Information and the Skewness of Music Sales*

Ken Hendricks
University of Texas at Austin

Alan Sorensen
Stanford University & NBER

March 20, 2007

Abstract

This paper studies the role of consumer learning in the demand for recorded music by examining the impact of an artist's new album on sales of past and future albums. Using detailed album sales data for a sample of 355 artists, we show that the release of a new album increases sales of old albums, and the increase is substantial and permanent—especially if the new release is a hit. Various patterns in the data suggest the source of the spillover is information: a new release causes some uninformed consumers to discover the artist and purchase the artist's past albums. We develop and estimate a structural learning model that allows us to quantify the impact of consumers' lack of information on market outcomes. Our estimates imply that the distribution of sales is substantially more skewed than it would be if consumers were more fully informed, and that sales of non-debut albums are roughly 25% higher than they would be without the benefit of information generated by previous albums.

*We thank Dirk Bergmann, Greg Crawford, Steve Durlauf, Phillip Leslie, Marc Rysman, and Michael Whinston for suggestions that prompted changes in the paper, as well as Don Engel, Michael Lopez, and Ralph Peer for invaluable conversations that helped clarify several institutional details about the music industry. Chris Muratore and Nielsen SoundScan were very helpful in providing the data, and Natalie Chun and Abe Dunn provided outstanding research assistance. We are responsible for any errors or shortcomings that remain in spite of all the invaluable input.

1 Introduction

In entertainment industries such as books, music, and movies, commercial success tends to be highly concentrated. Many new products flow into these markets each week, but only a small fraction turn out to be profitable. Even among the profitable products, the distribution of returns is extremely skewed: a large share of total industry profit is claimed by a small number of very successful products. The skewness may simply be a reflection of the products' relative qualities. However, it may also reflect a lack of information: if consumers are unaware or poorly informed about most products, then market demand depends not only on their preferences but also on their knowledge of the product space and the process by which they obtain this knowledge. In entertainment industries, this process is driven in part by commercial success: consumers buy only the products they hear about, and they only hear about the products that other consumers buy. As a result, a product's success reinforces itself, causing the distribution of success across products to be more highly concentrated.

In this paper we study the process of product discovery in the market for recorded music. Our objective is to quantify how much of the skewness in the distribution of album sales can be attributed to consumers' lack of awareness of albums. We ask, what would the distribution of album sales look like if consumers were fully informed about all available albums? Our empirical strategy for addressing this issue is based on measuring the impact of new album releases on sales of previous albums by the same artist. The promotional activity and radio airplay associated with a newly released album enhances consumer awareness, and cause some consumers to purchase the artist's past albums (which are referred to in the industry as "catalog" albums). We call this effect the *backward spillover*. In order to measure it, we constructed a dataset consisting of weekly sales histories for a sample of 355 artists in the period 1993-2002. We observe sales separately for each of the artists' albums, and each artist in the sample released at least two albums (including a debut) during the sample period.

Figure 1 shows two clear examples of the backward spillover. The figure plots the logarithm of weekly national sales for the first and second albums of two popular recording artists, from the time of the artist's debut until six months after the artist's third release. The vertical lines in each graph indicate the release dates of the second and third albums. In the weeks surrounding these release dates, sales of catalog titles increased substantially. In the case of the "Bloodhound Gang," a relatively obscure alternative rock band, the second album was considerably more popular than the first, and its release catapulted sales of the prior album to levels even higher than it had attained at the time of its own release, with the effect persisting for at least three years. For the "Foo Fighters," a more popular hard rock band with a very successful debut album, the impact of the second release was somewhat less dramatic, but still generated an increase in sales of the band's first album. In both examples, the backward spillover is significantly positive for both the second and third album releases. The effects appear to begin in the weeks just prior to the new album's release, and they persists for many months.

The backward spillover could result from changes in consumers' information about catalog albums, but it could also result from increases in consumers' utility for those albums. Utility can increase if preferences depend on the artist's popularity and the new release raises the artist's popularity,[1] or if consumer preferences over an artist's albums are supermodular.[2] We document three patterns in the backward spillover that suggest changes in information (not preferences) are the primary source of the spillover. First, increased sales of catalog albums start to appear roughly four weeks *prior* to the release of a new album, which we argue most likely reflects consumers learning about artists from pre-release radio airplay and other promotional activity. Second, the spillovers are larger when the new release is a hit, and especially large when the new release is a hit and the catalog album was not, which again is highly suggestive of consumers discovering artists who were previously unknown. Third, we show that backward spillovers are smaller in an artist's home market (i.e., the city where the artist began her career, and where there is presumably a larger stock of informed consumers), even though sales are on average higher in the home market. As we argue in more detail below, these patterns cannot be easily rationalized by consumption complementarities or social effects, since both alternatives imply a selection effect: when an artist releases a new album, the potential demand for catalog albums comes from consumers who knowingly chose *not* to purchase those albums previously.

We develop and estimate a structural model in which the backward spillover is driven by changes in information. The probability that a consumer purchases an album is the product of two probabilities: the probability that she knows about the album and the probability that she likes the album. We assume that a new release has no effect on her preferences for the artist's catalog albums but it can increase the likelihood that she discovers the artist. We use variation in the magnitudes of the backward spillovers across artists to estimate the probability of artist discovery, and then use our estimates to quantify the extent to which skewness in the distribution of album sales reflects

[1] See Becker and Murphy [5]) and Brock and Durlauf [8]) for models with social effects in consumption.

[2] Becker, Grossman, and Murphy [4] and Gentzkow [14] are two interesting empirical studies of supermodular preferences.

consumers' lack of information (as opposed to album quality). Our estimates imply that while almost all consumers learn about an artist with a major hit, only 32% of consumers learn about an artist whose album achieves the median level of sales. We estimate that if consumers were fully informed about the albums, sales would have been substantially less skewed. For example, sales of the top artist in our sample would have exceeded the median artist's sales by a factor of 30 instead of the observed factor of 90.[3]

Because album releases generate information, they create a larger fan base for the artist's future albums, raising sales of these albums (relative to what they would have sold if the current album had never been released). Our structural estimates also allow us to quantify the importance of this effect, which we call the *forward spillover*. We calculate the counterfactual sales of second albums in the absence of any forward spillover—i.e., we use our model to predict the success of artists' second albums if instead those albums had been debuts. The difference between counterfactual sales and observed sales is large: collectively, the second albums in our sample sold 25% more than they would have if they hadn't been preceded by another album. This finding implies that contractual relationships between artists and record labels are complicated by a significant hold-up problem, and rationalizes the pervasive use of long-term contracts in the industry.

We are not aware of prior empirical literature on information spillovers between products,[4] but our theoretical framework is similar to prior work on brand extension. Choi [11], Cabral [10], and Wernerfelt [21] have developed theoretical models that study the impact of information spillovers on firms' decisions about whether to release new products under existing brand names. When consumers are uncertain about product qualities, the strong reputation of an existing product increases demand for new products sold under the same brand (the forward spillover), and the release of a high-quality new product can improve the brand image and boost sales of the existing product (the backward spillover).[5] There is a voluminous theoretical literature on the hold-up problem in contracts, but we are not aware of any that have studied the effect of backward spillovers.

The paper is organized as follows. Section 2 describes the data, which consist of weekly album sales histories for a sample of 355 artists. In Section 3 we use these data to measure the backward

[3]Our method does not require that we estimate consumer preferences for an album, and in particular whether these preferences depend upon the album's popularity. Since social effects are an alternative source of skew in the distribution of sales, and indeed may be very important in the market for music, we interpret our results as indicating the amount of *additional* skew that results from consumers' lack of information.

[4]Benkard's [6] study of learning by doing in aircraft production shows that learning spills over across aircraft types, but we have not seen any empirical papers that analyze information spillovers on the demand side of a market.

[5]In Cabral's paper, for example, the "feedback reputation effect" is exactly analogous to what we call the backward spillover.

spillovers, and show that variation in the magnitudes of the backward spillovers suggests consumer learning is the source. Section 4 presents the structural learning model, and describes the two counterfactual exercises aimed at revealing the quantitative impacts of consumer learning. Section 5 concludes.

2 Data

Our data describe the album sales histories of 355 music artists who were active between 1993 and 2002. Weekly sales data for each artist's albums were obtained from Nielsen SoundScan, a market research firm that tracks music sales at the point of sale, essentially by monitoring the cash registers at over 14,000 retail outlets. SoundScan is the principal source of sales data for the industry, and is the basis for the ubiquitous Billboard charts that track artist popularity. Various online databases were also consulted for auxiliary information (e.g., about genres and record labels) and to verify album release dates.

The sample was constructed by first identifying a set of candidate artists who released debut albums between 1993 and 2002, which is the period for which SoundScan data were available. Sampling randomly from the universe of such artists is infeasible, largely because it is difficult to find information on artists who were unsuccessful. Instead, we constructed our sample by looking for new artists appearing on Billboard charts. The majority of artists in our sample appeared on Billboard's "Heatseekers" chart, which lists the sales ranking of the top 25 new or ascendant artists each week.[6] A smaller number of artists were found because they appeared on regional "New Artists" charts, and an even smaller number were identified as new artists whose debut albums went straight to the Top 200 chart. This selection is obviously nonrandom: an artist must have enjoyed at least some small measure of success to be included in the sample. However, although the sample includes some artists whose first appearance on the Heatseekers list was followed by a rise to stardom, we note (and show in detail below) that it also includes many unknown artists whose success was modest and/or fleeting.[7]

Because our primary objective is to study demand responses to newly released albums, we restrict our attention to major studio releases. Singles, recordings of live performances, interviews, holiday

[6] Artists on the Heatseekers chart are "new" in the sense that they have never before appeared in the overall top 100 of Billboard's weekly sales chart—i.e., only artists who have never passed that threshold are eligible to be listed as Heatseekers.

[7] The weekly sales of the lowest-ranked artist on the Heatseekers chart is typically around 3,000, which is only a fraction of typical weekly sales for releases by famous artists who have graduated from the Heatseekers category.

albums, and anthologies or greatest hits albums are excluded from the analysis.[8] The resulting sets of albums were compared against online sources of artist discographies to verify that we had sales data for each artist's complete album history; we dropped any artists for whom albums were missing or for whom the sales data were incomplete.[9] Since timing of releases is an important part of our analysis, we also dropped a small number of artists with albums for which we could not reliably ascertain a release date.[10] Finally, we narrowed the sample to artists for whom we observe the first 52 weeks of sales for at least the first two albums; we then include an artist's third album in the analysis if we observe at least the first 52 weeks of sales for that album (i.e., we include third albums if they were released before 2002).

After applying all of these filters, the remaining sample contains 355 artists and 888 albums. The sample covers three broad genres of music: Rock (227 artists), Rap/R&B/Dance (79 artists), and Country/Blues (49 artists). The artists in the sample also cover a broad range of commercial success, from superstars to relative unknowns. Some of the most successful artists in the sample are Alanis Morissette, the Backstreet Boys, and Shania Twain; examples at the other extreme include Jupiter Coyote, The Weakerthans, and Melissa Ferrick.

Table 1 summarizes various important aspects of the data. The first panel shows the distribution of the albums' release dates separately by release number. The median debut date for artists in our sample is May 1996, with some releasing their first albums as early as 1993 and others as late as 2000. There are 178 artists in the sample for whom we observe three releases during the sample period, and 177 for whom we observe only 2 releases. Note that while we always observe at least two releases for each artist (due to the sample selection criteria), if we observe only two we do not know whether the artist's career died after the second release or if the third album was (or will be) released after the end of the sample period. In what follows we will discuss this right-truncation problem whenever it has a material impact on the analysis.

[8] Greatest hits albums could certainly affect sales of previous albums—repackaging old music would likely cannibalize sales of earlier albums—but we are primarily interested in the impact of *new* music on sales of old music. Moreover, there are very few artists in our sample that actually released greatest hits albums during the sample period, making it difficult to estimate their impact with any statistical precision.

[9] The most common causes for missing data were that a single SoundScan report was missing (e.g., the one containing the first few weeks of sales for the album) or that we pulled data for the re-release of an album but failed to obtain sales for the original release.

[10] For most albums, the release date listed by SoundScan is clearly correct; however, for some albums the listed date is inconsistent with the sales pattern (e.g., a large amount of sales reported before the listed release date). In the latter case, we consulted alternative sources to verify the release date that appeared to be correct based on the sales numbers. Whenever we could not confidently determine the release date of an album, we dropped it along with all other albums by the same artist.

The second panel of the table illustrates the considerable heterogeneity in sales across albums. Production, marketing, and distribution costs for a typical album are in the ballpark of $500,000, so an album must sell roughly 50,000 units (assuming a wholesale price of $10 per unit) in order to be barely profitable; over half of the albums in our sample passed that threshold in the first year. However, although most of the albums in the sample were nominally successful, the distribution of success is highly skewed: as the table illustrates, sales of the most popular albums are orders of magnitude higher than sales of the least popular ones. For debut albums, for example, first-year sales at the 90^{th} percentile are ten times sales at the median, and over 100 times sales at the 10^{th} percentile.

