1 shows that albums that succeed in surviving past the initial weeks have a higher probability of further success on the charts. This is consistent with the rapid diffusion of information on music albums on the file sharing networks with initial positive feedback having swift impact. We now investigate whether information on an album's debut week on the chart and related user behavior on file sharing networks during the debut week might be useful for record companies in decisions relating to marketing or promoting their music products. That is, can music marketing decisions be improved with early monitoring of music album file sharing activities (say, using the first week's market and P2P sharing information)? This approach is motivated by Adler's (1985) concentration of success phenomenon, where we observe that the weekly ranking on the Billboard chart is highly correlated with the subsequent week's chart rank. Information diffusion through the P2P networks, in the form of sharing of a given album, captures actual consumer actions regarding the music album and may reveal a different characteristic of consumer attitude. Combining these two diverse sources of information would potentially enable us to efficiently monitor the performance and lifecycle of music albums on the charts. We detail our decision making approach below.

We consider two observable variables from an album's first week on the chart and estimate the probability of success in surviving past the initial two weeks. The first variable is the album's performance on the charts (*debut rank*) and the second variable of interest is the extent of online sharing of the album by users of a large P2P network (*shares*). The empirical model used is:

Album success $= b_0 + b_1$ debut rank $+ b_2$ shares.

The dependent variable is binary with a value of 0 if the album survives on the charts for 2 weeks or less, and 1 otherwise. The variable *debut rank* represents that the rank of the album on the first week the album debuts on the Billboard Top 100 chart. The variable *shares* is the average number of copies of the album available for downloading in the initial week when the album first debuts on the chart, and is collected from WinMX, the second most widely used P2P network, with a user base of over 1.5 million (Pastore, 2001; Graham, 2005a, b). Further details on this P2P file sharing measure are provided in Bhattacharjee et al. (2006a). It is possible that there are other potential variables that add to the explanatory power of the model. However, the focus is a given music company that needs to estimate an album's success probability with limited information available within a week of the album's debut and make a fast decision on subsequent marketing and promotional efforts for that album. Given the hedonic nature of the music album, consumer attitude towards the music product is only available after its consumption by consumers. As we detail below, our parsimonious empirical model is remarkably robust and has high explanatory power given only a week's information on an album's performance on the chart.

The data set utilized in the analysis is derived from a sample of 380 albums that debuted on the Billboard Top 100 Charts in 2003. (Recall that the stochastic model is estimated using data from 2002 and is used as a benchmark for initial guidance on the choice of the dependent variable—*album success*—cutoff value). Amongst these 380 albums from 2003, 218 albums survived more than 2 weeks while the remaining 162 albums survived no more than 2 weeks on the chart. This represents a baseline success rate of approximately 58%. The results of the logistic regression analysis are provided in Table 6. The overall measures of fit (Table 7) indicate a good fit for the overall model. The classification table (Table 8) that the predictive ability of the model to correctly identify successful albums is 82.9%, a significant enhancement over the baseline prediction of 58%.

An important finding is that user activity on file sharing, as measured by the variable *shares*, has a significant (negative) impact on an album's subsequent lifecycle. Sharing represents an important 'interactive human' activity on computer networks that can in fact be observed and monitored. Utilizing observable information from an album's market characteristics (*debut rank*) and human activity on computer networks (*shares*), we have shown that it is possible to accurately predict an album's

Table 6
Logistic regression parameter estimates

Parameter	Standard estimate	Wald χ^2	p Value
Intercept	2.8385	102.3403	<0.0001
Debut rank	0.0634	107.3405	<0.0001
Shares	0.00083	5.3655	<0.05

Table 7
Logistic regression fit values

Test	χ^2	p Value
Likelihood ratio	1901508	<0.0001
Score	164.3290	<0.0001
Wald	107.4926	<0.0001

Table 8
Logistic regression classification table

	Classified as successful	Classified as unsuccessful	Total
Successful albums	187	31	218
Unsuccessful albums	34	128	162

continued success on the charts. The insights from the stochastic model, coupled with the explanatory power of the parsimonious logistic regression model that utilizes the album's debut week information alone, helps to accurately predict and channel resources towards potential winners in the market. Our modeling and analysis represent important first steps to begin to establish the link between such individual level online activities to overall market performance.

6. Summary and conclusion

Technological advances can have significant impacts on economic markets. Here, we analyzed the lifecycle of albums on sales-based rankings, the Billboard Chart. Following earlier work related to the markets for artists, we proposed a stochastic process model. Brute-force estimation yielded excellent fits for all years. The nature of the estimated models indicates a shift after the technological innovations of MP3 and online file sharing that occurred over the 1998–1999 period. The 2000–2002 period is characterized by a much shorter lifecycle. Utilizing this initial information, we estimate their continued success on the charts by combining chart information and user activity data from P2P networks.

