An Empirical Study of Investment Externalities: The Case of Albums by the Same Recording Artist

Ken Hendricks
Department of Economics
University of Texas at Austin

Alan Sorensen
Graduate School of Business
Stanford University

April 12, 2005
PRELIMINARY AND INCOMPLETE
Please do not quote

Abstract

This paper studies the role of investment externalities on the career of a recording artist. We show that a new album increases sales of the old albums, and the increase is substantial and permanent. The externality is potentially a barrier to entry. It prevents artists from switching distributors and mitigates the distributor's hold-up problem. We show that the externality is mainly associated with first-time hits, which suggests that the source of the externality is informational and not preference complementarities.

1 Introduction

In this paper, we investigate the role of investment externalities on the career of an artist. We distinguish two kinds of investment externalities. The *forward externality* refers to the impact of investing in a new album on sales of future albums. At any point in time, consumers may be familiar with only a relatively small fraction of existing artists and to have heard only a small subset of the available albums. Since new releases typically get more playing time at radio stations than older releases, and artists frequently go on tour to promote a new album, the new release may enhance consumer awareness of the artist and increase the stock of potential consumers for her future albums. If albums are complements, then the new album makes consumer more likely to purchase future albums. On the other hand, if consumers have a taste for diversity, then the new release is likely to reduce demand for the artist's future albums. The *backward externality* refers to the impact of investing in a new album on sales of previously released albums by that artist. An album is classified as catalogue approximately 12-18 months after its release. If albums are complements, or if the new album increases the stock of informed consumers, then the new release could cause an increase in sales of catalogue albums. On the other hand, if consumers prefer new albums to old albums, then the new release could cannibalize sales of catalogue albums.

Figure 1 illustrates the backward externality. 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 catalogue titles increase 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. For the "Foo Fighters,"a more popular hard rock band with a very successful debut album, the impact of the second release is somewhat less dramatic, but still seems to have generated an increase in sales of the band's first album. In both examples, the backward externality is significantly positive. The effect appears to begin in the weeks just prior to the new album's release, and it persists for many months. In fact, for the "Bloodhound Gang," the effect persisted for at least three years.

The motivation for studying the impact of a new release on catalogue sales is that a positive backward externality is potentially a barrier to entry. We develop a simple model of the contracting problem between the artist and distributor with positive externalities and explore their implications for contract renegotiations and investment. We show how the backward externality locks in the artist and mitigates the effects of the hold-up problem created by the forward externality. If the artist tries to sell the rights to a new release, the incumbent distributor (i.e., the one that financed her previous albums) has an advantage. It internalizes the backward externality of its investment in the new release on catalogue sales, whereas its rivals do not. Hence, the incumbent's willingness to invest in the new album, and to pay for it, exceeds that of its rivals. Thus, even if artists and distributors cannot commit to long-term contracts, artists will not switch distributors, and this indeed appears to be true in our sample. But, if a new distributor cannot easily bid away established artists, then the only way it can enter the industry is by developing its own stable of artists, which can take a long time, and may be impossible to accomplish without a prior reputation from having produced hit artists before.

Our main empirical goals in this paper are to measure the backward externality, determine the conditions under which it occurs, and identify the source of the externality. We study these questions using a large sample of recording artists whose debut albums were released in the United States during the period 1993 to 2002. Many of the artists in our sample released as many as four albums during the sample period, allowing us to study the variation in backward externality over different pairs of albums (e.g., release of albums 3 and 4 on sales of albums 1 and 2, and album 4 on sales of album 3).

Our empirical strategy for quantifying the backward externality is taken from the literature on treatment effects. The release of a new album is interpreted as "treatment,"and we are interested in measuring the difference between catalogue sales of the artist with treatment and without treatment. Since the artist cannot be in both states at the same time, we only observe one of the outcomes. However, if release times are random, then catalogue sales for "untreated" artists with the same number of catalogue albums can be used to estimate the counterfactual sales for "treated" artists. Of course, the variation in release times across artists may not be entirely random. We use fixed effects to control for time-invariant factors such as genre and artist popularity that may influence release times, and conduct various checks to determine the robustness of our results. We find that the average treatment effects are positive, permanent, and both statistically and economically significant.

We find that the magnitude of the backward externality is much larger when the new release is the artist's first hit than when the new release follows a hit. Figure 1 illustrates this pattern. The debut album of the Bloodhound Gang did not sell well prior to the release of their second album, which was a big hit. The backward externality is huge. A large fraction of the consumers appear to have purchased both albums following the release of the second album. In the case of the Foo Fighters, both the debut album and the second album were hits, and the backward externality is quite small. This pattern sheds light on the source of the backward externality. We argue that our results support the hypothesis that the backward externality is due to preference learning and not preference complementarity.

The paper is organized as follows. In Section 2 we present the contracting model. Section 3 discusses the data. Section 4 outlines the empirical strategy for estimating the backward externality and reports the main estimation results. Section 5 examines the sources of the backward externality. Section 6 presents data on the incidence of artists switching between labels. Section 7 concludes.

2 Music Contract Model[1]

The standard contract between the recording artist and a label (or distribution company) is a royalty contract.[2] The recording artist agrees to deliver an album within a specified time period. The label agrees to give the artist an advance to finance recording costs but recoups the advance from the artist's royalties. The advance varies across contracts, but it typically ranges between $100,000 to $200,000 with very successful artists able to command substantially higher advances. The royalty rates are usually between 15 to 20 percent. The contract gives the label the option to extend the term of the contract for more albums at the same terms and conditions that apply to the first album. However, in practise, contract terms are frequently renegotiated after an artist has a hit.

The timeline of our model is as follows. Let θ_k denote the quality of album k, $k = 1, 2$, and let $\theta = (\theta_1, \theta_2)$. In period 0, the artist makes a random draw θ_1 from distribution H_{θ_1}. Both artist and label observe a signal S about θ_1 and sign a royalty contract that gives the artist a share $\alpha \in (0, 1)$ of any revenues from album 1. Let I denote the label that invests in album 1. In period 1, Label I chooses its investment $i_1$in album 1 and revenue $r(i_1, \theta_1)$ is realized. Both artist and label learn θ_1. Artist auctions rights to second album, possibly switching to new label E. We shall assume that the royalty rate on album 2 is fixed, so negotiations are only over the fixed transfer. After signing new album contract, artist draws θ_2 from distribution $H_{\theta_2|\theta_1}$. In period 2, Label i, $i = I, E$, chooses its investment i_2 in album 2 and revenues $R_j(i_{1,} i_2; \theta_1, \theta_2)$ for both albums are realized.

We will assume that the random variables S, θ_1, θ_2 are affiliated and that S, θ_2 are independent conditional on θ_1. The former assumption implies that the expected quality of album 1 is increasing in S, and the expected quality of album 2 is increasing in the quality of album 1. The latter assumption implies that S is uninformative about θ_2 once θ_1 is known.

Investment in an album consists mostly of fixed costs: recording, distribution, and promotion costs. Let F denote the minimum amount that a label has to invest in order to bring a CD to market. As mentioned above, the label recoups recording costs from the artist's royalties but, in what follows, we ignore this feature of the contract to simplify the presentation of the main ideas.

The revenue functions are reduced form. Obviously, the revenues of album 1 in period 1 are increasing in the label's investment in the album and the album's quality. However, the properties of the revenue functions in period 2 depend upon the underlying choice model. We impose the

[1]We are indebted to Michael Whinston for helping us formalize the contracting problem.

[2]See Krasilovsky et al's [4] book on contracts in the music industry.

following restrictions.

Assumption 1: (i) R_2 is increasing, concave, and supermodular in (i_1, i_2), nondecreasing in θ_1, θ_2; (ii) R_1 is nondecreasing in i_2, θ_2.

We say that investment in album 2 generates a *backward externality* if $R_1(i_1, i_2, \theta)$ is strictly increasing in i_2. Similarly, investment in album 1 generates a *forward externality* if $R_2(i_1, i_2, \theta)$ is strictly increasing in i_1. Note that we have not made any assumptions about the impact of label I's investment in album 1 and its quality on sales of album 1 in period 2. In the appendix, we examine two plausible choice models that yield revenue functions consistent with Assumption 1. In one model, preferences are supermodular, so albums are complements and, in the other, preferences are additive but consumers learn about their preferences for the albums.

The kind of choice model that Assumption 1 rules out is one in which albums are substitutes. In this model, higher sales of one album cannibalizes sales of the other album, and the label would have to worry about these cannibalization effects in choosing its investments.