The skewness of returns is even greater across artists than across albums, since artist popularity tends to be somewhat persistent. An artist whose debut album is a hit is likely to also have a hit with her second album, so absolute differences in popularity among a cohort of artists are amplified over the course of their careers. Across the artists in our sample, the simple correlation between first-year sales of first and second releases is 0.52. For second and third releases the correlation is 0.77. Most of an artist's popularity appears to derive from artist-specific factors rather than album-specific factors, but the heterogeneity in success across albums by a given artist can still be substantial.

Another interesting feature of the sales distributions is how little they differ by release number. To the extent that an artist's popularity grows over time, one might expect later albums to be increasingly successful commercially. However, while this pattern holds on average for albums 1 through 3, even for artists who ultimately have very successful careers it is often the case that the most successful album was the first.

Most albums' sales paths exhibit an early peak followed by a steady, roughly exponential decline. As indicated in the third and fourth panels of table 1, sales typically peak in the very first week and are heavily front-loaded: a large fraction of the total sales occur in the first four weeks after release. Debut albums are an exception: first releases sometimes peak after several weeks, which presumably reflects a more gradual diffusion of information about albums by new artists. The degree to which sales are front-loaded increases with each successive release.

Seasonal variation in demand for music CDs is substantial. Overall, sales are strongest from late spring through early fall, and there is a dramatic spike in sales during mid- to late-December. Not surprisingly, album release dates exhibit some seasonality as well. Table 2 shows the distribution of releases across months. Late spring through early fall is the most popular time to release a new album, and record companies appear to avoid releasing new albums in December or January. Albums that would have been released in late November or December are presumably expedited in order to capture the holiday sales period.

The last panel of Table 1 summarizes the delay between album releases. The median elapsed time before the release of the second album is more than two years, and the low end of the distribution is still more than one year. Figure 2 shows a more complete picture of the heterogeneity in release lags for adjacent albums. Note that we can only compute time-to-next-release conditional on there being a next release. If an artist's second album was released near the end of the sample period, we only observe a third release if the time to release was short. However, Figure 2 shows that the distribution of elapsed time between albums 1 and 2 is clearly very similar to the distribution between albums 2 and 3, which suggests the right-truncation problem is not very severe for third albums.[11]

In addition to the obvious right truncation problem, our sample selection is likely to be biased toward artists whose success came early in their careers. For an artist to be selected into our sample, it must be the case that (a) the artist appeared on a Billboard chart between 1993-2002, and (b) we have data on all the artist's CD sales, which means the artist's first release must have come after January 1993. Taken together, these conditions imply that artists who hit a Billboard chart early in the sample period must have done so on their first or second album (otherwise we would have excluded them due to lack of data on their previous releases). Moreover, of the artists debuting late in our sample period, only the ones with early success will make it into our sample, because only they will have appeared on a Billboard chart. So the selection pushes toward artists who start strong. While this means our data will overstate the tendency of artists' successes to come early in their careers, we do not see any obvious biases the selection will induce in the empirical analyses below. Moreover, a quick check of some out-of-sample data suggests the selection bias is not very severe. We compiled a list of 927 artists who appeared on the Heatseekers chart between 1997-2002 but who are not included in our sample. Of these artists, 73% made it to the chart on their first or second album, as compared to 87% for the artists in our sample. The difference is qualitatively consistent with the selection problem described above, but we do not think the difference is quantitatively large enough to undermine our main results.

[11] In a previous version of this paper we included fourth albums in the analysis. The right-truncation problem is much more salient for fourth albums.

3 Measuring the Spillovers

Before introducing the structural model, we first want to measure the spillovers and determine their source. In this section we analyze the magnitudes of the spillovers (and how the magnitudes vary across artists) using an empirical approach taken from the literature on treatment effects.[12] Our method exploits exogenous variation in albums' release times: a new album release by an artist is interpreted as the "treatment," and sales of "treated" artists are compared to the sales of control artists who have not yet released a new album. We follow the impact of a new release on sales of catalog albums for 39 weeks (13 pre- and 26 post-treatment), and refer to this period as the treatment "window."

3.1 Regression Model

In presenting the model, we focus on the first treatment episode: the release of album 2 and its impact on sales of album 1. Let y_{it}^0 denote the log of album 1 sales of artist i in period t without treatment, and let y_{it}^s denote the log of album 1 sales in period t when artist i is in the s^{th} period of treatment. For each artist, t indexes time since the debut album's release, not calendar time. By taking logs, we are implicitly assuming that treatment effects are proportional, not additive. There are two reasons for adopting this specification. One is that the distribution of album sales is highly skewed. The other is that the average treatment effect is likely to be nonlinear: a new release has a larger impact on total sales of catalog titles for more popular artists. By measuring the treatment effect in proportional terms, we capture some of this nonlinearity. However, it could bias our estimates of the treatment effects upwards since proportionate effects are likely to be higher for less popular artists, and there are many more of them. Proportionate effects may also be higher for popular artists who are treated later since their sales levels are likely to be a lot lower than popular artists who are treated earlier. We address these issues in discussing the results below.

Our objective is to estimate the average treatment effect on the treated (ATE) for each period of the treatment window. The ATE is simply the difference $y_{it}^s - y_{it}^0$. The main challenge in estimating the ATE is that, in each period, we observe only one outcome for each artist. Our approach to measuring this difference is to use the sales of not-yet-treated albums (i.e., albums whose artists have not yet released a second album) as the control group against which to compare sales of treated albums (i.e., albums whose artists have recently released a second album). Essentially, this

[12] See Wooldridge [22] for a summary.

approach assumes that for an album whose artist issues a new release at t, counterfactual sales (i.e., what sales would have been in the absence of the new release) can be inferred from the sales of all other albums at t for which there has not yet been a new release.

Our specific sampling and estimation procedure is as follows. Albums are included in the sample only until the last period of the treatment window: observations on sales *after* that window are not used in estimating the regressions. We adopt this approach to ensure that, at any given t, treated albums are being compared with not-yet-treated albums, rather than a mix of not-yet-treated and previously-treated albums. Thus, the sample in period t includes artists that have not yet released a new album and artists who had a new release in periods $t-1, t-2, .., $ or $t-S+1$ but excludes artists whose new release occurred prior to period $t-S+1$. Basically, we want the control group to measure what happens to sales over time before any new albums are released.[13]

The regression model is as follows:

$$y_{it} = \alpha_0 + \alpha_i + \lambda_t + \sum_{m=2}^{12} \delta_m D_{it}^m + \sum_{s=-13}^{25} \beta_s I_{it}^s + \epsilon_{it}, \tag{1}$$

where α_i is an artist fixed effect, the λ_t's are time dummies, and the D^m's are month-of-year dummies (to control for seasonality).[14] Here I_{it}^s is an indicator equal to one if the release of artist i's new album was s weeks away from period t, so β_s measures the new album's sales impact in week s of the treatment window. ($t = 0$ corresponds to the first week following the new release.) Intuitively, after accounting for time and artist fixed effects, we compute the difference in the average sales of album 1 between artists in treatment period s and artists who are not treated for each period, and then average these differences across the time periods. The stochastic error, ϵ_{it}, is assumed to be heteroskedastic across i (some artists' sales are more volatile than others') and autocorrelated within i (random shocks to an artist's sales are persistent over time). The time dummies (λ_t) allow for a flexible decay path of sales, but implicitly we are assuming that the shape of this decay path is the same across albums. Although differences in the level of demand are captured by the album fixed effects, differences in the shapes of albums' sales paths are necessarily part of the error (ϵ).

Including separate indicators for successive weeks of treatment allows us to check whether the

[13]We believe dropping post-treatment observations is the most appropriate approach, but it turns out not to matter very much: our estimates change very little if we include these observations.

[14]The results reported below are essentially unchanged if we control for seasonality with week-of-year dummies instead of month-of-year dummies.

new release's impact diminishes (or even reverses) over time, which is important for determining whether the effects reflect intertemporal demand shifts. We allow for a 39-week treatment window, beginning 13 weeks (3 months) *before* the release of the new album. The pre-release periods are included for two reasons. First, much of the promotional activity surrounding the release of a new album occurs in the weeks leading up to the release, and we want to allow for the possibility that the backward spillover reflects consumers' responses to these pre-release marketing campaigns. In some cases labels release singles from the new album in advance of the album itself, so that pre-release effects could also reflect advance airplay of the album's songs.[15] Second, including pre-release dummies serves as a reality check: we consider it rather implausible that a new album could have an impact on prior albums' sales many months in advance of its actual release, so if the estimated effects of the pre-release dummies are statistical zeros for months far enough back, we can interpret this as an indirect validation of our empirical model.

For the regression described above to yield consistent estimates of the treatment effect, the critical assumption is that the treatment indicators in a period are independent of the idiosyncratic sales shocks in that period. In other words, after controlling for time-invariant characteristics such as genre and artist quality that affect the level of sales in each period, we need the treatment (i.e., the release of a new album) to be random across artists. This is a strong but not implausible assumption. We suspect that the main factor determining the time between releases is the creative process, which is arguably exogenous to time-varying factors. Developing new music requires ideas, coordination, and effort, all of which are subject to the vagaries of the artist's moods and incentives. Nevertheless, the specific question for our analysis is whether release times depend on the sales patterns of previous albums in ways that album fixed effects cannot control.

One possibility is that release times are related to the *shape* of the previous album's sales path. For example, albums of artists that spend relatively more effort promoting the current album in live tours and other engagements will tend to have "longer legs" (i.e., slower decline rates) and later release times than albums of artists that spend more time working on the new album. To check this, we estimated Cox proportional hazard models with time-to-release as the dependent variable, and various album and artist characteristics included as covariates. Somewhat surprisingly, the time it takes to release an artist's new album is essentially independent of the success of the prior album

[15] One might wonder whether the relevant event is the release of the single or the release of the album. Although we have data on when singles were released for *sale*, this does not correspond reliably with the timing of the release on the radio. Radio stations are given advance copies of albums to be played on the air, and a given single may be played on the radio long before it is released for sale in stores. Moreover, even when a single has been released in advance of the album, the label's promotional activity is still focused around the release date of the album.

(as measured by first six months' sales) and of its decline rate, after conditioning on genre.[16] These results seem to validate our assumption that release times are exogenous—at least with respect to the level and rate of change in the prior album's sales. However, subtle relationships between sales-path shapes and release times may still exist. If so, the potential problem is that our regression only controls for the average rate of decline in album sales, so our estimates of the treatment effect will be biased if deviations from that average are systematically related to release times.

In order to address this issue, we can estimate the regression model of equation (1) using the first difference of *ln(sales)* as the dependent variable: i.e., we estimate

$$\Delta y_{it} = \widetilde{\alpha}_0 + \widetilde{\alpha}_i + \widetilde{\lambda}_t + \sum_{m=2}^{12} \widetilde{\delta}_m D_{it}^m + \sum_{s=-13}^{25} \widetilde{\beta}_s I_{it}^s + \widetilde{\epsilon}_{it} \ , \tag{2}$$

where $\Delta y_{it} \equiv y_{it} - y_{it-1}$. This model estimates the impact of new releases on the percentage rate of *change* (from week to week) in previous albums' sales. The advantage of this specification is that heterogeneity in sales levels is still accounted for (the first differencing sweeps it out), and the fixed effects, $\widetilde{\alpha}_i$, now control for unobserved heterogeneity in albums' decline rates. Taking this heterogeneity out of the error term mitigates concerns about the endogeneity of treatment with respect to the shape of an album's sales path.

3.2 Spillover Estimates

We estimate the regressions in (1) and (2) separately for each of two treatments: the impact of the second and third releases on sales of the previous album.[17] In constructing the samples for estimating the regression we impose several restrictions. First, we exclude the first eight months of albums' sales histories, in order to avoid having to model heterogeneity in early time paths. Recall that although most albums peak very early and then decline monotonically, for some "sleeper" albums we do observe accelerating sales over the first few months. By starting our sample at eight months, we ensure that the vast majority of albums have already reached their sales peaks, so that the λ_t's have a better chance at controlling for the decay dynamics. A second restriction involves truncating the other end of the sales histories: we exclude sales occurring more than four years beyond the relevant starting point. This means that if an artist's second album was released more

[16] A table showing the detailed results of this exercise is included in a previous version of this paper [16].

[17] Here we report results only for adjacent album pairs, but we have also measured the impact for non-adjacent pairs (e.g., the impact of album 3's release on sales of album 1). The effects for non-adjacent pairs are positive, statistically significant, and persistent, but slightly smaller than for adjacent album pairs.

than four years after the first, then that artist is not included in the estimation of the impact of second releases on first albums, and (similarly) if an artist's third release came more than four years after the second, then that artist is excluded from the regressions estimating the impact of album 3 on album 2.