In our stochastic lifecycle analysis, we utilized the entire set of outcomes (Billboard Top 100 rankings) for the comparison periods: 1995–1997 and 2000–2002. We find that music as a digital good has been significantly impacted by market changes brought about by easier information dissemination and product access to potential consumers. While overall album survival has decreased in the 2000–2002 period, the chances of survival increase dramatically after an album has survived beyond the first week during this period. This indicates a pattern that the "good" albums survive more. This also suggests that the new environment brought on by technological and other market innovations is not conducive to lower quality music albums. Easier sampling and information dissemination hurts the lower quality albums. In general, the lifecycle of lower quality products will tend to diminish faster under this new environment. The analysis also suggests that albums face a shorter lifecycle overall. Does it indicate front-end sales to a larger degree? If so, what are the implications for copyright laws that have limits of 70 years and longer on information goods such as music, books, etc.?

Our conclusions also emphasize the robustness of our stochastic lifecycle model and consistency of results within each of the three-year periods, together with the differences between the two periods. Even with significant "churn" in the music market and related environment between the 1995–97 and 2000–2002 periods, our simple yet robust model effectively captures the stochastic component of the chart lifecycle process. Our parsimonious logistic regression model likewise has a high explanatory power while simply using two complementary elements of a music album's debut week information. This model approach is applicable to other stochastic processes that follow a similar pattern.

Acknowledgements

Appendix A. Steady-state characterization of the stochastic process

T_k, the total number of music albums that have appeared on the chart at the end of week k, can be expressed as

$$T_k = \sum_{m=1}^{k} C_{m,1}, \tag{A.1}$$

$TD_{k,w}$, the total number of albums that were on the chart for exactly w weeks before falling off the charts at the end of week k, can be expressed recursively as

$$TD_{k,w} = TD_{k-1,w} + (1 - p_w)C_{k,w}. \tag{A.2}$$

Note that $TD_{w,w-j} = 0 \; \forall \; j \geqslant 1$. Therefore (A.2) can be expressed as

$$TD_{k,w} = (1 - p_w)\sum_{m=1}^{k-w} C_{m+w,w} + TD_{w,w}. \tag{A.3}$$

Further, we have

$$TD_{w,w} = (1 - p_w)C_{1,1}\prod_{j=1}^{w-1} p_j, \tag{A.4}$$

$$C_{m+w,w} = C_{m+1,1}\prod_{j=1}^{w-1} p_j. \tag{A.5}$$

Using (A.4) and (A.5), (A.3) can be expressed as

$$TD_{k,w} = (1 - p_w)\prod_{j=1}^{w-1} p_j \sum_{m=1}^{k-w} C_{m+1,1} + (1 - p_w)\prod_{j=1}^{w-1} p_j C_{1,1}. \tag{A.6}$$

Simplifying, we obtain

$$TD_{k,w} = (1 - p_w)\prod_{j=1}^{w-1} p_j \sum_{m=1}^{k-w+1} C_{m,1}. \tag{A.7}$$

From (A.1) and (A.7), we have

$$\frac{TD_{k,w}}{T_k} = (1 - p_w)\prod_{j=1}^{w-1} p_j \left(\frac{\sum_{m=1}^{k-w+1} C_{m,1}}{\sum_{m=1}^{k} C_{m,1}} \right). \tag{A.8}$$

For finite values of w,

$$\lim_{k \to \infty} \left(\frac{TD_{k,w}}{T_k} \right) = (1 - p_w) \prod_{j=1}^{w-1} p_j, \qquad (A.9)$$

which is independent of k, thus yielding our steady state.

References

Adler, M., 1985. Stardom and talent. American Economic Review 75, 208 212.

Albert, S., 1998. Movie stars and the distribution of financially successful films in the motion picture industry. Journal of Cultural Economics 22, 249 270.

Associated Press, 2003. Norwegian hacker cracks iTunes code. CNN, November 27, 2003. ⟨http://www.cnn.com/2003/TECH/internet/11/27/itunes.code.ap/index.html⟩.

Bergstein, B., 2005. Copy protection still a work in progress. Associated Press, Novemeber 18, 2005. ⟨http://news.yahoo.com/s/ap/20051119/ap_on_hi_te/music_copy_protection;_ylt ArOQdPcARo.I0hnGv20mpcdk24cA;_ylu X3oDMTBidHQxYjh2BHNlYwN5bnN0b3J5⟩.