2.1 Period 2 Behavior

If the artist switches to outside label E, investment in album 2 will be

$$i_2^E(i_1, \theta_1) = \arg\max_{i_2} E_{\theta_2|\theta_1}[(1-\alpha)R_2(i_1, i_2; \theta) - i_2].$$

Recall that investment has to be at least equal to F. Hence, if θ_1 is sufficiently low, then the expected revenues from album 2 may not exceed recording and distribution costs, in which case, the optimal decision is not to invest in the album. Let θ_E denote the value of θ_1 at which album 2 breaks even. It follows from Assumption 1 that θ_E is unique and that i_2^E is strictly increasing in $\theta_1 \geq \theta_E$.

If the artist stays with label I, investment will be

$$i_2^I(i_1, \theta_1) = \arg\max_{i_2} E_{\theta_2|\theta_1}[(1-\alpha)(R_1(i_1, i_2; \theta) + R_2(i_1, i_2; \theta)) - i_2].$$

Let θ_I denote the value of θ_1 at which album 2 just profitable for Label I. Uniqueness follows from Assumption 1. If investment in album 2 generates a backward externality, then $\theta_I < \theta_E$ and $i_2^E(i_1, \theta_1) < i_2^I(i_1, \theta_1)$ for $\theta_1 \geq \theta_I$. In the absence of a backward externality, the investments by the two labels are the same and the conditions under which they are willing to invest in album 2 are also the same.

2.2 Bargaining

In negotiating a new album contract or, alternatively, renegotiating the old contract, the artist's threat point is label E's willingness to pay for the rights to album 2. Assuming the artist can get label E to bid its maximum willingness to pay, the artist can "hold up" the incumbent label for a payoff that is at least equal to

$$d_E(i_1, i_2^E(i_1,\theta_1),\theta_1) = E_{\theta_2|\theta_1}[R_2(i_1, i_2^E(i_1,\theta_1);\theta)] - i_2^E(i_1,\theta_1)$$

for $\theta_1 \geq \theta_E$. If the artist switches, then Label I's payoff in this range of album 1 quality is

$$d_I(i_1, i_2^I(i_1,\theta_1),\theta_1) = E_{\theta_2|\theta_1}[(1-\alpha)(R_1(i_1, i_2^E(i_1,\theta_1);\theta)].$$

However, if the artist does not switch, then Label I's maximum payoff is

$$\begin{aligned} V(i_1, i_2^I(i_1,\theta_1),\theta_1) &= E_{\theta_2|\theta_1}\{(1-\alpha)(R_1(i_1, i_2^I(i_1,\theta_1);\theta) + R_2(i_1, i_2^I(i_1,\theta_1);\theta))\} \\ &\quad -d_E(i_1, i_2^E(i_1,\theta_1),\theta_1) - i_2^I(i_1,\theta_1). \end{aligned}$$

Let

$$S(i_1,\theta_1) = V(i_1, i_2^I(i_1,\theta_1),\theta_1) - d_I(i_1, i_1^E(i_1,\theta_1),\theta_1)$$

denote the bargaining surplus. Here S is strictly positive if and only if investment in album 2 generates a backward externality. When this is the case, the artist never switches labels and the backward externality represents a barrier to entry.

2.3 Period 1 Investment

Label I's investment in album 1 is

$$i_1^I = \arg\max_{i_1} E_{\theta_1|S}[(1-\alpha)r(i_1,\theta_1) + \delta V(i_1, i_2^I(i_1,\theta_1),\theta_1) - i_1].$$

Suppose the revenue functions are additively separable in i_1 and i_2. That is,

$$R_1(i_1,i_2) = a_1 i_1 + a_2 i_2,\ R_2(i_1,i_2) = b_1 i_1 + b_2 i_2.$$

Then the investment in album 2 by either label is independent of Label I's investment in album 1. Similarly, increases in the magnitude of a_2, the backward externality parameter, has no marginal impact on investment in album 1. Hence, the amount invested in album 1 does not depend upon

Label I's investment in album 2.[3] However, the magnitude of the backward externality does affect the threshold signal below which investing in album 1 is not expected to be profitable. An increase in a_2 increases the difference between i_2^I and i_2^E and increases the bargaining surplus available to Label I. As a result, album 1 is profitable at lower values of the signal on θ_1. It is in this sense that a positive backward externality mitigates the hold-up problem, even when investments are not complementary, and affects the label's investment in album 1. If investments are complementary, then the backward externality also affects the investment in album 1.

If the artist is able to commit not to renegotiate her contract, eliminating the hold-up problem, marginal artists are more likely to obtain investment.

2.4 Discussion

We do not directly observe the investment levels or album quality. This limits our ability to determine the properties of the reduced form revenue functions. However, we do observe when the second album is released, and the sales of album 1 before and after its release. Therefore, using sales as proxies for album quality, and treating i_j as a binary variable that is equal to one if the label invests in album j and zero otherwise, we can study the properties of

$$\Delta(\theta_1, \theta_2) = [R_1(1, 1; \theta_1, \theta_2) - R_1(1, 0; \theta_1, \theta_2)].$$

The key empirical issue is predicting the revenues of album 1 in the counterfactual world where album 2 is not released. Following the treatment literature, we will use album 1 sales of other, comparable artists to forecast $R_1(1, 0; \theta_1, \theta_2)$. If Δ is estimated to be positive, then the backward externality is present. We also examine the variation in Δ with respect to θ_1 and θ_2 to determine the potential source of the backward externality.

3 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

[3] We thank Michael Whinston for making this point.

the basis for the ubiquitous Billboard charts that track artist popularity. Various online databases, most notably *allmusic.com*, were also consulted for auxiliary information 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.[4] 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 Heatseeker 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. (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.)

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 albums, and anthologies or "greatest hits" albums are excluded from the analysis because they rarely contain any new music that could be expected to affect demand for previous albums.[5] The resulting sets of albums were compared against online sources of artist discographies to verify that we had sales data for each artists' complete album history; we dropped any artists for whom albums were missing or for which the sales data were incomplete.[6] Since timing of releases is an

[4]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.

[5]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.

[6]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.

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.[7] 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 artists' third and fourth albums in the analysis if we observe at least the first 52 weeks of sales for those albums (i.e., we include third and fourth albums if they were released before 2002).

After applying all of these filters, the remaining sample contains 355 artists and 962 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.

For each album in the sample, we observe weekly sales from the time of its release through the end of 2002. The key feature of the data is that sales are reported at the album level, so that we can observe the sales of prior albums when a new album is released. Both cross-sectional and time-series variation can be exploited to measure the sales responses: for a given album, we observe both that album's sales history prior to the new release and also sales paths for other comparable artists who did not release new albums.

Table 1 summarizes various important patterns in 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 74 artists in the sample for whom we observe 4 releases during the sample period, another 104 for whom we observe 3 releases, 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.

The second panel of the table illustrates the considerable heterogeneity in sales across albums.

[7] 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.

Recording and distribution costs for a typical album are in the ballpark of $200,000-$300,000, so an album must sell roughly 15,000 units (at around $16 per unit) in order to be barely profitable; most 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 first release are ten times sales at the median and over 100 times sales of the album 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 highly successful second release, so that 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 (third and fourth) releases the correlation is 0.77 (0.70). 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 for 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 appears to hold 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. In our sample, among the 74 artists for whom we observe four releases, 42 had the greatest success with either the first or second release.

Figure 2 shows "typical" sales paths for first and second releases, as well as the typical timing of later releases. (The paths depicted are kernel regressions of monthly sales on time since release; the vertical lines marking subsequent release dates reflect median time to release.) Although there is obviously heterogeneity across albums—not every album's sales path looks like the one in the picture—the figure conveys the predominant pattern: an early peak followed by a steady, roughly exponential decline. As indicated in the third and fourth panels of table, 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 seems to increase 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 lists 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. The distribution of time between albums 2 and 3 is very similar. Fourth albums appear to be released more quickly, but this likely reflects sample selection. We can only compute time-to-next-release conditional on there being a next release, and since most of the third albums in our sample were released near the end of the sample period, we only observe a fourth release if the time to release was short. This right truncation applies to the other albums as well, but we do not expect the problem to be as severe in those cases. Figure 3 shows a more complete picture of the heterogeneity in release lags across albums, including elapsed time between non-adjacent albums. The distribution of elapsed time between albums 1 and 2 is clearly very similar to the distribution between albums 2 and 3, but the right truncation is obvious in cases involving the release of album 4.

3.1 Sample selection issues

In addition to the obvious right truncation problem, our sample selection is likely to be biased toward artists who succeeded 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 been after 01 Jan 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 central empirical analysis of section 4. 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.

4 Empirical Analysis of the Backward Externality

In this section we first discuss our empirical strategy for estimating the backward externality. Our approach is taken from the treatment literature [8] and exploits exogenous variation in release times of albums. We investigate the plausibility of the exogeneity assumption by estimating a duration model. After reporting the results of this model, we report the estimates of the treatment effects. Finally, we examine the robustness of the results to unobserved heterogeneity and the proportional specificiation.