Table 3 presents estimates of the regressions (1) and (2), with standard errors corrected for heteroskedasticity across artists and serial correlation within artists. (Estimated AR(1) coefficients are listed at the bottom of the table.) The columns of the table represent different treatment episodes (album pairs), and the rows of the table list the estimated effects for the 39 weeks of the treatment window (i.e., the $\hat{\beta}_s$'s). Since the dependent variable is the logarithm of sales, the coefficients for specification (1) can be interpreted as approximate percentage changes in sales resulting from the new release, and for specification (2) they represent effects on the percentage rate of change in sales from week to week. The number of coefficients listed in Table 3 makes it somewhat difficult to read, so we summarize the results graphically in Figures 3 and 4. Figure 3 shows the estimated effects from specification (1), along with 95% confidence bands, for each of the album pairs. As can be seen in the figure, the estimates of the effects for each of the weeks following the release of a new album are always positive, substantive, and statistically significant. The largest spillover is between albums 2 and 1, with estimates ranging between 40-55%. The spillover of album 3 onto album 1 is smaller, with estimates ranging roughly between 15-35%. Figure 4 shows a comparison of the results from the two specifications. The solid line plots the cumulative impact implied by the estimated weekly coefficients from the first-differenced model (2), and the dashed line indicates the estimated effects from the levels regression (1). The implied effects are qualitatively and quantitatively very similar, which we interpret as reassuring evidence that our results are driven by real effects, not by subtle correlations between current sales flows and the timing of new releases.[18]

In each treatment episode, the estimated impact of the new album three months prior to its actual release is statistically indistinguishable from zero. As discussed above, this provides some reassurance about the model's assumptions: three months prior to the treatment, the sales of soon-to-be-treated albums are statistically indistinguishable from control albums (after conditioning on album fixed effects and seasonal effects). In general, small (but statistically significant) increases start showing up 4-8 weeks prior to the new album's release, growing in magnitude until the week

[18]We also checked the robustness of the estimates by splitting the sample in each treatment based on the median treatment time. As expected, the patterns are the same but the estimated effects are smaller for the albums that are treated early and larger for albums treated later. (This pattern makes sense because our model assumes the effects are proportional: albums treated later will tend to have lower sales flows at the time of treatment, so the proportional impact of the new release will tend to be larger than for albums with high sales flows.) The estimates are always strongly significant.

of the release ($t = 0$ in the table), at which point there is a substantial spike upward in sales.

The estimated effects are remarkably persistent: especially for the impact of album 2 on album 1, the spillovers do not appear to be transitory. It is important to note, however, that the increasing coefficients in some specifications do not imply ever-increasing sales paths, since the treatment effects in general do not dominate the underlying decay trend in sales. (In order to save space, the table does not list the estimated time dummies, which reveal a steady and almost perfectly monotonic decline over time.)

The main conclusion that we draw from the above results is that the backward spillover is on average significantly positive and permanent. There is no evidence of new albums shifting demand for catalog albums from future periods towards the release period. If the spillover represents consumers who would have eventually purchased the catalog title anyway (i.e., even if the new album were never released), then the changes in sales would be transitory and eventually would become negative. We have tried longer treatment windows. In some cases, the treatment effect does die out eventually but in none of the cases does the treatment effect turn negative. Thus, the release of a new album generates permanent increases in demand for past albums, inducing purchases by customers who would not have otherwise purchased.

3.3 Spillover Variation

Although it is clear from our results that backward spillovers are significant, it is less clear why the spillovers occur. In this subsection we analyze variation in the magnitudes of the spillovers as a means of understanding their source. First, we split our sample based on whether the albums were "hits," and examine how the backward spillover depends on the relative success of the new album *vis a vis* the catalog album. We define a hit as an album that sold 250,000 units or more in its first year; 30% of the albums in our sample meet this criterion.[19] We then divide our sample into four categories—hits followed by hits, hits followed by non-hits, non-hits followed by hits, and non-hits followed by non-hits—and summarize the backward spillovers for each of the four categories in Table 4.

The table is based on estimates of the regression model computed separately for each subgroup.[20]

[19]As a point of reference, the RIAA certifies albums as "Gold" if they sell more than 500,000 units. Also, among the albums we categorize as hits, at least 90% had peak sales high enough to appear on Billboard's Top 200 chart (vs. less than 10% among those we categorize as non-hits).

[20]We use the first-differences model in equation (2). Some of the estimated sales increases are smaller if we estimate the model in levels, but the qualitative patterns are essentially the same.

These are then used to calculate the implied total change in sales for the "median" album. Specifically, we calculate the median weekly sales 14 weeks prior to the median release time, and the median weekly decline over the 39 weeks that follow. (In these calculations, we use only albums whose artists have not yet released the next album, so that the median sales flows and median decline rates will not reflect any of the backward spillovers.) For example, in the group of 53 artists whose first two albums were both hits, the median time between the first and second releases is 108 weeks. Among first albums for which there was not yet a second release, the median weekly sales at week 94 (=108-14) was 1,888, and the median decline rate over weeks 95-134 was 2.1% per week. So we take a hypothetical album, with weekly sales beginning at 1,888 and declining at 2.1% per week, and apply the percentage increases implied by our estimated coefficients. The predicted total increase in sales over the 39-week period is 22,161, or roughly $350,000 in additional revenues (using a retail price of $16 per unit).

The patterns in Table 4 establish that the backward spillover is always larger when the new album is a hit, whether the previous album was a non-hit or a hit. The largest percentage increase occurs when a non-hit album is followed by a hit: for an artist whose second album was her first hit, we estimate that weekly sales of her first album more than double when the new album is released. The smallest increase occurs when a hit is followed by a non-hit. The same patterns hold when we examine the impact of the third release on the sales of album 2. The spillovers are large when the new album is a hit, but negligible otherwise. The numbers are slightly smaller than those for the previous album. An important lesson from Table 4 is that although on average (across all types of albums) the backward spillovers are of modest economic significance, they are in fact quite large for the artists that matter: those who have hits or have the potential to produce hits.

In addition to splitting our sample to compare national sales across artists, we can also split the sample geographically to compare sales across markets for a given artist. An especially informative comparison is between an artist's home market (i.e., the city where the artist's career began) and other markets. Because new artists tend to have geographically limited concert tours—in many cases performing only in local clubs—artists in their early careers are more popular in their home markets.

We were able to determine the city of origin for 325 of the 339 artists included in the regression analysis of Table 3; 268 of these artists originated in the U.S., so we can observe sales in the home market and compare them to sales in other markets across the nation. SoundScan reports album sales separately for 100 Designated Market Areas (DMAs), each one corresponding to a major metropolitan area such as Los Angeles or Boston. We determined each artist's city of origin, and labeled the nearest DMA to be the artist's home market.[21] It is easy to verify that artists are indeed more popular in their home markets: over 80% of debut albums had disproportionately high sales in the artist's home market, meaning that the home market's share of national first-year sales was higher than the typical share for other artists of the same genre. On average, the home market's share of national sales was 8 percentage points larger than would have been predicted based on that market's share of overall sales within the artist's genre.

Are backward spillovers smaller in artists' home markets? Using the market-level data, we estimate a variant of the regression model in (1):

$$y_{imt} = \alpha_0 + \alpha_i + \sum_{g=1}^{4} \theta_{gm} G_i^g + \lambda_1 t + \lambda_2 t^2 + \psi H_{im} + \sum_{k=2}^{12} \delta_k D_{it}^k + \sum_{s=-13}^{26} I_{it}^s (\beta_s + \gamma H_{im}) + \epsilon_{imt} \quad (3)$$

where y_{imt} is log sales of artist i's album in market m in week t; G_i^g is a dummy equal to one if artist i is in genre g (so the θ_{gm}'s are market×genre fixed effects); the D_{it}^k's are month-of-year dummies, the I_{it}^s's are the treatment dummies, and H_{im} equals one if market m is artist i's home market. The key differences between this model and the one described in equation (1) are that (i) we use market-level sales data, and control for heterogeneity in sales across markets using market×genre fixed effects;[22] (ii) we measure whether sales are on average higher in the artist's home market (i.e., the parameter ψ); and (iii) we allow the spillover effects to differ for home markets vs. other markets (via the parameter γ).

Table 5 reports the results. The estimates of ψ confirm that on average sales are much higher in an artist's home market than in other markets. For the debut album, the coefficient of 0.814 implies that sales are over twice as high in the home market than in other markets, other things being equal. Notably, the home market advantage is smaller for later albums. Also, in spite of the fact that artists' albums are on average more successful in their home markets, the backward spillovers are on average *smaller* in home markets. The estimates of γ are similar across the album pairs,

[21] Roughly 20% of the artists are solo artists, and for these we were only able to find the city of birth—which is not necessarily the city in which the artist first began performing. However, it is plausible that solo artists are more well-known in their birth cities than in other cities nationwide, even if they began their performing careers elsewhere. In any case, all of our analyses deliver the same conclusions if we exclude solo artists.

[22] Note that we can alternatively include market×artist fixed effects. Doing so means we cannot estimate ψ, the coefficient on H_{im}, because H_{im} is collinear with the market×artist effect for the home market. Adopting this specification yields results for all the other parameters that are virtually identical to those we report for the model with market×genre effects.

indicating that backward spillovers are 10-14 percentage points smaller in an artist's home market than in other markets.

3.4 Interpretation

The preceding analysis establishes several stylized facts about the backward spillover: (1) it starts to appear several weeks prior to the release of the new album and increases throughout the pre-release period; (2) it peaks in the week of the release and thereafter remains roughly constant as a percentage of sales; (3) it is large and economically significant when the new release is a hit; (4) it is large when the catalog album was a hit but especially large (in percentage terms) when the catalog album was not a hit; (5) it is smaller in an artist's home market, even though sales are on average substantially higher in the home market.

These facts are easily explained if the backward spillover results from consumers discovering artists. The pre-release effects most likely reflect promotional activity and radio airplay that occurs prior to the new album's release in stores. As a simple check, we estimated the regressions separately for artists who released singles in advance of the new album vs. artists who released singles after the release of the album.[23] The pre-release effects on the catalog album's sales are much larger for the artists with advance singles: artists with pre-album singles had roughly 30-40% sales increases of album 1 in the three weeks prior to the release of album 2, while artists with post-album singles had increases of 5-20%. This pattern suggests the pre-release sales increases come primarily from consumers who discover the artist as a result of pre-release promotion and airplay.

The post-release pattern can be explained by an arrival rate of new consumers that is proportional to the stock of uninformed consumers, which falls over the release period. The magnitude of the spillover then depends on the capacity of the album to draw new consumers to the artist's market, which in turn depends on how well the new release does and on the number of uninformed consumers. The spillover is larger when the new release is a hit because uninformed consumers are more likely to discover the artist and, upon discovery, purchase the artist's albums. The variation in the spillover of a hit release with the success of the catalog album reflects the interaction of two effects: the number of uninformed consumers is lower when the catalog was a hit, but the probability of purchase conditional on artist discovery is higher.

[23]Of the album releases represented in our sample, 23% were preceded by the release of a single, while another 16% had single releases occurring after the album release.

The fact that spillovers are smaller in artists' home markets can be attributed to home markets having smaller stocks of uninformed consumers. Potential buyers in the home market are already familiar with the artist, so the new album generates fewer additional sales of the catalog album. Also, the fact that the home market advantage is smaller for later albums is consistent with the notion that awareness of the artist becomes less geographically concentrated as the the artist's career progresses.

Although the data strongly suggest that the backward spillovers are an information phenomenon, it is of course important to consider alternative explanations. One possibility is that the spillovers reflect responses to price changes. Although we cannot rule this out directly (because we do not have price data for the albums in our sample), it is clear that prices cannot explain the spillovers. Variation in price across titles and over time is very limited, and although discounts are occasionally "pushed down" to the retail level by distributors, these discounts are usually for new albums rather than catalog titles. According to two retail store managers with whom we had conversations, even when catalog albums are discounted, the timing of the sales is not systematically related to new releases by the same artist.[24] Moreover, even if there were meaningful variation in prices, price effects would still have difficulty explaining the results in Tables 4 and 5 regarding hits and home markets.

A more plausible alternative explanation of the backward spillovers is that they result from complementarities in consumption. If consumers have supermodular preferences over albums by the same artist, then a new release increases the utility of the artist's catalog albums. Consumers who were previously not willing to buy a catalog album may do so when they can consume it together with the new album. [25] The pre-release effects would have to be interpreted as purchases by consumers who anticipate buying the new album when it is released. That is, even though the benefits of joint consumption cannot be obtained until both albums are available, the consumer decides to buy the catalog album immediately to obtain the additional benefits of consuming the catalog album before the new release.

[24]In a previous version we reported price data for a sample of CDs offered at a major online retailer. Comparing prices for three groups of albums—new releases, catalog titles by artists with new releases, and catalog titles by artists without new releases—we found that although new releases tended to be discounted, the price distributions for the other two groups were indistinguishable. Catalog titles by artists who recently released a new album were no more likely to be discounted than other catalog titles.