Bhattacharjee, S., Gopal, R.D., Lertwachara, K., Marsden, J.R., 2006a. Impact of legal threats on individual behavior: an analysis of music industry actions and online music sharing. The Journal of Law and Economics 49, 91 114.

Bhattacharjee, S., Gopal, R.D., Lertwachara, K., Marsden, J.R., 2006b. Whatever happened to payola? An empirical analysis of online music sharing. Decision Support Systems, in press.

Bradley, M., 2005. Sony aims at pirates and hits users. Christian science monitor, November 9, 2005. ⟨http://www.csmonitor.com/2005/1109/p14s01 stct.html⟩.

Bradlow, E.T., Fader, P.S., 2001. A bayesian lifetime model for the hot 100 billboard songs. The Journal of the American Statistical Association 96, 368 381.

Chung, K.H., Cox, R.A.K., 1994. A stochastic model of superstardom: an application of the Yule distribution. Review of Economics and Statistics 76, 771 775.

Clarke, G., 2005. DVD Jon hacks media player file encryption. The Register, September 2, 2005. ⟨http://www.theregister.co.uk/2005/09/02/dvd_jon_mediaplayer/⟩.

Cox, A.K., Chung, K.H., 1991. Patterns of research output and author concentration in the economics literature. Review of Economics and Statistics 73, 740 747.

Craver, S.A., Wu, M., Liu, B., Stubblefield, A., Swartzlander, B., Wallach, D.W., Dean, D., Felten, E.W., 2001. Reading between the lines: lessons from the SDMI challenge. In: Proceedings of the 10th USENIX Security Symposium, August 2001, Washington, DC.

Felten, E.W., 2003. A skeptical view of DRM and fair use. Communications of the ACM 46 (4), 56 61.

Felten, E.W., 2005. Inside risks: DRM and public policy. Communications of the ACM 48 (7), 112.

Goodley, S., 2003. Disharmony over music pirates on the Internet. The Telegraph, January 9, 2003. ⟨http://www.telegraph.co.uk⟩.

Graham, J., 2005a. Morpheus takes a stand again. USA TODAY, March 10, 2005. ⟨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: WinMX most popular music download spot, while iTunes and LimeWire share second position. The NPD Group, June 7, 2005. ⟨http://www.npd.com/dynamic/releases/press_050607.htm⟩.

Halderman, J.A., 2002. Evaluating new copy prevention techniques for audio CDs. In: Proceedings of the ACM Workshop on Digital Rights Management, November 2002, Washington, DC.

Levene, M., Fenner, T., Loizou, G., Wheeldon, R., 2002. A stochastic model for the evolution of the web. Computer Networks 39, 277 287.

Pastore, M., 2001. Net users finding P2P music alternatives. ClickZ Network, October 11, 2001. ⟨www.clickz.com/stats/sectors/software/article.php/901921⟩.

Patrizio, A., 1999. Why the DVD hack was a cinch. Wired News, November 2, 1999. ⟨http://www.wired.com/news/technology/0,1282,32263,00.html⟩.

Reuters, 2002. CD crack: magic marker indeed. Wired News, May 20, 2002. ⟨http://www.wired.com/news/technology/0,1282,52665,00.html⟩.

Reuters, 2005. Sony tests technology to limit CD burning. CNET News, 1 June 2005. ⟨http://news.cnet.co.uk/digitalmusic/0,39029666,39189658,00.htm⟩.

Rosen, S., 1981. The economics of superstars. American Economic Review 71, 845 858.

Russinovich, M., 2005. Sony, rootkits and digital rights management gone too far. Mark's Sysinternals blog, October 31, 2005. ⟨http://www.sysinternals.com/blog/2005/10/sony rootkits and digital rights.html⟩.

Schneier, B., 2000. Secrets and Lies: Digital Security in a Networked World. Wiley, New York.

Simon, H.A., 1955. On a class of skew distribution functions. Biometrika 42, 425 440.

Smith, E., 2005. Sony BMG faces civil complaint over CD software. The Wall Street Journal, November 22, 2005.

Yule, G.U., 1924. A mathematical theory of evolution, based on the conclusions of Dr. J.C. Willis, F.R.S. Philosophical Transactions of the Royal Society B. 213, 21 87.

Ziemann, G., 2002. RIAA's statistics don't add up to piracy. Azoz.com, December 11, 2002. ⟨http://www.azoz.com/music/features/0008.html⟩.