4.1 Empirical Strategy

A new album release by an artist is interpreted as the "treatment." Releasing a new album is an irreversible act: once treated, the catalogue albums remain treated. We will follow the impact of a new release on sales of catalogue albums for S periods, and refer to this number as the length of the treatment "window."(In the models estimated below, S is 39 weeks: 13 pre- and 26 post-treatment.) Each new release is analyzed as a separate treatment episode. For each episode, time is measured in terms of the number of periods since the last new album was released.

Without loss of generality, we focus on the first treatment episode. Let y_{it}^0 denote log of album 1 sales of artist i in period t without treatment and let y_{it}^s denote log of album 1 sales in period t when artist i is in the s^{th} period of treatment. Our objective is to estimate the average treatment

[8] See Wooldridge [5]

effect (ATE) for each period of the treatment window:

$$ATE_s = E[y_i^s - y_i^0],\ s = 1, .., S.$$

Notice that, 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 nonlinear: a new release has a larger impact on total sales of catalogue 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. We address this issue in section 4.4 below.

The main challenge in estimating the ATE's for artist i is that, in each period, we observe only one outcome for that artist. The observed outcome for artist i in period t is

$$y_{it} = y_{it}^0 + \sum_{s=1}^{S} w_{i,t-s+1}[y_{it}^s - y_{it}^0],$$

where $w_{i,t-s+1}$ is an indicator variable that is equal to one if artist i enters treatment in period $t-s+1$ and zero otherwise. The probability model generating outcomes for artist i in period t is given by:

$$y_{it}^s = \mu^s + \phi(t) + \nu_i + v_{it}^s,\ s = 0, 1, 2, .., S.$$

Here μ^s is the mean of the distribution of log sales in time period t for artists in the s^{th} period of treatment, $\phi(t)$ is a function that that captures the common, downward trend in an artist sales, ν_i measures the impact of unobserved artist characteristics on sales in every period, and v_{it}^s is the idiosyncratic shock to album 1 sales of artist i when she is in treatment period s at time period t. The artist-specific effect does not vary across the treatment window. Substituting the above equations, the observed outcome for artist i in period t is given by

$$\begin{aligned} y_{it} &= \mu^0 + \phi(t) + \nu_i + v_{it}^0 \\ & \sum_{s=1}^{S} w_{i,t-s+1}[(\mu^s - \mu^0)] + (v_{it}^s - v_{it}^0)]. \end{aligned}$$

The ATE for treatment period s is the difference in means, $\mu^s - \mu^0$.

We use the outcomes of a random sample of artists as proxies for the missing sales data on each artist. For each artist, t indexes time since the debut album's release, not calendar time. 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: our 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.[9]

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). 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 absorbed in the album fixed effects, differences in the shapes of albums' time paths are necessarily part of the error (ϵ). Including separate indicators for successive weeks of treatment allows us to check whether the 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.

[9] 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.

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 externality reflects consumers' responses to these pre-release marketing campaigns. 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.

We estimate the regression in 1 separately for each of six "treatments": the impact of the second, third, and fourth releases on sales of the first album; the impact of the third and fourth releases on sales of the second album; and the impact of the fourth release on sales of the third album. In constructing the samples for estimating the regression in 1 we impose several restrictions. First, in the first treatment, we exclude the first six 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 six 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. For later treatments, we restrict the sample to begin six months after the release of the previous album. So, for example, in estimating the impact of album 4 on album 2, we use album 2's sales beginning six months after the release of album 3. In essence, we want to consider the impacts of the various releases separately, in each case taking the flow of sales just prior to the new release as given. 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 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 two regressions estimating the impact of album 3 on albums 1 and 2.

4.2 Exogeneity of time to release

The regressions described above yield consistent, unbiased estimates of the treatment effect if the treatment indicators 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, we need the treatment to be random across artists. This is a strong but not implausible assumption. Selection effects would arise if the distributor uses a release rule that depends upon the flow of sales. For example, suppose the rule is to release the second album if sales of album 1 reaches a certain (random) threshold. In this case, the probability of treatment is more likely following a bad sales shock. However, it is not clear that such rules make sense from a decision-theoretic perspective. If the new release does not cannibalize sales of catalogue titles, then the distributor (and artist) should try to release a new album as soon as possible.

In general, variation in the elapsed time between releases (shown for our sample in Figure 3) could have several sources. 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. Live tours and other engagements may delay the production of an artist's next album. However, it is possible release times vary for strategic reasons. In some cases artists may delay the production of new music as a bargaining tactic. (Most recording contracts grant the record company an option to produce future albums by the artist under the same terms as applied to previous albums. Artists' leverage for negotiating more favorable terms in these contracts is to threaten to withhold new music.) Even when artists cooperate, it is possible that their record companies try to strategically time releases of new albums. Anecdotally, some record company executives seem to believe that new releases cannibalize sales of previous albums,[10] in which case strategic incentives come into play. The seasonality in album release dates shown in Table 2 suggests that at least some discretion is exercised by the record companies.

To better understand the sources of variation in release times, we estimate Cox proportional hazard models with various album and artist characteristics included as covariates. Table 3 presents the results. Somewhat surprisingly, the time it takes to release an artist's second album is essentially independent of the success of the first album (as measured by first six months' sales) after conditioning on the other covariates. On the other hand, third albums are released more quickly when the second album was a success. (The coefficient on first six months' sales is positive and statistically significant, meaning that successful second albums have a higher hazard rate for the release of album 3.) Release lags are significantly shorter for Country artists, and the coefficients on "years since 1993" reveal a general time trend toward longer lags between second and third (and third and fourth) albums.

[10] This belief is demonstrably false in our sample, as shown below.

Although some of the anecdotal evidence suggests release times should generally be regarded as endogenous, 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 for. Note that even if release times depended on the success of previous albums, this would not pose a threat to our empirical strategy: the inclusion of album fixed effects absorbs any differences in sales levels across albums. The validity of our approach depends more critically on whether release times are related to the *shape* of the previous album's sales path. Because our regression only controls for the average rate of decline in album sales, our estimates of the treatment effect will be biased if deviations from that average systematically affect release times. The numbers in Table 3 suggest this kind of dependence may be important for the releases of albums 3 and 4, where it is clear that albums with faster decline rates are associated with longer delays before the subsequent release. In principle, this could mean that albums tend to get treated when their sales are high relative to predictions based on average decline rates, in which case our estimates of the new release's impact would be biased upward. However, we suspect this effect is relatively unimportant in a quantitative sense. In section 4.4 we address the issue directly using an alternative regression specification that controls for heterogeneity in decline rates, and the results confirm that our main findings are robust.

4.3 Results

Table 4 presents estimates of equation (1), 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 can be interpreted as approximate percentage changes in sales resulting from the new release.

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

Figure 4 shows the estimated effects graphically: each panel plots the estimated coefficients (and 95% confidence bands) for one 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 externality is between albums 2 and 1, with estimates ranging between 40-55%. The externalities for the remaining pairs of albums are smaller, ranging mostly between 15-35%; however, overall the magnitudes are remarkably similar between album pairs.

For most album pairs (and especially the impact of album 2 on album 1) the estimated effects are remarkably persistent: the externalities do not appear to be transitory. The only apparent exception is the impact of album 3 on album 2, for which the coefficients decline somewhat at the end of the treatment window. 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 economic significance of the estimates of proportional changes reported in Table 4 is not immediately obvious since sales of the albums decline steadily with time. Table 5 shows the implied increases in total sales over the 39-week treatment window. The increases are reported at the 10^{th}, 50^{th}, and 90^{th} percentiles of the sales distribution at the time of the respective new release. (In each case we compute numbers at the median release period.) We report these percentiles because the level of sales across albums is extremely heterogeneous, so the proportional effects listed in Table 4 imply quite different increases in absolute sales for different artists. During the sample period, the retail price of albums is typically between $10-$18.

The implied effects appear to be very large for hit albums—for example, a 90^{th} percentile first album sells over 49,000 extra units in the months following the release of the second album—and the numbers are economically meaningful: multiplied by a CD price of $16, the estimated effects imply revenue increases of over $750,000 for 90^{th} percentile albums. On the other hand, the implied sales increases are inconsequential for the very small albums: even a large proportional increase doesn't mean much when applied to relatively low sales flows. The numbers in the table make clear that the biggest externalities are between adjacent albums—e.g., albums 2 and 1, 3 and 2, or 4 and 3. Interestingly, the implied effects are still economically large at the release of the fourth album.

4.4 Robustness

4.4.1 Allowing for heterogeneity in decline rates

The critical assumption underlying our empirical approach is that release times are exogenous with respect to the sales dynamics of previous albums. However, the proportional hazard models discussed in section 4.2 indicate that albums with slowly declining sales are treated earlier on average: new albums are released sooner when the current album has "long legs." If the timing of the treatment is endogenously related to whether current sales flows are low or high relative to overall averages, then our estimates of the externality could be biased.