[25]The complementarity could be interpreted as a characterization of fans: e.g., when consumers listen regularly to an artist's music, they become accustomed to it or invested in the image associated with it, and therefore more likely to purchase more music from that artist. Such complementarities would be similar to those modeled by Becker, Grossman, and Murphy [4] to describe cigarette addiction, and by Gentzkow [14] to describe consumption of online and print editions of a newspaper.

Another possible source of the spillovers is social effects: a consumer's utility from an album may depend on the number of other consumers purchasing the artist's albums.[26] A new release that sells well could increase utility (and hence sales) of catalog albums if (a) the social effects operate to some extent at the artist level, and (b) the social effects associated with the new album exceed the social effects generated by the catalog albums themselves when they were released. Under this explanation, the pre-release effects reflect consumers anticipating the new release will be a hit.

Like the learning model, consumption complementarities or social effects would predict persistent backward spillovers, since in both cases the new release directly changes the probability of wanting to purchase the catalog album. However, the feature that distinguishes these models from consumer learning is a selection effect. In the consumption complementarity or social effects models, the potential demand for catalog albums when the artist releases a new album comes from consumers who chose *not* to purchase the catalog albums previously. Consequently, the probability that these consumers will purchase the catalog album following a new release is significantly lower than it would have been for a randomly selected consumer. The selection effect makes it difficult for these models to explain the patterns described in Table 4. In the case of supermodular preferences, the very large spillovers on catalog albums that did poorly when they were released can only be explained by a distribution of preferences in which low levels of sales are associated with a high density of consumers near the margin. As long as album purchases are tail events, almost any ordinary distribution (e.g., exponential, normal, lognormal) would imply the opposite: the density of consumers near the margin is high when sales are high. A similar issue arises when trying to explain why the backward spillover is smaller in an artist's home market (where the artist's sales are otherwise higher). Also, the fact that artists tend to be more successful in their home markets would need to be explained. It could reflect an artist selection effect, in the sense that artists may choose their home markets based on their musical styles (or choose their musical styles based on their home markets), but the magnitude seems implausibly large.

For a social effects model, the main difficulty is explaining why spillovers are still quite large when the catalog album was more successful than the new release. If the catalog album was more popular than the new release, then the social effects should have been even stronger for that album. Any consumers who were not induced to buy the catalog album at the time of its release will not be induced to do so at the time of the new release either. As a simple check, we estimated the backward spillovers for artists with two consecutive hits, but for whom the second album was less popular than the first. The spillovers are no smaller for this group than for those whose second hit

[26]See Becker and Murphy [5] and Brock and Durlauf [8] for insightful overviews of models with social effects.

was more popular than the first.

We also checked whether the spillovers vary by genre. Intuitively, one might expect social effects to be strongest for music aimed at teenagers and young adults. We estimated the spillovers separately for "teen-oriented" genres (such as "Indie Rock" and "Teen Pop"), and found no evidence that the spillovers are larger for these genres than for other genres.

4 A Learning Model of Spillovers

Based on the empirical results described above, we conclude that the principal source of the backward spillovers is learning by previously uninformed consumers. One broad implication of our findings is that the distribution of sales in music is significantly more concentrated than it would be in the but-for world of complete information. In other words, the high concentration of returns across artists partly reflects the way in which consumers learn about albums and rtists, rather than their actual preferences. In this section, we develop a structural model of spillovers based on artist discovery that allows us to quantify how much of the skewness in sales is due to a lack of information. In particular, we calculate counterfactual sales under the assumption that all consumers are fully informed about all albums, and compare these to actual sales.

A second implication of consumer learning is that non-debut albums benefit from a forward spillover: notoriety generated by the artist's previous albums carries over to the new release. We measure the importance of these forward spillovers by calculating counterfactual sales of album 2 under the assumption that there was no previous album. In other words, we use our structural model to predict the success of artists' second albums if instead those albums had been debuts.

4.1 Model

We consider a simple model in which the artist releases three albums, $k = 1, 2, 3$, sequentially in periods $t = 1, 2, 3$. Unlike in the previous section, where we analyzed weekly sales dynamics, in this model we are interested in cumulative sales. We define a period to be the time between releases of two albums, and refer to it as the "release period" of the new album. The probability that a consumer buys an album in its release period is the product of two probabilities: the probability that he likes the album conditional on knowing about the album, and the probability that he knows about the album. The latter probability is the key unobservable in our model: if known, then album sales in the counterfactual world of complete information are simply equal to observed sales divided by this probability. In essence, our approach is to specify a functional form for this probability and estimate it from the impact of new releases on catalog sales under the assumption that album preferences do not change.

Survey evidence indicates that consumers learn about albums and artists primarily by hearing their albums on the radio or by seeing music videos on television.[27] We assume that if a consumer learns about an artist, then she learns about all of the artist's albums and her utility for those albums. Thus, our model of album learning is really a model of artist discovery. Conditional on consumer i not knowing about the artist at the beginning of period t, let I_{it} denote the binary random variable that is equal to 1 if consumer i discovers the artist during period t, and zero otherwise. We define the "discovery probability" to be

$$\Pr\{I_{it} = 1\} = \frac{ae^{bE(S_{tt})}}{(1-a) + ae^{bE(S_{tt})}}$$

where a and b are positive parameters and $E(S_{tt})$ is expected sales of the new release in period t. If the new album is expected to sell very little, then playing time is essentially zero and the probability that consumer i learns about the artist is a; as expected sales get very large, the probability converges to 1. This functional form is motivated by the fact that consumers learn about artists primarily through radio airplay, which we assume is proportional to expected sales of the new release.[28]

Appealing to the law of large numbers, the fraction of consumers who know about the artist at the end of release period t is given by

$$q_t = q_{t-1} + (1-q_t)\frac{ae^{bE(S_{tt})}}{(1-a) + ae^{bE(S_{tt})}}. \tag{4}$$

where q_{t-1} is the fraction of consumers who know about the artist at the beginning of period t. We assume that once consumers discover an artist, they do not forget—so the stock of informed

[27] In one national survey of music consumers conducted in 1994 [19] consumers were asked what motivated their recent music purchases, and the most common response was having heard the music on the radio. A more recent survey in 2006 [13] produced a similar finding: 55% of consumers said they learn about new music primarily from FM radio.

[28] Most radio airplay is devoted to songs from an artist's newest album. For example, we checked the Billboard "Hot 100 Airplay" chart for July 17, 1999, and found that 74 of the 75 listed songs were from the respective artist's newest album. However, even if the release of a new album leads to increased promotion or airplay of the artist's old albums, this is consistent with our assumption as long as the increase is proportional to expected sales of the *new* album.

consumers is strictly increasing over release periods. We shall also assume that radio stations have perfect foresight, so expected sales are equal to actual sales: $E(S_{tt}) = S_{tt}$.

If a consumer discovers the artist during release period t, the probability she likes catalog album k enough to buy it is denoted by p_{kt}. Sales of catalog album k in release period t ($t > k$) can then be written as

$$S_{kt} \simeq p_{kt}[q_t(S_{tt}) - q_{t-1}(S_{t-1,t-1})]N. \tag{5}$$

where N is the number of potential consumers. The term in parentheses is the fraction of new consumers who discover the artist in period t, and we use the law of large numbers to argue that the fraction of these consumers who buy album k is approximately equal to p_{kt}. Thus, the backward spillover is driven by these new consumers who discover the artist as a result of the promotion of the new release. Sales of the new release in period t are given by

$$S_{tt} \simeq p_{tt} q_t(S_{tt})N \tag{6}$$

where p_{tt} is the probability that a consumer who knows about the artist buys the new release in period t. Note that sales of the new release are reinforcing: more sales increases the fraction of informed consumers, which further increases sales. It is straightforward to show that a solution to (6) always exists, that there is either one or three solutions (generically), and the minimum and maximium solutions are increasing in album quality (i.e., p_{tt}). Here multiple equilibria can arise because of the logistic learning curve and the lack of coordination among radio stations in choosing playing time.

If artist discovery is random and album preferences do not change, then the probability that a consumer likes an album does not depend upon when she discovers the artist. In other words, the fraction of newly informed consumers who buy catalog album k in a later release period is approximately the same as the fraction of informed consumers who bought the album when it was released. We assume this is true except possibly for a decline in mean utility over time that is common across albums. Specifically, we write the preference probability as

$$p_{kt} = p_{kk} e^{-\gamma T_{kt}} \ ,$$

where T_{kt} is the length of time between the release of album k and the release of album t. This specification allows consumers to have a taste for "newness," and allows the spillover to decline as a function of time between releases.

Because N is very large, we treat equations (5) and (6) as equalities. Substituting for p_{kt} and q_t and taking logs, we obtain a pair of estimating equations:[29]

$$\log\frac{S_{12}}{S_{11}} = \log\left(\frac{(1-q_0)(1-a)}{q_0(1-a)+ae^{bS_{11}}}\right) - \gamma T_{12} + \log(a) + bS_{22} - \log\left(1-a+ae^{bS_{22}}\right) + \eta_1 \quad (7)$$

$$\log\frac{S_{23}}{S_{22}} = \log\left(\frac{(1-q_0)(1-a)^2}{q_0(1-a)+ae^{bS_{11}}+ae^{bS_{22}}(1-a+ae^{bS_{11}})}\right) - \gamma T_{23} + \log(a) + bS_{33} - \log\left(1-a+ae^{bS_{33}}\right) + \eta_2 \quad (8)$$

Note that the market size, N, cancels. Here q_0 is the baseline awareness of an artist (which for the present exercise we assume to be constant across artists), and we include the η's as artist-specific error terms.

In taking this simple model to the data, it will be convenient to standardize the length of the period over which to measure albums' sales. In the results reported below, we calculate sales over a one-year period: S_{kk} is measured as first-year sales of album k, and S_{kt} is measured as cumulative sales of album k during the first year of release t. The definition of a release period as one year is long enough for the sales dynamics to have run their course: almost anyone who was going to learn about the new release and buy it before the release of the next album will have done so within the first year. It introduces some measurement error into the model, since the fraction of informed consumers at the end of an album's first year is not the same as the fraction of informed consumers at the time of the next release. However, the error is small: on average, first year sales represent 85% of cumulative sales at the time of the next release.[30] Time-between-releases (T_{kt}) is measured from end of the first year of album k's release to the beginning of the first year of album t's release.

[29] In what follows, we focus on spillovers between adjacent album releases, k and $k-1$, where $k = 2, 3$. However, it is also possible to use the spillover of album 3 onto album 1.

[30] An alternative approach would be to let the release periods be artist-specific, measuring S_{kt} as cumulative sales of album k all the way up to the time of album t's release. However, it turns out not to make a meaningful difference, precisely because album sales after the first year are so low.

Given the above definition of release periods, the estimating equations generated by our model are essentially nonlinear regressions of spillover sales (S_{kt}) on first-year sales of the catalog album (S_{kk}) and first-year sales of the new release (S_{tt}). Straightforward estimation methods will yield unbiased estimates of the learning parameters as long as the error terms (η), which represent approximation errors and unobserved shocks to spillover sales, are orthogonal to S_{kk} and S_{tt}. This condition is almost surely true for S_{kk}, since the catalog album's first-year sales are predetermined at the time of the new album's release, and typically the time between release periods is several months. Indeed, by the time the new album is released, sales of the catalog album are typically flat. Promotional activities that occur after the new release (radio airplay, television appearances, concert tours, etc.) will tend to increase sales of both the new and catalog albums. As long as these activities are a direct result of the new album release, they represent exactly the effects S_{tt} is supposed to capture as a proxy. On the other hand, promotional activities that would have occurred irrespective of the new release would be problematic from an econometric standpoint, since they could generate a spurious positive correlation between S_{kt} and S_{tt}. We suspect this issue is unimportant, however, since airplay and other promotions almost always focus on the artist's new release.[31]

In summary, our model is based on two key assumptions. First, learning is driven by promotion and airplay of the *new* album only. This assumption makes the system recursive, decoupling the equation for spillover sales from the equation for the sales of the new album. We can then estimate the learning parameters a and b from the spillovers (equations (7) and (8)). Second, consumer preferences for a catalog album are constant (except possibly for age effects). In particular, it cannot depend upon the popularity of the new album.

In the appendix we consider an alternative model in which we essentially reverse these assumptions: we treat the discovery probability as fixed, but allow preferences for the catalog album to be a function of the new album's success, so that the spillovers are generated by social effects. Unlike the learning model, the social effects model fits the data rather poorly, and yields implausible parameter estimates. We take this as further evidence that social effects are not the source of the spillovers.

[31] Another possible endogeneity problem that we considered is the timing of Christmas effects. All sales variables include a Christmas effect, because we measure sales over a one-year period. However, if the new album is released just prior to Christmas, the holiday sales spike may be larger (for both the new album and the catalog album) than it would be for an album released many months after Christmas, which could generate a spurious positive correlation between S_{kt} and S_{tt}. This does not appear to be empirically important, however: including controls for the season of the new album's release does not meaningfully change our estimates of the learning parameters in equations (7) and (8).

4.2 Results

Table 6 reports nonlinear least squares estimates of equations (7) and (8).[32] Note that estimating the equations separately serves as a check of the model's validity. The model generates separate predictions about the spillovers between the two album pairs, but those predictions are defined in terms of the same parameters. So if the model is a valid description of the data generating process, we should expect the estimates from the two equations to be quantitatively and qualitatively consistent with each other. The similarity of the estimates in columns 1 and 2 of the table suggest our learning model's predictions are robust, especially when compared with the predictions of a model in which the spillovers come from social effects. (In Appendix A we conduct a parallel exercise for a model based on social effects, and find stark inconsistencies between estimates based on albums 1 and 2 vs. 2 and 3.)

Because the counterfactual exercises below will pertain to albums 1 and 2, we will focus our discussion on the first column of Table 6. Our estimate of q_0 implies that on average 18 percent of potential buyers are aware of an artist before she releases her debut album. Although this number may seem somewhat high, it is plausible given that most artists tour extensively (playing small concerts in clubs, or performing as the opening act for a larger band) before ever releasing an album. The shape of the learning function is determined by the parameters a and b, with a representing the baseline learning rate, and b representing the rate at which learning increases with sales. Figure 5 illustrates the learning function implied by our estimates of equation (7). Initially, learning increases at an increasing rate as a function of sales; but eventually the function becomes concave and the fraction of informed consumers approaches one. The inflection point is at 2.54 million sales. As noted above, the logistic learning curve can potentially give rise to multiple equilibria; however, this turns out to be irrelevant given our parameter estimates. Due in particular to the relatively high estimated values of q_0 and a, for the albums in our sample the relationship described in equation (6) has only one fixed point.

Our estimates imply that learning is nearly complete for artists with extremely successful albums. For example, an artist whose debut album sells 10 million copies (which would classify it as a huge hit, and earn it the RIAA "Diamond" award) would be known to 99% of consumers. At the other end of the success spectrum, the majority of consumers remain uninformed: if a debut album

[32] Note that the sample sizes are somewhat smaller than those reported in Table 3. In this model we are aggregating sales over time, and defining the sales periods to be one year. The difference in sample size arises because we exclude artists for whom the one-year release periods overlapped—i.e., for whom the new release came between 8-12 months after the previous release.

sells fewer than 500,000 copies in the first year, our estimates suggest only a third of potential consumers will have discovered the artist in that year.

The estimated value of a suggests that with each album released, at least 16 percent of previously uninformed consumers will discover the artist even if the album has zero sales. Because learning is cumulative in our model, this means an artist could become a household name by releasing a long sequence of very low-quality albums. However, the numbers imply that such an artist would need to release 13 such albums to achieve 90% awareness (i.e. 90% of consumers being aware of the artist). By contrast, a successful artist can become famous with only two or three hit albums. For example, a sequence of three "triple-platinum" albums (sales of 3 million each) would lead 94% of consumers to learn of the artist.

4.3 Counterfactual analyses

Using our estimates of the structural model, we can calculate counterfactual sales under the assumption that all consumers are fully informed about every debut album—i.e., $q_1 \equiv 1$. Specifically, letting $\hat{S}_{11}$ denote counterfactual (complete information) sales, we calculate

$$\hat{S}_{11} = p_{11}N = \frac{S_{11}}{q_1(S_{11})} \ ,$$

where the second equality follows from the fact that observed sales S_{11} are equal to $p_{11}q_1(S_{11})N$.

Figure 6 compares the observed first-year sales of debut albums (S_{11}) to the sales that would have occurred if consumers were fully informed ($\hat{S}_{11}$). The counterfactual distribution of sales is still quite skewed, but it is substantially less concentrated than the distribution of actual sales. The Gini coefficient is .647, as opposed to .724 for actual sales.[33] Perhaps more importantly, among the artists in our sample, fewer than half (48%) sold more than 100,000 units of their debut albums. Our estimates imply that if consumers had been fully informed, nearly three quarters (72%) would have met or exceeded this threshold.

Indeed, the central implication of our estimates is that albums of "mid-range" artists are substantially undersold, in the sense that many would-be buyers remain uninformed. We estimate that the median artist in our sample would have sold nearly 200,000 additional units of the debut album if all consumers had been aware of it. Of course, one might argue that some of these potential

[33] As a point of reference, the income distribution in the United States has a Gini coefficient of around .47.

sales will occur when later releases by the artist generate new information. However, substantial learning will take place only if one of those later releases is a major hit. Furthermore, artists whose early albums are mediocre are more likely to have their careers truncated due to low sales. For example, if an artist's first two albums are only moderately successful, her label may decline to produce any future albums—even though with full information the artist would eventually become a success.[34]

Figure 6 also shows that hit albums are *not* substantially undersold: our estimates imply that almost all consumers are aware of artists with major hits. Albums at the bottom end of the success spectrum are also not undersold, but for a different reason. Even though most consumers are unaware of these albums, the albums' qualities are sufficiently low that sales would be minimal even if everyone were fully informed.

While our estimates indicate that hit albums are not undersold, the model itself also implies that hit albums are not "oversold" either. This is an important distinction from the standard herding model, which has been suggested as an alternative explanation for the skewed distribution of music sales.[35] In a herding model, consumers rationally ignore their own information to follow the herd: they buy what others buy. If consumers behave this way in the market for music, it would imply that not only are some albums undersold, but others are oversold—in the sense that their sales are greater than they would be if everyone made purchase decisions independently based on their own information and preferences. However, we suspect the latter phenomenon is less relevant for recorded music. In markets such as restaurants and books, potential consumers only observe what other consumers buy, so they draw inferences about a product's quality only from the knowledge of its overall popularity. But in music, when other consumers buy an album, the songs on that album get played more frequently on the radio, generating signals that inform consumers' about their *own* preferences for the album. As a result, consumers are less likely to herd on a bad album.[36]

The relatively large differences shown in Figure 6 may of course reflect the dramatic nature of the counterfactual being considered (full information). However, the general implication of our findings is that the distribution of returns in the market for music may be significantly reshaped by technologies that facilitate the diffusion of information. In particular, by making it easier for lesser-known artists to gain exposure, internet technologies for sharing and sampling music should

[34] We do not mean to suggest that all unsuccessful artists are potential stars, but rather that some potential stars' careers may be truncated because consumers were unaware of their music.

[35] See Banerjee [3] and Bikhchandani *et al* [7] for examples of herding models.

[36] This may partly explain why book sales are much more skewed than music sales. (See Sorensen [18] for some evidence and discussion of the skewed distribution of sales for hardcover fiction.)

tend to "flatten" the distribution of success (i.e., make it less skewed). However, our results suggest that struggling artists out in the "long tail" of this distribution are *not* those who benefit most from increased information; instead, moderately successful artists (i.e., nominally profitable ones that don't have any major hits) are the primary beneficiaries. Because our data mostly predate the widespread use of internet music technologies, we cannot directly test whether the internet has in fact changed the shape of returns as predicted; we leave this as a question for future research.[37]

Our estimates of the learning model also allow us to calculate the sales that artists' second albums would have garnered in the absence of a forward spillover—i.e., if the second albums had instead been the debut albums. Letting $\hat{S}_{22}$ denote the counterfactual sales of album 2 if it had instead been the debut album, we can write

$$\hat{S}_{22} = p_{22} q_1(\hat{S}_{22}) N = \frac{q_1(\hat{S}_{22})}{q_1(S_{11}) + [1 - q_1(S_{11})] q_2(S_{22})} \, ,$$

where the second equality follows from the fact that

$$S_{22} = p_{22} \left[q_1(S_{11}) + [1 - q_1(S_{11})] q_2(S_{22}) \right] N \, .$$

We calculate $\hat{S}_{22}$ by finding the root of equation (4.3). These calculations imply that forward spillovers have a substantial impact on sales. The median difference between S_{22} and $\hat{S}_{22}$ (observed album 2 sales, and predicted sales of album 2 if it had instead been the debut album) is 16,450. The largest differences, which occur for artists whose first albums were big hits and whose second albums were smaller hits, are over 1 million. We estimate that the artists in our sample collectively sold 29.48 million more units on their second albums than they would have if those albums were debuts—a difference of roughly 25%.

The presence of a large forward spillover has significant implications for contracts between artists and record labels. Forward spillovers require the recording rights for a new album to be bundled with recording rights for future albums. Investments in new albums yield returns on future albums, and if the album rights are not bundled, these returns will not be fully captured by the investing label: other labels can free-ride and selectively bid for new albums by artists whose previous

[37] A recent paper by Brynjolfsson, Hu, and Simester [9] shows some evidence that internet marketing can make the distribution of sales less concentrated. They show that the sales distribution for a mid-size clothing retailer is significantly less skewed for its internet sales channel than for its catalog channel, and argue that the difference reflects the lower costs of product search on the internet.

albums did well. Hence, in the absence of a long-term contract, the artist will be able to capture some of those investment returns. This is the familiar holdup problem. It reduces the willingness of the label to invest in a new album, leading to underinvestment (and possibly no investment) in that album. Long-term contracts resolve the holdup problem. Because our estimates imply that forward spillovers are quantitatively important, they help explain why virtually all contracts between artists and labels are initially long-term contracts.

These considerations also have relevance to current policy. The Recording Industry Association of America (RIAA) and American Federation of Television and Radio Artists (AFTRA) have repeatedly lobbied Congress to end long-term contracting, as was done in the movie industry in the 1940s (see Terviö (2004)). Our results suggest that eliminating the label's option to extend the terms of the contract for more albums would likely lead to significant inefficiencies. Fewer albums would be produced, and a higher proportion of the albums would be by established artists.

In practice, artists may not be able to commit to a long-term contract. Contract terms are almost always renegotiated after an artist has a successful album. Artists gain bargaining leverage following a hit album because outside labels are willing to pay for the next album. The artist can exploit this leverage by strategically withholding or delaying new recordings, or by (with the help of a lawyer) getting out of a recording contract.[38] However, the artist's inability to commit may be offset to some extent by a lock-in effect caused by the backward spillover. The backward spillover implies that in order to ensure that investments in the new album are efficient, the rights on catalog albums have to be bundled with the rights on the new album. Otherwise, the label that owns the recording rights to the new album will not internalize the impact of its investment on sales of catalog. The incumbent label, who owns the recording rights to catalog, will have an advantage in bidding for the rights to the new album: it is willing to invest more and pay more for those rights than an outside label. Thus, the backward spillover tends to lock in the artist.

Renegotiation outcomes seem to confirm the importance of this lock-in effect. Artists almost always stay with their incumbent labels. In our sample, fewer than 10% of artists ever switched between major labels, and most of the observed switches were due to termination by the incumbent label. Furthermore, artists who negotiate "reversions" in their initial contracts—i.e., clauses stipulating that the rights to the masters revert to the artist after some number of years—typically lease their catalogs to the record label that is producing and distributing their current albums.

[38] A more thorough description of contracting practices is provided in a previous version of this paper [16]. Our understanding of these practices is based largely on conversations with Don Engel, one of the more successful lawyers who specializes in renegotiating contracts. (His press pseudonym is "Busta Contract.")

5 Conclusion

We have shown that the release of a new album generates substantial, persistent increases in the sales of previous albums by the same artist. The evidence strongly suggests that these backward spillovers are generated by consumer learning: a new album release causes some consumers to discover artists and albums about which they were previously uninformed.

Cross-sectional variation in the spillovers allows us to make quantitative inferences about the importance of consumer learning and its impact on market outcomes. Our structural estimates imply that the distribution of sales is substantially more skewed than it would be if consumers were more fully informed. In particular, mid-range artists' albums are dramatically undersold (to the tune of hundreds of thousands of units) relative to what would happen if consumers were all fully informed. We also find that non-debut albums benefit from a large forward spillover, selling tens of thousands more units than they would without the information generated by the prior albums.

More generally, our results illustrate the importance of product discovery in markets with frequent inflows of new products. Other entertainment industries like books, movies, and video games are obvious examples of such markets.[39] Another example is personal computers: Goeree [15] argues that the rapid pace of technological change in computers leads consumers to be less than fully informed about the set of available products. Our findings indicate that the distribution of success in these markets may look very different from what it would be in a world with fully informed consumers.

Also, as in the studies of Ackerberg [1, 2] and Goeree, our results imply an important role for informative advertising. This is particularly interesting in our context, since music is a product for which it is often argued that "prestige" effects of advertising are important. Our interpretation is that new albums generate backward spillovers not because the "buzz" associated with them persuades informed consumers to reconsider their purchase decisions, but because they serve as a form of informative advertising. This of course raises the possibility that similar sales increases might be obtained through direct advertising—i.e., with respect to promoting catalog sales, marketing expenditures may be a substitute for new album releases. However, we suspect that it would be difficult to generate significant spillover sales from direct advertising alone. The reason is that

[39] The highly publicized success of the author Dan Brown provides a clear illustration of backward spillovers and learning in the market for books. In 2003, Brown published a novel that was wildly successful, and its success catapulted one of Brown's earlier novels (initially published in 2000) onto bestseller lists as well—even though the earlier novel had previously sold very few copies and was generally unknown.

radio airplay is the most important form of promotion (because it is the primary channel through which consumers learn about new music), and airplay cannot be bought—at least not legally. By focusing their promotional efforts around the time of a new album's release, the record labels are apparently exploiting economies in advertising.