In order to address this concern, we 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 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.

Table 6 lists estimated coefficients from this model for the six album pairs considered in Table 4, and Figure 5 shows the estimated patterns graphically. (The figure plots the cumulative impact implied by the estimated weekly coefficients.) The implied effects are qualitatively and quantitatively very similar to those reported in Table 4. Given this reassuring comparison, we are confident that our main results are driven by real effects, not by subtle correlations between current sales flows and the timing of new releases.

4.4.2 Proportionality assumption

[To be written.]

5 Source of the Backward Externality

Our data clearly indicate that sales flows of an artist's existing albums increase when the artist releases a new title, and these increases are large and persistent. Why does this happen? One possible explanation is prices: if retailers routinely lower the prices of catalog albums by artists with new releases, then the increase in sales that we have been calling the backward externality is in fact not an externality; it simply reflects down-sloping demand curves. We do not have any publicly available price data for the albums in our sample, so we cannot directly rule out this explanation; however, we have various reasons to doubt that the backward externalities are driven by discounts. First, retail CD prices are remarkably rigid in general. Variation in price across titles and over time is limited, with almost all albums being simply classified into a few pricing tiers. Discounts are occasionally "pushed down" to the retail level by distributors, but 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 does not seem to be systematically related to new releases by the same artist.

Ideally we'd like to know about discounting practices during our sample period, but we do not have meaningful data on prices from that period.[11] However, one thing we can check is whether current retailers engage in discounting practices that could generate backward externalities. To this end, we identified 20 artists who released new titles in February 2005, and downloaded Amazon.com's prices on each of these artist's bestselling CDs. (We excluded singles, imports, and EP's, as well as a few high-priced special-edition anthologies.) To assemble a control group, we matched each of the 20 artists to two "similar artists" (where the similarity was suggested by Amazon.com), and again obtained prices for the bestselling CDs of each of these artists. This yielded 20 artists with recent releases, 40 artists without recent releases, and a total of 264 albums combined.

Table 7 compares the prices for three groups of albums: new releases, catalog titles by artists with new releases, and catalog titles by artists without new releases. The table clearly indicates that although new releases tend to be discounted, the price distributions for the other two groups are quite similar. A Kolmogorov-Smirnov test fails to reject that the prices of catalog titles have the same distribution regardless of whether the artist has a new release, and a simple t-test fails to reject that the mean prices are equal. Another interesting feature of Amazon.com is that two

[11] The Federal Trade Commission has collected price data for our sample period in order to study the proposed merger between Time-Warner and EMI in 2000. We are currently working to have those data released.

prices are typically posted: the "list price" and the Amazon.com price. If we define discounting to mean offering a price below list price, we find that 91 of 160 (57%) catalog titles without new releases were discounted, as compared to 47 of 84 (56%) catalog titles by artists who did have new releases. These numbers are obviously based on a small sample, and they describe pricing patterns over three years after the end of our sample period. However, we consider it highly unlikely that retail pricing policies in the 1990's were somehow more fluid or sophisticated than in 2005, and we are confident that the backward externality estimated in sections 4.3 and 4.4.1 are not merely responses to discounting.

Two explanations we consider more plausible are preference complementarities and preference learning. In the appendix we outline detailed models of both; here we briefly summarize the ideas. The preference complementarity explanation posits that the externalities result from direct complementarities in the consumption of albums by the same artist: i.e., owning one album by an artist endogenously increases the utility from purchasing other albums by that artist. This is essentially a story about "fans": 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. In this framework, the backward externality occurs because new releases make previous albums more attractive to consumers whose utility from both albums is high enough to induce purchase, but whose utility from the previous album alone was just below the purchase threshold.

The preference learning explanation assumes no direct complementarities in consumption. Instead, the backward externality results from consumers updating their beliefs about a previous album's quality when they hear tracks from the new release. In this model, consumers are either not aware of the previous album or too uncertain about whether they will like it to buy it. The release of a new album generates new signals (e.g., new tracks are played on the radio), so new releases enhance awareness and those that lead to favorable updates of beliefs about past albums will generate a backward externality.

Importantly, note that both of these models predict persistent backward externalities, since in both cases the new release directly changes the probability of wanting to purchase the catalog album. Other information-based explanations (e.g., based on artist discovery) may predict transitory demand increases, as the "buzz" and increased airplay around the time of a new release could accelerate the arrival of consumers at the store. However, if these consumers are ones who would have eventually purchased the catalog title anyway (i.e., even if the new album were never released), then the increases could not be as persistent as the ones we see in our data. In order to generate permanent increases in demand for past albums, a new release must induce purchases by customers who would not have otherwise purchased.

One feature of the estimates that favors a preference learning explanation is the presence of pre-release effects. Record companies commonly concentrate their marketing and promotion efforts in the weeks leading up to a new release, and often release singles to be played on the radio several weeks before the release of the full album. In the preference learning model, pre-release promotion naturally leads some consumers to update their beliefs about catalog albums, inducing some to go out and buy the old album even before the new one is released. Conversely, the preference complementarity model, strictly interpreted, suggests that the benefits of joint consumption can't be obtained until both albums are available.

A sharper distinction in the predictions of the two models relates to the timing of artists' successes. In the preference learning story, the impact of a new release is largest when it leads to favorable updates of beliefs about the quality of past albums. This suggests that the backward externality will be largest when the new release is a hit, while the previous releases were not. Moreover, once an artist has a hit album, there may be few potential consumers who have not heard the catalog albums and even for those that have, there may be little left to learn about one's preferences for that artist, since the airplay and publicity associated with a hit yield sufficient signals about the style and quality of the artist's music. Therefore, if an artist's first album was a hit, we may not expect a large backward externality from the second release even if the second release is an even bigger hit.

The predictions of a preference complementarity model depend upon the distribution of preferences in the population. Since sales of any single album, even hits, are a relatively small fraction of the number of potential consumers, preferences for an album are highly skewed, with purchases coming from consumers in the right tail of the distribution. Under this assumption, the backward externality is much larger when the catalog album and new release are both hits than when they are both duds. The intuition is that there is more mass on the marginal types for hits than for duds. In the latter case, the marginal types are in the tails of the joint distribution where there is very little probablity mass, particularly when the consumer's preferences are correlated across albums. The asymmetric cases are also easily ranked. If album 2 is a hit, then lots of consumers who did not buy album 1 in period 1 will be willing to buy it together with album 2 due to the complementarity effect. If album 2 is a dud, then few consumers buy album 2 and, as a result, change their minds about the utility of album 1. The more difficult comparison is two hits versus a dud followed by a hit. The ranking depends upon functional forms, but our analysis suggests that the largest externalities are likely to occur when the catalog title and new release are both big hits.

In order to check these distinctions empirically, we focus on the impact of album 2's release and divide the artists into four groups: those for whom (i) both albums 1 and 2 were hits, (ii) album 1 was a hit but album 2 was not, (iii) album 1 was not a hit but album 2 was a hit, and (iv) neither album was a hit. For the purposes of the exercise, we define a hit to be any album that sold over 14,000 units in a week at some point in its lifetime; this is approximately the threshold for breaking into the top 100 on Billboard's weekly Top 200 chart. We then re-estimate the regression model (1) for each of these subgroups separately. (In each case, the control group of as-yet-untreated artists consists only of artists within the same subgroup.) Table 8 shows the results. The striking result is that the externality is largest when the new release is a hit and the previous release was not. Given that our regression model estimates proportional effects, this is perhaps not surprising: the same absolute increase in sales implies a much larger percentage change when applied to a smaller sales flow. However, the difference in the backward externality across subgroups is dramatic even when considered in terms of absolute sales (as in the bottom panel of the table). These patterns are perfectly consistent with the preference learning model, but not with preference complementarities. The former model predicts a large backward externality when the new release generates favorable updates—i.e., when the new album is better than the previous one;[12] the latter predicts that the backward externality will be largest when both albums are hits. We conclude that the most likely source of the backward externalities is an information mechanism in which new releases help consumers learn their preferences about previously released albums. We discuss some important implications of this finding in the concluding section below.

6 Conclusion

We find that the backward externality of a new release on sales of catalogue albums are on average positive, substantial, and permanent. The effect of the second album on sales of the debut album in the 39 week period beginning 13 weeks prior to the release week is approximately 40-50% per week, and the percentages show no indication of declining with time. The effects of the third and fourth albums on catalogue albums are smaller but still significant, approximately 20-30%

[12]The result that externalities are larger for "non-hits on non-hits" than for "hits on hits" is not clearly predicted by the preference learning model. However, even within the category represented in column 4 of the table, the effects are largest when the second album is better than the first (in the sense of having a higher sales peak).

per week. The implied total increases in sales of catalogue albums are economically significant, particularly on the latest catalogue album.