Appendix A

The model of Section 4 is based on the assumption that consumer learning is the source of the backward spillover. In this appendix we consider an alternative model, in which the spillover arises from social effects in consumers' preferences.

In the learning model, the probability that a consumer buys an album is the product of two probabilities: the probability that he likes the album conditional on knowing about the album, and the probability that he knows about the album. In section 4.1 we assumed that the discovery probability is a function of the new album's sales, while the preference probability is fixed (except for possible age effects). Here we assume the opposite: we set the discovery probability equal to 1 (i.e., all consumers are informed), but let preferences depend on the new album's sales.[40]

The artist releases three albums, $k = 1, 2, 3$, and consumer i's utility from purchasing album k in period t, $k \leq t$, is given by

$$u_{ikt} = \delta_k + \alpha_{kt} S_{tt} + \varepsilon_{ik} \tag{A.1}$$

where δ_k is a measure of album quality (net of price), S_{tt} is the sales of the artist's new release in period t, and ε_{ik} is an idiosyncratic preference shock. Here we assume that utility does not decline over time. As in the learning model, equation (A.1) assumes that social effects in later periods are driven by sales of the *new* albums released in those periods, not by sales of the catalog album. In other words, the social effects are a flow, not a stock. The rationale is again that almost all promotion and radio airplay is associated with newly released albums, so the "buzz" associated with a particular artist will be approximately proportional to the sales of that artist's new album, and not a function of promotion and airplay of previous albums in previous periods. Of course, it may be easier to create a "buzz" on a new album if the previous album was successful. This effect operates indirectly in our model through the correlation in album qualities.

We will assume that consumer i buys album k in its release period if utility exceeds zero. If ε is a logit error, then the probability that consumer i buys album k in its release period is given by

$$p_{kk}(S_{kk}) = \frac{e^{\delta_k + \alpha_{kk} S_{kk}}}{1 + e^{\delta_k + \alpha_{kk} S_{kk}}}.$$

We appeal to the law of large numbers to argue that, in equilibrium, S_{kk} satisfies

[40] Allowing preferences to depend on the album's popularity is in line with the social effects models developed by Becker and Murphy [5] and Brock and Durlauf [8].

$$S_{kk} \simeq p_{kk}(S_{kk})N \ . \tag{A.2}$$

As with the learning model, it is easy to show that a solution always exists, that there are either one or three solutions, and that the minimum and maximum solutions are increasing in δ.

If positive, the probability of purchasing album k in period $t = k+1$ is given by[41]

$$\begin{aligned} p_{k,k+1}(S_{kk}, S_{k+1,k+1}) &= \Pr\{\delta_k + \alpha_{k,k+1}S_{k+1,k+1} + \varepsilon_{ik} > 0 | \delta_k + \alpha_{kk}S_{kk} + \varepsilon_{ik} < 0\} \\ &= \Pr\{-\delta_k - \alpha_{k,k+1}S_{k+1,k+1} < \varepsilon_{ik} < -\delta_k - \alpha_{kk}S_{kk}\} \\ &= \left[1 + e^{\delta_k + \alpha_{kk}S_{kk}}\right]^{-1} - \left[1 + e^{\delta_k + \alpha_{k,k+1}S_{k+1,k+1}}\right]^{-1} \end{aligned} \tag{A.3}$$

and spillover sales are given by

$$S_{k,k+1} \simeq p_{k,k+1}(S_{kk}, S_{k+1,k+1})N \ .$$

For positive spillovers to occur, the social effects of album $k+1$ need to exceed those of album k. This is possible only if social effects focus at least in part on the artist rather than the album, since in the latter case $\alpha_{k,k+1}$ is zero by definition. If social effects are pure artist effects, then their impact on album k is the same as on album $k+1$, so $\alpha_{k,k+1} = \alpha_{k+1,k+1}$. In general, we would expect $\alpha_{k,k+1} < \alpha_{k+1,k+1}$. Also, note that if $\alpha_{k,k+1} = \alpha_{kk}$, the spillover effect is positive if and only if album $k+1$ sales exceed album k sales.

To estimate the model, we use the equation for p_{kk} to obtain the relation

$$\log\left(\frac{p_{kk}}{1-p_{kk}}\right) = \delta_k + \alpha_{kk}S_{kk} \ ,$$

which, treating equation (A.2) as an equality, implies that

$$\delta_k = \log\left(\frac{S_{kk}}{N - S_{kk}}\right) - \alpha_{kk}S_{kk} \ .$$

Substituting this expression for δ_k into equation (A.3), we get

$$p_{k,k+1}(S_{kk}, S_{k+1,k+1}) = \left[1 + \frac{S_{kk}}{N-S_{kk}}\right]^{-1} - \left[1 + \frac{S_{kk}}{N-S_{kk}} e^{\alpha_{k,k+1}S_{k+1,k+1} - \alpha_{kk}S_{kk}}\right]^{-1} \ .$$

[41] To be consistent with the analysis in Section 4, we focus on adjacent album pairs.

which implies the estimating equation:

$$S_{k,k+1} = N\left[1+\frac{S_{kk}}{N-S_{kk}}\right]^{-1} - N\left[1+\frac{S_{kk}}{N-S_{kk}}e^{\alpha_{k,k+1}S_{k+1,k+1}-\alpha_{kk}S_{kk}}\right]^{-1} + \eta \tag{A.4}$$

where η represents an artist-specific error term.

Table A.1 reports nonlinear least squares estimates of this equation. The first column reports estimates obtained from album 2's spillover onto album 1; the second column uses album 3's spillover onto album 2. We treat the market size, N, as a parameter to be estimated; in unreported results we fixed this parameter at various levels and obtained similar estimates for the social effects parameters (α_{kk} and $\alpha_{k,k+1}$). We also estimated a model with a correction for time between releases (as in the learning model of section 4), which also yielded similar results.

The table illustrates the sense in which the social effects model yields a poor fit of the data. We estimate that α_{12} exceeds α_{11}, which means the popularity of album 2 has more influence on demand for album 1 than does the popularity of album 1 itself. The result that α_{12} is also greater than α_{22} means that the popularity of album 2 has a greater impact on sales of album 1 than on sales of album 2. We consider these to be implausible implications—or at least implications that are inconsistent with conventional theory about how social effects work.

Why do the data reject the social effects model in this way? Essentially, due to the selection effect described in Section 3.4, a social effects model has a difficult time explaining why we sometimes see significant backward spillovers even when the new album was less popular than the catalog album. The fact that we observe such spillovers in the data forces the estimate of $\alpha_{k,k+1}$ to be significantly higher than the estimate of α_{kk}.

References

[1] Ackerberg, D., "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics*, vol. 32 no. 2, Summer 2001.

[2] Ackerberg, D., "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, vol. 44 no. 3, August 2003.

[3] Banerjee, A. "A Simple Model of Herd Behavior," *Quarterly Journal of Economics*, v. 107 n. 3, August 1992, pp. 797-817.

[4] Becker, G., Grossman, M., and K. Murphy, "An Empirical Analysis of Cigarette Addiction," *American Economic Review*, v. 84 n. 3, June 1994, pp. 396-418.

[5] Becker, G., and Murphy, K., *Social Economics: Market Behavior in a Social Environment*, Harvard University Press, 2000.

[6] Benkard, C. L., "Learning and Forgetting: The Dynamics of Aircraft Production," *American Economic Review*, v. 90 n. 4, September 2000, pp. 1034-1054.

[7] Bikhchandani, S., Hirshleifer, D. and Welch, I., "A Theory of Fads, Fashions, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 100, October 1992, pp. 992-1026.

[8] Brock, W. and S. Durlauf, "Interactions-Based Models," in *Handbook of Econometrics*, vol. 5., J. Heckman and E. Leamer eds., Amsterdam: North-Holland, 2001.

[9] Brynjolfsson, E., Hu, Y., and D. Simester, "Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales," working paper, MIT Sloan School of Management, December 2006.

[10] Cabral, L.M.B., "Stretching Firm and Brand Reputation," *Rand Journal of Economics*, 31, 4, Winter 2000, pp. 658-673.

[11] Choi, J.P., "Brand Extension as Informational Leverage," *Review of Economic Studies*, v. 65 (1998), pp. 655-669.

[12] Crawford, G. and M. Shum, "Uncertainty and Learning in Pharmaceutical Demand," *Econometrica*, vol. 73, no. 4, July 2005.

[13] "Details About the AP/Rolling Stone Poll on Downloading Music," Associated Press, Feb. 2., 2006.

[14] Gentzkow, M., "Valuing New Goods in a Model with Complementarity: Online Newspapers," forthcoming, *American Economic Review*.

[15] Goeree, M. S., 2005, "Advertising in the U.S. Personal Computer Industry," Claremont McKenna working paper, 2005.

[16] Hendricks, K. and A. Sorensen, "Information Spillovers in the Market for Recorded Music," NBER working paper #12263, 2006.

[17] Krasilovsky, M.W., Sidney Shemel, John M. Gross, 2000, *This Business of Music: The Definitive Guide to the Music Industry,* 9th Edition, Billboard Books.

[18] Sorensen, A., "Bestseller Lists and Product Variety: The Case of Book Sales," forthcoming, *Journal of Industrial Economics*.

[19] *Soundata National Music Consumer Study*, National Association of Recording Merchandisers, 1995.

[20] Terviö, M., "Mediocrity in Talent Markets," University of Berkeley working paper, December, 2004.

[21] Wernerfelt, B., "Umbrella Branding as a Signal of New Product Quality: An Example of Signalling by Posting a Bond." *RAND Journal of Economics*, v. 19 (1988), pp. 458-466.

[22] Wooldridge, J., 2002, *Econometric Analysis of Cross-section and Panel Data*, MIT Press.

Table 1: Summary Statistics

	N	Mean	Std. Dev.	Percentiles .10	Percentiles .50	Percentiles .90
Date of release:						
album 1	355	13may1996	102	22aug1993	05may1996	28feb1999
2	355	20jul1998	108	23jul1995	02aug1998	27may2001
3	178	03jun1999	90	13oct1996	04aug1999	05aug2001
First year sales:						
album 1	355	312,074	755,251	7,381	78,360	781,801
2	355	367,103	935,912	10,705	55,675	951,956
3	178	450,716	867,630	7,837	71,674	1,461,214
overall	888	361,864	854,420	9,095	67,558	996,460
First 4 weeks / First year:						
album 1	355	.121	.111	.016	.085	.265
2	355	.263	.137	.086	.263	.441
3	178	.305	.131	.134	.305	.500
overall	888	.214	.148	.031	.198	.419
Peak sales week:						
album 1	355	31.9	47.8	0	15	87
2	355	7.83	23.1	0	0	28
3	178	4.05	13.1	0	0	12
overall	888	16.7	36.3	0	1	46
Weeks between releases:						
1 & 2	355	114	53.5	58	107	179
2 & 3	178	111	46.7	58	104	169

Table 2: Seasonality in release dates

	Percent of releases occurring			
Month	Album 1 (n=355)	Album 2 (n=355)	Album 3 (n=178)	Overall (n=888)
Jan	3.94	3.10	3.37	3.49
Feb	8.17	4.23	3.93	5.74
Mar	13.24	9.58	11.80	11.49
Apr	9.01	8.45	8.99	8.78
May	11.83	9.01	7.30	9.80
Jun	7.61	12.68	6.74	9.46
Jul	8.45	9.01	10.11	9.01
Aug	11.55	9.58	10.67	10.59
Sep	7.32	11.27	11.80	9.80
Oct	12.39	10.70	16.29	12.50
Nov	5.92	11.83	6.74	8.45
Dec	0.56	0.56	2.25	0.90