The incidence of the backward externality is strongly related to the relative success of the catalogue and new release. Indeed, the backward externality appears to be largely a hit-driven phenomenon, and perhaps even a first-hit phenomenon. When album 2 is the first hit, album 1 sales begin increasing two months prior to album 2's release, and the percentage increases in the post-release period range from 60–100%. By contrast, when album 1 is a hit, the impact of album 2 on album 1 sales is relatively weak. Pre-release sales are negligible, and post-release increases range from 6-15% per week. The results of the third album on the second are similar. These results suggest that the source of the backward externality is preference learning rather than preference complementarity.

An important implication of our results is that the backward externality represents a barrier to entry. An entrant can bid for new releases of artists whose careers are on the decline but not on new releases by artists whose careers are on the rise. This is not a good situation for the entrant, particularly if the incumbent label is better able to forecast the artist's peak. The entrant could try to purchase the rights to the artist's catalogue from the incumbent label. But the incumbent label is unlikely to sell, knowing that doing so facilitates the entry of a firm that will become its competitor in the market for music and in the market for new artists.

A second important implication of our empirical results is that the skewness in the distribution of sales across albums and artists may reflect the process through which consumers learn about their preferences. Consumers learn about albums by listening to them on radio and at concerts, and by talking to friends about albums that they have heard and concerts that they have attended. Consequently, if consumers only buy what they hear, and they only hear what others buy, then consumers are likely to buy the same albums, and the distribution of album sales is endogenous. We intend to explore this issue in a subsequent paper.

References

[1] Becker, G., Michael Grossman, Kevin M. Murphy, "An Empirical Analysis of Cigarette Addiction", American Economic Review, Vol. 84, No. 3, June, 1994, pp. 396-418.

[2] Gandal, N., Sarit Markovich, Michael Riordan, "Ain't it "Suite": Strategic Bundling in the PC Office Software Market", mimeo, April, 2004.

[3] Gentzkow, M., 2004, "Valuing New Goods in a Model with Complementarity: Online Newspapers", mimeo, 2004.

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

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

Appendix

In this appendix we present two choice models which generate revenues functions that are consistent with Assumption 1. They also offer two possible explanations for the backward externality.

6.1 Preference Learning

Let θ_{ij} denote consumer i's standalone utility of album j, $j = 1, 2$. We assume that it consists of a component that is common to all consumers and an idiosyncratic component:

$$\theta_{ij} = \theta_j \eta_{ij},$$

where $\theta_j(i_j)$ is a measure of the quality of album j. It may depend upon the label's investment in the album. Let F denote the joint distribution of (η_{i1}, η_{i2}) with mean one and covariance matrix Σ. We should think of the η_{ij}'s as being highly skewed with most of the mass near zero (e.g., exponential). Let F_j denote the marginal distribution of η_{ij}. The probability that consumer i is informed about album j is given by $q_j(\theta_j, i_j)$. It is an increasing function of the quality of album j and marketing in album j. We assume that if consumer i is informed about album 2, then he is also informed about album 1. Similarly if consumer i is informed about album 1 then he is also informed about album 2. Consumer i's purchase rule is to buy album j iff $\theta_{ij} > p \Leftrightarrow \eta_{ij} > \frac{p}{\theta_j}$. The multiplicative structure implies that an increase in the quality of the album has the same effect as a decrease in price.

The probability that consumer i purchases album 1 in the first period is

$$r(\theta_1, i_1) = q_1(\theta_1, i_1)[1 - F_1(p/\theta_1(i_1))].$$

The probability that consumer i purchases album 1 in the second period is

$$R_1(\theta_1, \theta_2, i_1, i_2) = q_2(\theta_2, i_2)(1 - q_1(\theta_1, i_1))[1 - F_1(p/\theta_1(i_1))]$$

and the probability of purchasing album 2 in period 2 is

$$R_2(\theta_1, \theta_2, i_1, i_2) = (q_1(\theta_1, i_1) + q_2(\theta_2, i_2)(1 - q_1(\theta_1, i_1))[1 - F_2(p/\theta_2(i_2))].$$

This model has the following properties. Consumers more likely to learn about album 1 in period 1, and more likely to purchase it, if it is a hit than a dud. Hence, r is increasing in θ_1 and i_1. Consumers who do not learn about album 1 in period 1 are more likely to discover the album if album 2 is a hit than a dud. Hence, R_1 is increasing in θ_2 and i_2. The effect of θ_1 and i_1 on R_1 is ambigous. The fraction of uninformed consumers is lower on hits than duds, but the purchasing probability is higher. The former effect is likely to dominate for hits. Higher values of θ_1 and i_1 increases R_2 since more consumers are informed about the album. Finally, R_2 is increasing in θ_2 and i_2 since it raises the probability that consumers hear the album and, upon hearing it, purchase the album.

6.2 Preference Complementarity

As in the previous model, we assume that the standalone utility of album j is the product of a common component, θ_j, which measures the quality of the album, and an idiosnycratic component, η_{ij}, that measures the utility of the match between consumer i and album j. The quality of an album depends upon the label's investment in the album. Everyone knows their preferences for an album when it is released. Utility of consumer i in period 1 is given by

$$U_{i1} = \theta_{i1} - p$$

if he buys album 1 and zero otherwise. If he does not buy album 1 in period 1, then consumer i's utility in period 2 is given by

$$U_{i2}(d_1, d_2) = d_1\theta_{i1} + d_2\theta_{i2} + d_1 d_2 \rho\theta_{i1}\theta_{i2} - d_1 p - d_2 p$$

where d_j is a binary variable that is equal to 1 if album j is purchased (in period 2) and zero otherwise and $\rho \geq 0$. If he has bought album 1 in period 1, then consumer i's utility in period 2 is consumer i's utility is given by

$$U_{i2}(d_2) = \theta_{i1} + d_2\theta_{i2} + d_2\rho\theta_{i1}\theta_{i2} - d_2 p.$$

We assume that consumers buy album 1 in period 1 iff $\theta_{i1} > p$.

Consider first the period 2 choice problem of consumers who did not buy album 1 in period 1, that is, $\theta_{i1} < p \Leftrightarrow \eta_{i1} > \frac{p}{\theta_1}$. Since the utility of album 1 can exceed p if and only if consumer i buys album 2, it follows that either consumer i buys only album 1 or both albums or none. We distinguish two regions. When the standalone utilities of album 2 is less than p, the choice is between buying both albums or none. In this case, the indifference boundary is given by

$$\theta_{i2} = \frac{2p - \theta_{i1}}{(1 + \rho\theta_{i1})} \Leftrightarrow \eta_{i2} = \frac{2p - \theta_1\eta_{i1}}{\theta_2(1 + \rho\theta_1\eta_{i1})}.$$

When the standalone utility of album 2 exceeds p, then the choice is between buying album 2 only or buying both. In this case, the indifference boundary is given by

$$\theta_{i2} = \frac{p - \theta_{i1}}{\rho\theta_{i1}} \Leftrightarrow \eta_{i2} = \frac{\frac{p}{\theta_1} - \eta_{i1}}{\rho\theta_2\eta_{i1}}.$$

Note that the two boundaries intersect at

$$\eta_{i1} = \frac{p}{\theta_1(1 + \rho p)}, \eta_{i2} = \frac{p}{\theta_2}.$$

Thus, the boundary for the region in which consumer i buys both albums conditional on not buying album 1 in period 1 is continuous. Let $\gamma(\eta_{i1})$ denote the boundary. Plotting the boundary in (η_{i1}, η_{i2}) space, it is easily verified that it shifts toward the origin with increases in θ_1 or θ_2. (The boundary for purchasing album 1 also shifts towards the origin.)

Consider next the period 2 choice problem of consumers who have bought album in period 1, that is, $\theta_{i1} > p$. In this case, the choice is between buying album 2 or not, and the indifference boundary is given by

$$\eta_{i2} = \alpha(\eta_{i1}) \equiv \frac{p - \theta_1 \eta_{i1}}{\theta_2(1 + \rho\theta_1\eta_{i1})}.$$

The revenue functions are as follows:

$$r(\theta_1, i_1) = [1 - F_1(p/\theta_1)],$$

$$R_1(\theta_1, \theta_2, i_1, i_2) = \int_0^{p/\theta_1} \int_{\gamma(\eta_{i1})}^{\infty} dF(\eta_{i1}, \eta_{i2}),$$

$$R_2(\theta_1, \theta_2, i_1, i_2) = \int_0^{p/\theta_1} \int_{\gamma(\eta_{i1})}^{\infty} dF(\eta_{i1}, \eta_{i2}) + \int_{p/\theta_1}^{\infty} \int_{\alpha(\eta_{i1})}^{\infty} dF(\eta_{i1}, \eta_{i2}).$$

It is not difficult to check that these functions satisfy Assumption 1.

6.3 Quantitative Predictions

Do the two models have different predictions about the variation in the magnitude of the backward externality? We distinguish two states for each album, hit (H) and not hit (N), and rank the magnitude of the backward externality across the four possible pairs of states.

The ranking for the preference learning model seems fairly clear. If high quality creates lots of exposure, then the externality is largest when album 1 is not a hit, and album 2 is a hit. It is smallest when album 1 is a hit and album 2 is not a hit. To summarize: $NH > NN > HH > HN$. This is the ordering that we see in the data.

The ranking for the preference complementarity model depends upon the distribution of idiosyncratic shocks. Suppose they are perfectly correlated, $\eta_{i1} = \eta_{i2} = \eta_i$, so that all of the probability mass is concentrated on the diagonal. Let $\overline{\eta}$ denote the intersection of the diagonal with $\gamma(\eta)$, i.e., the solution to the equation, $\eta = \gamma(\eta)$. An intersection always exists but it may not occur in the region of externality, that is, $\overline{\eta}$ may exceed p/θ_1. This situation arises if album 1 is hit and album 2 is a dud. In this case, there is no externality. The intuition is that the marginal types on album 1 have very low values of the idiosyncratic component on album 2, and hence do not want to buy both albums. When $\overline{\eta} < \frac{p}{\theta_1}$, then

$$R_1 = F(p/\theta_1) - F(\overline{\eta}(\theta_1, \theta_2)).$$

Increases in θ_2 lowers $\overline{\eta}$ and increases revenues. An increase in θ_1 has an ambiguous effect since it decreases both endpoints of the segment of the diagonal that lies in the region of backward externality. The magnitude of the externality depends primarily upon the location of the interval

$[\overline{\eta}, p/\theta_1]$ in the support of the distribution. The fact that the fraction of consumers that buy any album is relatively small suggests that the distribution of values for a specific album is highly skewed. In this case, if both albums are duds, then the interval lies in the tail of the distribution where the mass of marginal types is low. Hence, the difference in probability is likely to be quite small. If both albums are hits, then the interval lies much closer to zero where the mass of marginal types is much higher. The difference in probability is likely to be much higher. If album 1 is a dud and album 2 is a hit, then the marginal types on album 1 like album 2 a lot and will want to buy both. However, the mass of marginal types remains low since the interval lies in the tail of the distribution. In summary, when distribution of values are highly skewed, the magnitudes of the backward externality can be ranked as follows: $HH > NH > NN > HN = 0$.

The same basic intuition carries over to the case where the idiosyncratic shocks are independent. Reversing the order of integration, the revenue function can be expressed as

$$R_1(\theta_1, \theta_2, i_1, i_2) = \int_0^{p/\theta_1} [1 - F_2(\gamma(\eta_{i1}))] dF_1(\eta_{i1}).$$

Recall that $\gamma(\eta_{i1})$ is decreasing in η_{i1}, ranging from ∞ to $\frac{p}{(1+\rho p)\theta_2}$. Thus, the integrand is an increasing function of η_{i1}, ranging from 0 to $1 - F_2(p/(1+\rho)\theta_2)$. If album 2 is a dud, then the marginal types for album 2 are in the far right tail of the distribution of F_2 where there is little mass. Hence, the integrand is close to zero, and magnitude of the backward externality is small, regardless of the quality of album 1. If album 2 is a hit, then the marginal types for album 2 shift left where there is more probability mass. Hence, the integrand can take on much larger values, causing the magnitude of the backward externality to increase. Increases in the quality of album 1 causes the interval of integration to decrease but it also shifts $\gamma(\eta_{i1})$ into the part of the distribution with more mass. Under the assumption that the latter effect dominates, the ordering is as follows: $HH \simeq NH > NN > HN > 0$.

[Use the exponential distribution to verify these rankings.]

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
4	74	08jan2000	73	19apr1998	09feb2000	28oct2001
overall						
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
4	74	316,335	579,869	6,137	87,898	912,078
overall	962	358,362	836,366	8,938	68,059	976,853
First 4 weeks / First year:						
album 1	355	.121	.111	.0161	.0846	.265
2	355	.263	.137	.0855	.263	.441
3	178	.305	.131	.134	.305	.5
4	74	.312	.144	.119	.294	.523
overall	962	.222	.15	.0341	.208	.431
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
4	74	5.42	16.6	0	0	19
overall	962	15.8	35.3	0	1	44
Weeks between releases:						
1 & 2	355	114	53.5	58	107	179
2 & 3	178	111	46.7	58	104	169
3 & 4	74	93.1	36.8	50	88	154

Table 2: Seasonality in release dates

	Percent of releases occurring				
Month	Album 1 (n=355)	Album 2 (n=355)	Album 3 (n=178)	Album 4 (n=74)	Overall (n=962)
Jan	3.94	3.10	3.37	2.70	3.43
Feb	8.17	4.23	3.93	1.35	5.41
Mar	13.24	9.58	11.80	10.81	11.43
Apr	9.01	8.45	8.99	6.76	8.63
May	11.83	9.01	7.30	8.11	9.67
Jun	7.61	12.68	6.74	14.86	9.88
Jul	8.45	9.01	10.11	10.81	9.15
Aug	11.55	9.58	10.67	12.16	10.71
Sep	7.32	11.27	11.80	14.86	10.19
Oct	12.39	10.70	16.29	6.76	12.06
Nov	5.92	11.83	6.74	5.41	8.21
Dec	0.56	0.56	2.25	5.41	1.25

Table 3: Determinants of elapsed time between releases

	Elapsed time between:		
	1 and 2	2 and 3	3 and 4
First six months' sales	-0.005	0.028	0.010
	(0.014)	(0.010)	(0.018)
Decline rate (prev. album)	0.127	0.341	0.437
	(0.075)	(0.105)	(0.174)
Country	0.858	0.556	0.235
	(0.171)	(0.212)	(0.332)
Rap	-0.036	-0.156	0.754
	(0.138)	(0.201)	(0.291)
Years since 1993	0.090	-0.309	-0.369
	(0.031)	(0.048)	(0.080)
N	355	355	178
log likelihood	-1719.788	-920.496	-334.709

Estimated coefficients from Cox proportional hazard models, with standard errors in parentheses. A positive coefficient means that an increase in the corresponding covariate is associated with an increased hazard rate (i.e., less time between releases). The estimation accounts for right-censoring when the next album was not released before the end of our sample period.

Table 4: Estimated Effects of New Releases on Sales of Catalog Albums

Month (relative to release date)	Album pair					
	2→1	3→1	4→1	3→2	4→2	4→3
t=-13	-0.004	0.015	0.016	0.043	-0.010	-0.004
	(0.017)	(0.027)	(0.045)	(0.025)	(0.039)	(0.037)
t=-12	0.020	0.026	0.001	0.010	0.025	0.048
	(0.022)	(0.033)	(0.053)	(0.032)	(0.046)	(0.047)
t=-11	0.035	0.016	0.035	-0.047	-0.016	0.013
	(0.026)	(0.036)	(0.056)	(0.036)	(0.048)	(0.051)
t=-10	0.051	0.027	0.102	0.022	0.038	0.042
	(0.028)	(0.038)	(0.056)	(0.038)	(0.049)	(0.053)
t=-9	0.054	0.017	0.066	0.050	0.055	0.061
	(0.030)	(0.039)	(0.056)	(0.040)	(0.050)	(0.055)
t=-8	0.065	0.029	0.059	0.055	0.026	0.075
	(0.031)	(0.040)	(0.058)	(0.041)	(0.050)	(0.055)
t=-7	0.107	0.027	0.024	0.071	-0.001	0.034
	(0.032)	(0.040)	(0.058)	(0.042)	(0.051)	(0.056)
t=-6	0.129	0.068	0.121	0.087	0.052	0.055
	(0.033)	(0.041)	(0.058)	(0.043)	(0.051)	(0.057)
t=-5	0.184	0.115	0.165	0.119	0.067	0.063
	(0.034)	(0.040)	(0.059)	(0.043)	(0.051)	(0.057)
t=-4	0.239	0.130	0.172	0.175	0.135	0.103
	(0.034)	(0.041)	(0.058)	(0.044)	(0.051)	(0.057)
t=-3	0.279	0.165	0.146	0.241	0.135	0.127
	(0.035)	(0.041)	(0.057)	(0.044)	(0.052)	(0.058)
t=-2	0.324	0.227	0.206	0.256	0.182	0.220
	(0.035)	(0.041)	(0.058)	(0.044)	(0.052)	(0.058)
t=-1	0.397	0.272	0.307	0.330	0.258	0.206
	(0.036)	(0.041)	(0.060)	(0.045)	(0.052)	(0.059)
t=0	0.444	0.336	0.386	0.358	0.306	0.248
	(0.036)	(0.041)	(0.060)	(0.045)	(0.052)	(0.059)
t=1	0.422	0.249	0.334	0.306	0.255	0.263
	(0.036)	(0.042)	(0.059)	(0.045)	(0.052)	(0.059)
t=2	0.417	0.297	0.296	0.305	0.199	0.162
	(0.037)	(0.042)	(0.060)	(0.045)	(0.052)	(0.059)
t=3	0.398	0.293	0.284	0.280	0.216	0.219
	(0.037)	(0.042)	(0.059)	(0.045)	(0.053)	(0.060)
t=4	0.430	0.283	0.324	0.263	0.281	0.132
	(0.037)	(0.042)	(0.060)	(0.046)	(0.052)	(0.060)
t=5	0.429	0.270	0.305	0.249	0.203	0.227
	(0.037)	(0.042)	(0.059)	(0.046)	(0.052)	(0.060)
t=6	0.464	0.277	0.347	0.271	0.253	0.194
	(0.037)	(0.042)	(0.059)	(0.046)	(0.052)	(0.060)
t=7	0.481	0.285	0.308	0.258	0.239	0.156
	(0.038)	(0.042)	(0.060)	(0.046)	(0.052)	(0.061)
t=8	0.489	0.312	0.278	0.270	0.216	0.167
	(0.038)	(0.042)	(0.061)	(0.046)	(0.053)	(0.061)