Table 3: Estimated Effects of New Releases on Sales of Catalog Albums

Week (relative to release date)	Baseline model (1)		First-differenced model (2)	
	2→1	3→2	2→1	3→2
t=-13	-0.006	0.041	-0.024	0.041
	(0.017)	(0.025)	(0.016)	(0.024)
t=-12	0.022	0.013	0.010	-0.029
	(0.022)	(0.032)	(0.016)	(0.024)
t=-11	0.044	-0.048	0.012	-0.060
	(0.025)	(0.035)	(0.017)	(0.024)
t=-10	0.059	0.024	0.010	0.073
	(0.028)	(0.037)	(0.016)	(0.024)
t=-9	0.066	0.052	-0.000	0.031
	(0.029)	(0.039)	(0.016)	(0.024)
t=-8	0.078	0.055	0.008	0.011
	(0.030)	(0.040)	(0.016)	(0.024)
t=-7	0.124	0.074	0.044	0.029
	(0.031)	(0.040)	(0.017)	(0.024)
t=-6	0.148	0.090	0.022	0.028
	(0.031)	(0.041)	(0.017)	(0.024)
t=-5	0.201	0.121	0.054	0.038
	(0.032)	(0.041)	(0.016)	(0.024)
t=-4	0.260	0.177	0.057	0.060
	(0.032)	(0.042)	(0.016)	(0.024)
t=-3	0.301	0.242	0.042	0.073
	(0.033)	(0.042)	(0.016)	(0.024)
t=-2	0.346	0.257	0.050	0.022
	(0.033)	(0.042)	(0.016)	(0.025)
t=-1	0.419	0.332	0.079	0.089
	(0.033)	(0.042)	(0.017)	(0.025)
t=0	0.471	0.361	0.055	0.040
	(0.033)	(0.043)	(0.017)	(0.024)
t=1	0.449	0.311	-0.018	-0.038
	(0.034)	(0.043)	(0.016)	(0.025)
t=2	0.443	0.310	-0.007	0.003
	(0.034)	(0.043)	(0.016)	(0.025)
t=3	0.425	0.286	-0.026	-0.023
	(0.034)	(0.043)	(0.016)	(0.025)
t=4	0.455	0.271	0.018	-0.014
	(0.034)	(0.043)	(0.016)	(0.025)
t=5	0.455	0.254	-0.019	-0.022
	(0.034)	(0.044)	(0.016)	(0.024)
t=6	0.492	0.277	0.013	0.019
	(0.034)	(0.044)	(0.016)	(0.025)
t=7	0.509	0.263	-0.003	-0.021
	(0.035)	(0.044)	(0.017)	(0.025)

(continued next page)

Table 3: (continued)

Week (relative to release date)	Baseline model (1) 2→1	Baseline model (1) 3→2	First-differenced model (2) 2→1	First-differenced model (2) 3→2
t=8	0.516	0.273	-0.008	0.006
	(0.035)	(0.044)	(0.016)	(0.025)
t=9	0.474	0.268	-0.050	-0.014
	(0.035)	(0.044)	(0.016)	(0.025)
t=10	0.490	0.312	0.014	0.029
	(0.035)	(0.044)	(0.017)	(0.025)
t=11	0.489	0.339	-0.003	0.015
	(0.035)	(0.045)	(0.016)	(0.025)
t=12	0.495	0.336	-0.007	-0.022
	(0.035)	(0.045)	(0.017)	(0.025)
t=13	0.530	0.289	0.023	-0.051
	(0.035)	(0.045)	(0.017)	(0.025)
t=14	0.562	0.299	0.021	0.015
	(0.035)	(0.045)	(0.017)	(0.025)
t=15	0.530	0.255	-0.037	-0.027
	(0.036)	(0.045)	(0.017)	(0.025)
t=16	0.517	0.244	-0.013	-0.002
	(0.036)	(0.046)	(0.016)	(0.025)
t=17	0.533	0.213	0.019	-0.017
	(0.036)	(0.046)	(0.016)	(0.025)
t=18	0.532	0.223	-0.003	0.013
	(0.036)	(0.046)	(0.016)	(0.024)
t=19	0.545	0.161	0.007	-0.060
	(0.036)	(0.046)	(0.016)	(0.025)
t=20	0.561	0.172	0.008	0.014
	(0.037)	(0.047)	(0.016)	(0.025)
t=21	0.515	0.178	-0.050	0.005
	(0.037)	(0.047)	(0.017)	(0.025)
t=22	0.547	0.168	0.030	-0.009
	(0.037)	(0.047)	(0.016)	(0.024)
t=23	0.561	0.183	0.010	0.019
	(0.037)	(0.047)	(0.016)	(0.024)
t=24	0.566	0.154	-0.007	-0.027
	(0.037)	(0.047)	(0.016)	(0.025)
t=25	0.581	0.137	0.001	-0.013
	(0.037)	(0.047)	(0.017)	(0.025)
# albums	338	173	338	173
# observations	33,581	17,073	33,509	17,038
$\hat{\rho}$	.800	.736	-.220	-.270

Estimates of the regressions described in equations 1 and 2, with standard errors in parentheses corrected for heteroskedasticity across albums and autocorrelation within albums. Estimated coefficients for time and seasonal dummies are suppressed to save space. Each column represents an album pair: e.g., the column labeled 3→2 lists the estimated effects of album 3's release on the sales of album 2. $t = 0$ is the first week following the release of the new album. The $\hat{\rho}$'s are the estimated AR(1) coefficients, reflecting the degree of serial correlation in demand shocks for a given album.

Table 4: Spillovers and hits

Album 1, Album 2:	Hit, Hit	Hit, Not	Not, Hit	Not, Not
N	53	45	34	206
Median # weeks to release 2	108	124	101	104
Median weekly sales (album 1) prior to release:	1,888	318	342	154
Median weekly decline around release:	-0.021	-0.018	-0.018	-0.011
Estimated total change in sales:	22,161	660	14,557	883
Percentage change in sales:	42.7	7.2	148.5	17.6
Average of (sales before next release)/(first 4 years' sales):	0.73	0.85	0.55	0.62
Album 2, Album 3:	Hit, Hit	Hit, Not	Not, Hit	Not, Not
N	49	13	12	99
Median # weeks to release 3	105	117	95	103
Median weekly sales (album 1) prior to release:	1,555	466	844	85
Median weekly decline around release:	-0.013	-0.026	0.004	-0.010
Estimated total change in sales:	19,884	1,110	20,788	687
Percentage change in sales:	40.6	9.5	56.4	24.6
Average of (sales before next release)/(first 4 years' sales):	0.73	0.84	0.59	0.65

Hits are defined as albums that sold over 250,000 units nationally in the first year. Albums that didn't clear this threshold are the "Not" albums (i.e., not hits). The estimated total changes and percentage changes in sales reflect increases over the 39-week treatment window.

Table 5: Sales and spillovers in the artist's home market

	2→1	3→2
Home market ($\hat{\psi}$)	0.814	0.647
	(0.006)	(0.008)
Home market $\times$ new release period ($\hat{\gamma}$)	-0.105	-0.104
	(0.010)	(0.013)
# observations	2,727,890	1,437,340
# artists	268	142

Estimates of the regression model described in equation (3); the dependent variable is log sales. $\hat{\psi}$ measures the average difference in log sales between the artist's home market vs. other markets, and $\hat{\gamma}$ measures the average difference in the backward spillover in the artist's home market vs. other markets. Other coefficients are omitted to save space.

Table 6: Estimated parameters of learning model

		Equation (7) (2⟶1)	Equation (8) (3⟶2)
"Baseline" awareness:	q_0	0.180	0.267
		(0.049)	(0.087)
Learning function parameters:	a	0.161	0.244
		(0.072)	(0.164)
	b	0.065	0.036
		(0.013)	(0.012)
Time between releases:	γ	0.701	0.503
		(0.073)	(0.106)
	N	311	165
	R^2	.743	.823

Asymptotic standard errors in parentheses.

Table A.1: Estimated parameters of social effects model

		Equation (A.4) (2⟶1)	Equation (A.4) (3⟶2)
Market size:	N	8.779	12.907
		(0.324)	(0.386)
Social effects parameters:	α_{kk}	-0.003	-0.010
		(0.005)	(0.004)
	$\alpha_{k,k+1}$	0.079	0.033
		(0.005)	(0.005)
	Sample size	311	165
	R^2	.760	.699

Asymptotic standard errors in parentheses.