(continued next page)

Table 4: (continued)

Month (relative to release date)	Album pair					
	2→1	3→1	4→1	3→2	4→2	4→3
t=9	0.446	0.310	0.270	0.263	0.222	0.170
	(0.038)	(0.042)	(0.060)	(0.047)	(0.053)	(0.061)
t=10	0.463	0.339	0.235	0.308	0.224	0.159
	(0.038)	(0.043)	(0.061)	(0.047)	(0.053)	(0.061)
t=11	0.465	0.322	0.195	0.336	0.133	0.132
	(0.038)	(0.043)	(0.060)	(0.047)	(0.053)	(0.062)
t=12	0.470	0.358	0.280	0.331	0.154	0.107
	(0.038)	(0.043)	(0.060)	(0.047)	(0.053)	(0.062)
t=13	0.506	0.307	0.314	0.286	0.156	0.144
	(0.039)	(0.043)	(0.062)	(0.047)	(0.053)	(0.063)
t=14	0.539	0.312	0.309	0.299	0.241	0.113
	(0.039)	(0.043)	(0.061)	(0.048)	(0.054)	(0.063)
t=15	0.507	0.230	0.309	0.257	0.231	0.187
	(0.039)	(0.043)	(0.061)	(0.048)	(0.054)	(0.063)
t=16	0.495	0.212	0.249	0.244	0.151	0.160
	(0.039)	(0.043)	(0.063)	(0.048)	(0.054)	(0.064)
t=17	0.509	0.217	0.308	0.216	0.193	0.116
	(0.039)	(0.043)	(0.062)	(0.048)	(0.054)	(0.064)
t=18	0.511	0.236	0.261	0.223	0.232	0.040
	(0.040)	(0.044)	(0.063)	(0.049)	(0.054)	(0.064)
t=19	0.524	0.180	0.421	0.165	0.247	0.204
	(0.040)	(0.043)	(0.063)	(0.049)	(0.055)	(0.065)
t=20	0.540	0.247	0.360	0.179	0.260	0.197
	(0.040)	(0.044)	(0.062)	(0.049)	(0.055)	(0.065)
t=21	0.492	0.241	0.286	0.188	0.192	0.204
	(0.040)	(0.044)	(0.062)	(0.049)	(0.055)	(0.066)
t=22	0.524	0.246	0.295	0.178	0.202	0.240
	(0.040)	(0.044)	(0.062)	(0.050)	(0.055)	(0.066)
t=23	0.538	0.218	0.279	0.193	0.140	0.132
	(0.041)	(0.044)	(0.062)	(0.050)	(0.055)	(0.066)
t=24	0.544	0.225	0.286	0.165	0.244	0.192
	(0.041)	(0.044)	(0.062)	(0.050)	(0.056)	(0.066)
t=25	0.558	0.243	0.296	0.146	0.208	0.143
	(0.041)	(0.044)	(0.063)	(0.050)	(0.056)	(0.067)
# albums	339	162	66	173	70	74
# observations	38,996	18,528	6,589	19,841	7,002	7,465
$\hat{\rho}$	.838	.694	.547	.781	.566	.688

Estimates of the regression described in equation 1, 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 4→2 lists the estimated effects of album 4'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 5: Implied Total Increases in Sales

	Level of sales prior to new release (percentile)		
Album pair	0.10	0.50	0.90
2→1	907	5,296	49,203
3→1	139	1,290	10,974
4→1	74	711	8,432
3→2	446	2,455	26,850
4→2	102	1,071	8,759
4→3	314	2,003	26,220

Total increase in sales over the 39-week treatment window, as implied by the estimates in Table 4. So, for example, the release of the third album increases sales of the second album by 2,455 units if the second album's sales were at median levels prior to the new release.

Table 6: Estimated effects of new releases: first-differenced model

Month (relative to release date)	Album pair					
	2→1	3→1	4→1	3→2	4→2	4→3
t=-13	-0.019	0.020	0.032	0.041	0.018	0.011
	(0.016)	(0.026)	(0.044)	(0.023)	(0.040)	(0.037)
t=-12	0.009	0.020	-0.004	-0.035	0.048	0.046
	(0.016)	(0.027)	(0.045)	(0.023)	(0.041)	(0.038)
t=-11	0.008	-0.005	0.037	-0.058	-0.037	-0.050
	(0.016)	(0.027)	(0.048)	(0.023)	(0.041)	(0.038)
t=-10	0.012	0.019	0.074	0.069	0.057	0.060
	(0.016)	(0.027)	(0.045)	(0.023)	(0.041)	(0.038)
t=-9	0.001	-0.004	-0.034	0.029	0.016	0.030
	(0.016)	(0.027)	(0.045)	(0.023)	(0.041)	(0.038)
t=-8	0.011	0.021	0.001	0.012	-0.022	0.032
	(0.016)	(0.027)	(0.046)	(0.023)	(0.042)	(0.038)
t=-7	0.044	0.009	-0.019	0.026	-0.009	-0.013
	(0.016)	(0.027)	(0.048)	(0.023)	(0.042)	(0.036)
t=-6	0.025	0.054	0.107	0.027	0.069	0.030
	(0.016)	(0.027)	(0.046)	(0.023)	(0.042)	(0.036)
t=-5	0.058	0.049	0.041	0.037	0.024	0.041
	(0.016)	(0.026)	(0.045)	(0.023)	(0.042)	(0.038)
t=-4	0.058	0.025	0.013	0.060	0.066	0.043
	(0.016)	(0.026)	(0.045)	(0.023)	(0.042)	(0.038)
t=-3	0.044	0.044	-0.025	0.072	-0.005	0.027
	(0.016)	(0.027)	(0.045)	(0.023)	(0.042)	(0.038)
t=-2	0.054	0.077	0.064	0.019	0.049	0.105
	(0.016)	(0.027)	(0.045)	(0.023)	(0.042)	(0.038)
t=-1	0.082	0.061	0.103	0.087	0.083	-0.012
	(0.016)	(0.026)	(0.046)	(0.023)	(0.042)	(0.038)
t=0	0.055	0.076	0.089	0.038	0.060	0.057
	(0.016)	(0.026)	(0.048)	(0.023)	(0.042)	(0.038)
t=1	-0.014	-0.081	-0.046	-0.041	-0.046	0.011
	(0.016)	(0.026)	(0.046)	(0.023)	(0.042)	(0.038)
t=2	-0.003	0.046	-0.030	0.001	-0.057	-0.096
	(0.016)	(0.027)	(0.046)	(0.024)	(0.042)	(0.038)
t=3	-0.022	-0.008	-0.017	-0.024	0.018	0.061
	(0.016)	(0.027)	(0.046)	(0.023)	(0.042)	(0.038)
t=4	0.021	-0.012	0.035	-0.017	0.074	-0.070
	(0.016)	(0.026)	(0.045)	(0.023)	(0.042)	(0.036)
t=5	-0.015	-0.010	-0.003	-0.022	-0.071	0.101
	(0.016)	(0.026)	(0.046)	(0.023)	(0.041)	(0.036)
t=6	0.016	-0.006	0.030	0.015	0.049	-0.024
	(0.016)	(0.027)	(0.046)	(0.024)	(0.041)	(0.036)
t=7	0.002	-0.006	-0.042	-0.022	-0.027	-0.041
	(0.016)	(0.027)	(0.046)	(0.024)	(0.041)	(0.036)
t=8	-0.004	0.019	-0.027	0.005	-0.025	0.010
	(0.016)	(0.027)	(0.048)	(0.024)	(0.041)	(0.037)