Figure 1: Album sales paths for two examples

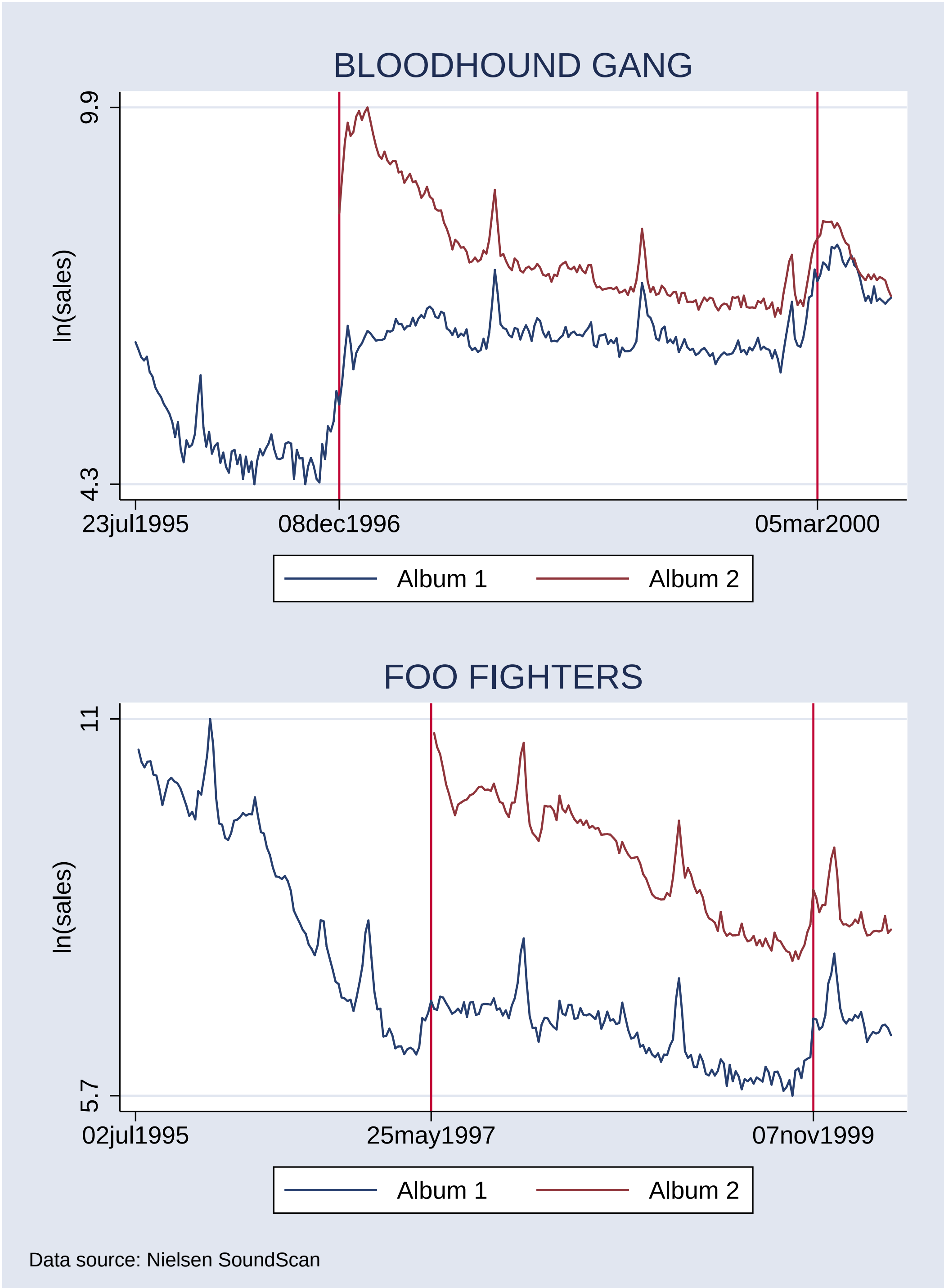

Figure 2: Distributions of Elapsed Time Between Releases

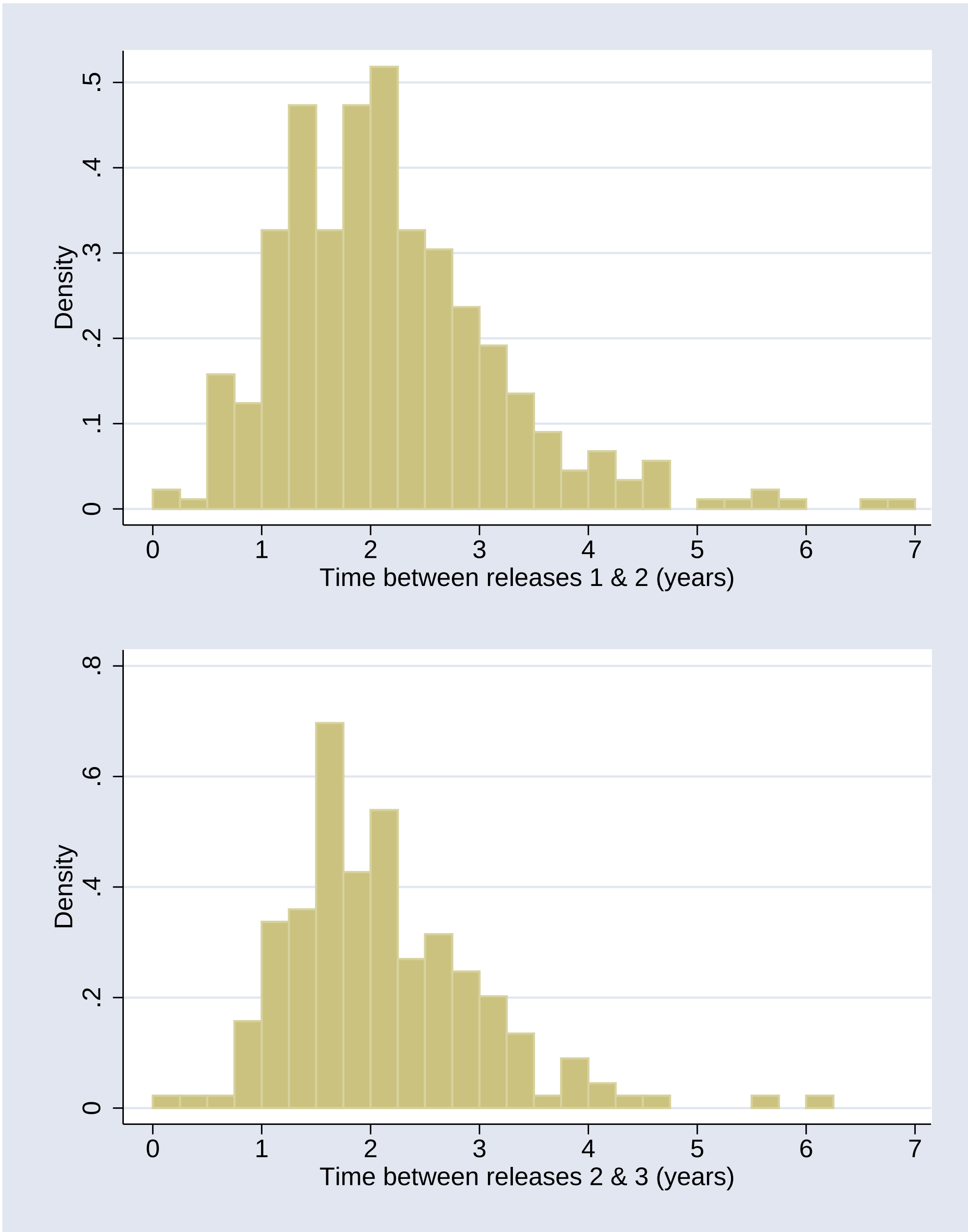

Figure 3: Time patterns of backward spillovers

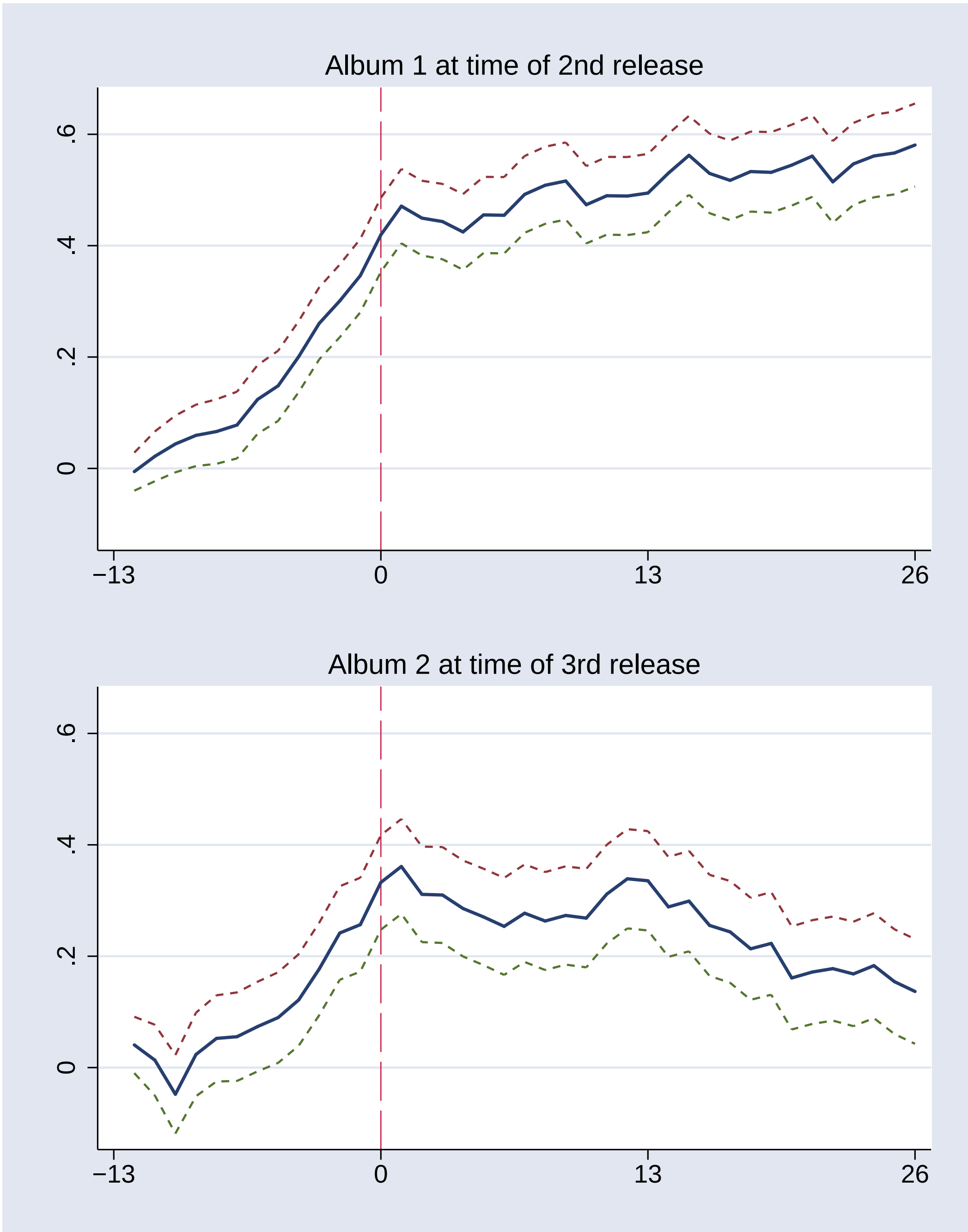

Figure 4: Time patterns of backward spillovers: first-differences model

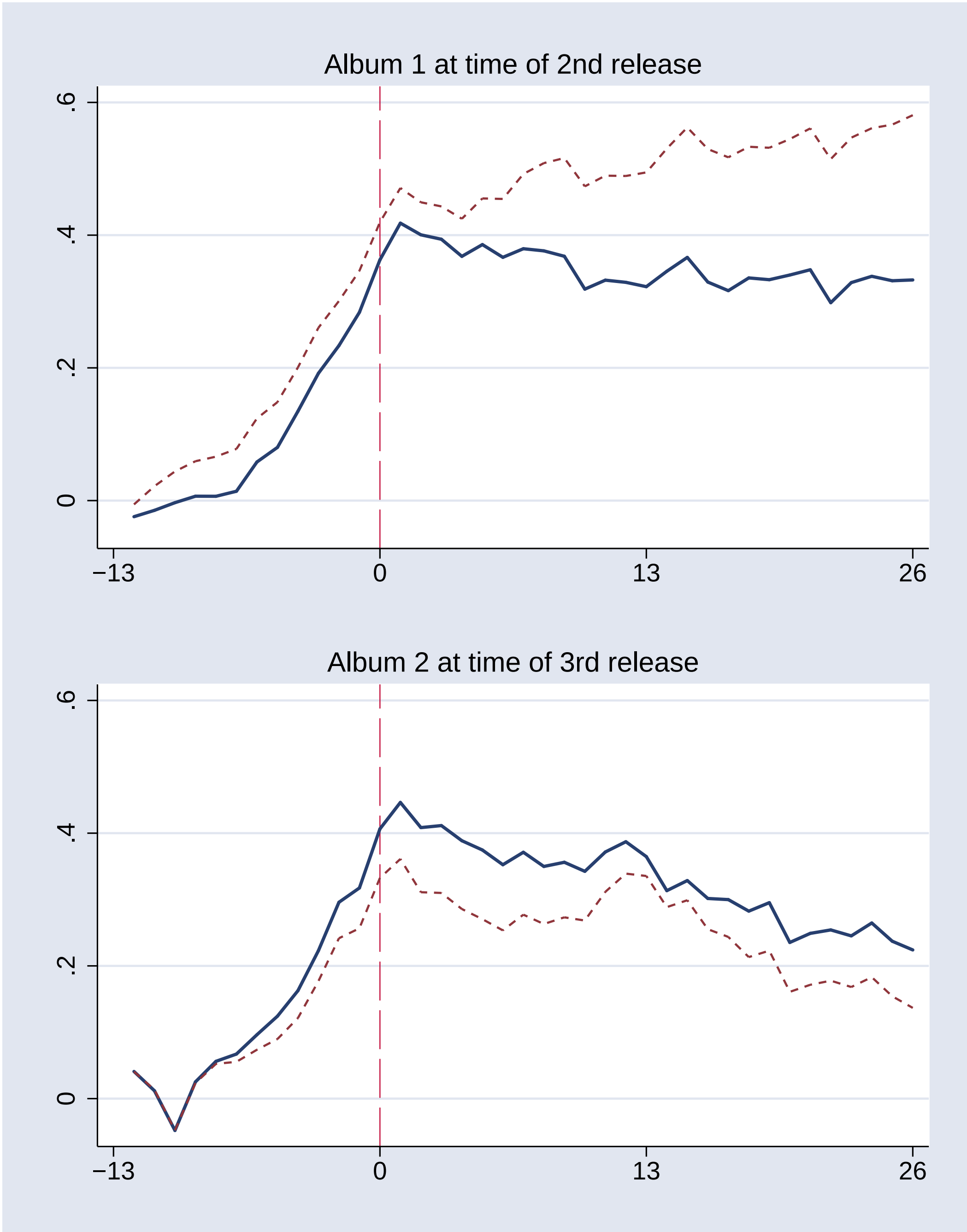

Figure 5: Estimated learning function

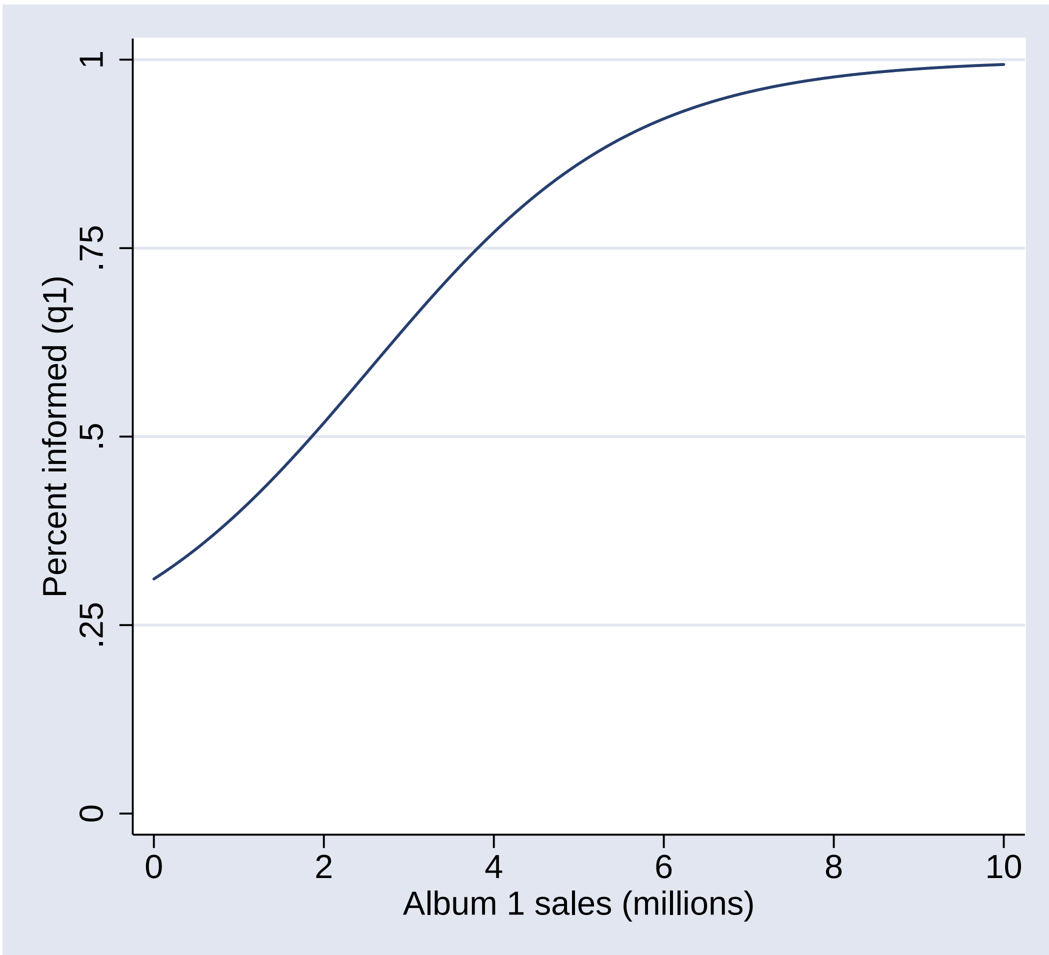

Figure 6: Counterfactual sales distribution for debut albums