(continued next page)

Table 6: (continued)

Month (relative to release date)	Album pair					
	2→1	3→1	4→1	3→2	4→2	4→3
t=9	-0.047	-0.017	-0.007	-0.018	0.001	-0.011
	(0.016)	(0.026)	(0.048)	(0.023)	(0.041)	(0.037)
t=10	0.018	0.007	-0.016	0.029	0.007	0.017
	(0.016)	(0.026)	(0.048)	(0.023)	(0.041)	(0.038)
t=11	0.003	-0.036	-0.014	0.014	-0.072	-0.010
	(0.016)	(0.027)	(0.046)	(0.023)	(0.041)	(0.038)
t=12	-0.002	0.016	0.089	-0.027	0.028	-0.011
	(0.016)	(0.027)	(0.045)	(0.023)	(0.041)	(0.038)
t=13	0.027	-0.059	0.021	-0.051	-0.004	0.034
	(0.016)	(0.027)	(0.046)	(0.023)	(0.041)	(0.038)
t=14	0.026	0.005	-0.011	0.012	0.068	-0.030
	(0.016)	(0.027)	(0.046)	(0.024)	(0.041)	(0.038)
t=15	-0.033	-0.062	-0.011	-0.028	-0.025	0.053
	(0.016)	(0.027)	(0.046)	(0.024)	(0.041)	(0.038)
t=16	-0.008	-0.003	-0.048	-0.005	-0.074	-0.043
	(0.016)	(0.027)	(0.047)	(0.024)	(0.042)	(0.038)
t=17	0.023	0.024	0.062	-0.019	0.050	-0.029
	(0.016)	(0.027)	(0.047)	(0.023)	(0.042)	(0.036)
t=18	0.004	0.026	-0.034	0.008	0.059	-0.033
	(0.016)	(0.027)	(0.047)	(0.023)	(0.042)	(0.036)
t=19	0.012	-0.056	0.168	-0.060	0.027	0.170
	(0.016)	(0.026)	(0.049)	(0.024)	(0.042)	(0.038)
t=20	0.012	0.069	-0.050	0.013	0.026	-0.014
	(0.016)	(0.026)	(0.046)	(0.024)	(0.042)	(0.036)
t=21	-0.046	-0.014	-0.062	0.005	-0.066	0.010
	(0.016)	(0.026)	(0.046)	(0.023)	(0.040)	(0.036)
t=22	0.033	-0.000	0.017	-0.012	0.020	0.038
	(0.016)	(0.026)	(0.046)	(0.023)	(0.040)	(0.036)
t=23	0.013	-0.027	-0.007	0.016	-0.061	-0.100
	(0.016)	(0.026)	(0.046)	(0.023)	(0.042)	(0.036)
t=24	-0.003	0.009	-0.004	-0.031	0.094	0.082
	(0.016)	(0.026)	(0.046)	(0.023)	(0.042)	(0.038)
t=25	0.005	0.019	0.003	-0.017	-0.043	-0.035
	(0.016)	(0.027)	(0.046)	(0.023)	(0.042)	(0.038)
# albums	339	162	66	173	70	74
# observations	38,923	18,476	6,554	19,806	7,995	7,454

Estimates of the regression described in equation 2, with $\Delta ln(sales)$ as the dependent variable. 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.

Table 7: Prices for catalog titles at the time of a new release

			Percentiles			
	N	mean	std. dev.	.10	.50	.90
New release titles	20	13.24	1.50	11.49	13.49	14.49
Catalog titles by artists with new releases	84	13.77	2.28	10.99	13.98	16.98
Catalog titles by artists without new releases	160	14.04	2.50	10.99	13.99	17.98

Based on prices at Amazon.com in February 2005.

Table 8: Externalities and hits: Albums 1 and 2

Album 1 and 2 outcomes:	Hit, Hit	Hit, Not	Not, Hit	Not, Not
N	59	41	51	188
Median # weeks to release 2	110	127	109	100
Weekly sales prior to treatment period:				
10th percentile	439	90	65	31
50th percentile	2,313	247	284	142
90th percentile	28,115	1,354	1,682	690
Estimated total change in sales:				
10th percentile	57	-33	2,301	753
50th percentile	298	-90	10,113	3,451
90th percentile	3,626	-493	60,000	16,768

Hits are defined as albums that at some point sold over 14,000 units nationally in one week. Albums that didn't clear this threshold are the "Not" albums (i.e., not hits). The estimated total change in sales reflects the increase over the 39-week treatment window, as in Table 5.

Figure 1: Album sales paths for two examples

Figure 2: Typical sales paths

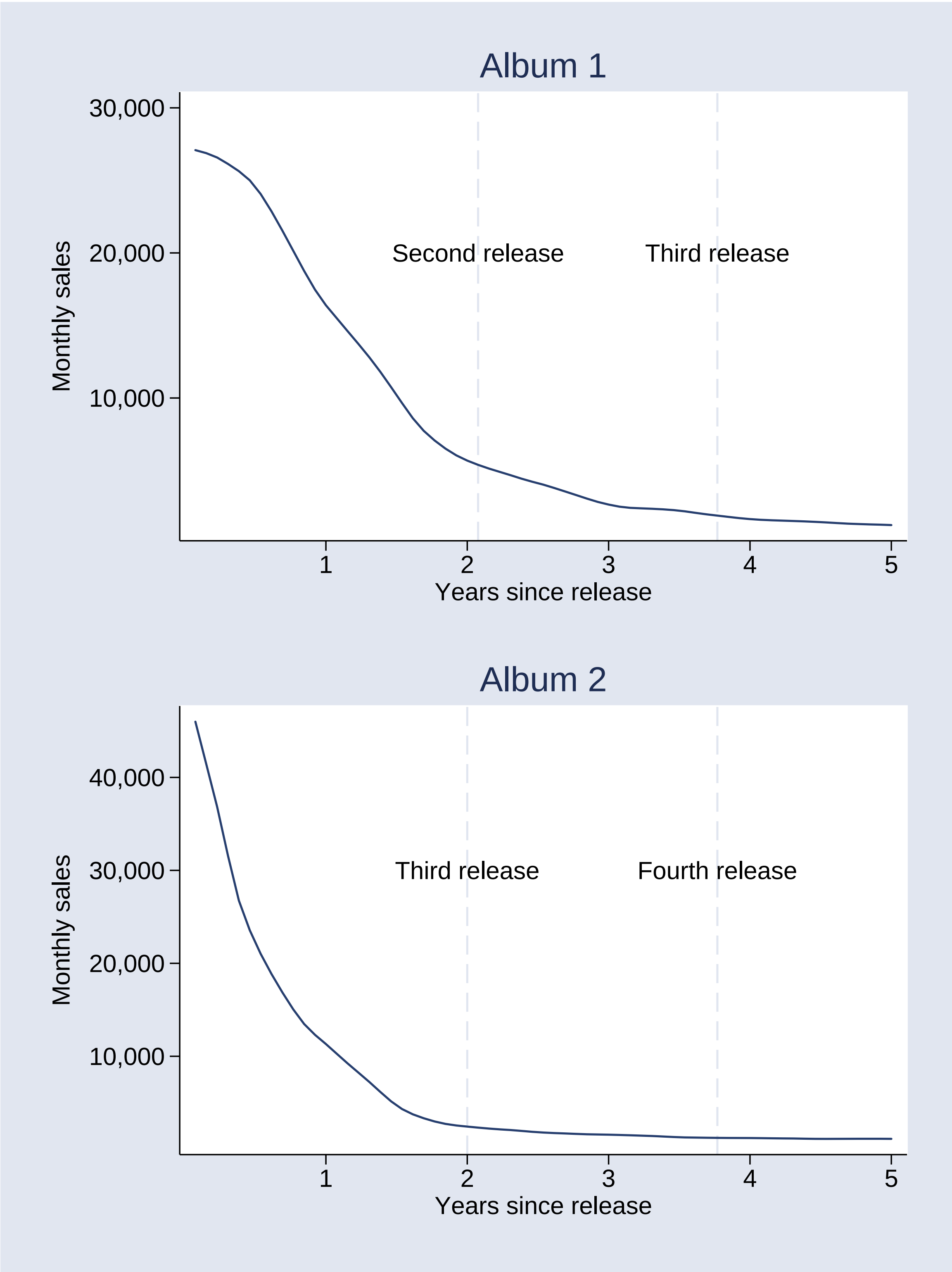

Figure 3: Distributions of Elapsed Time Between Releases

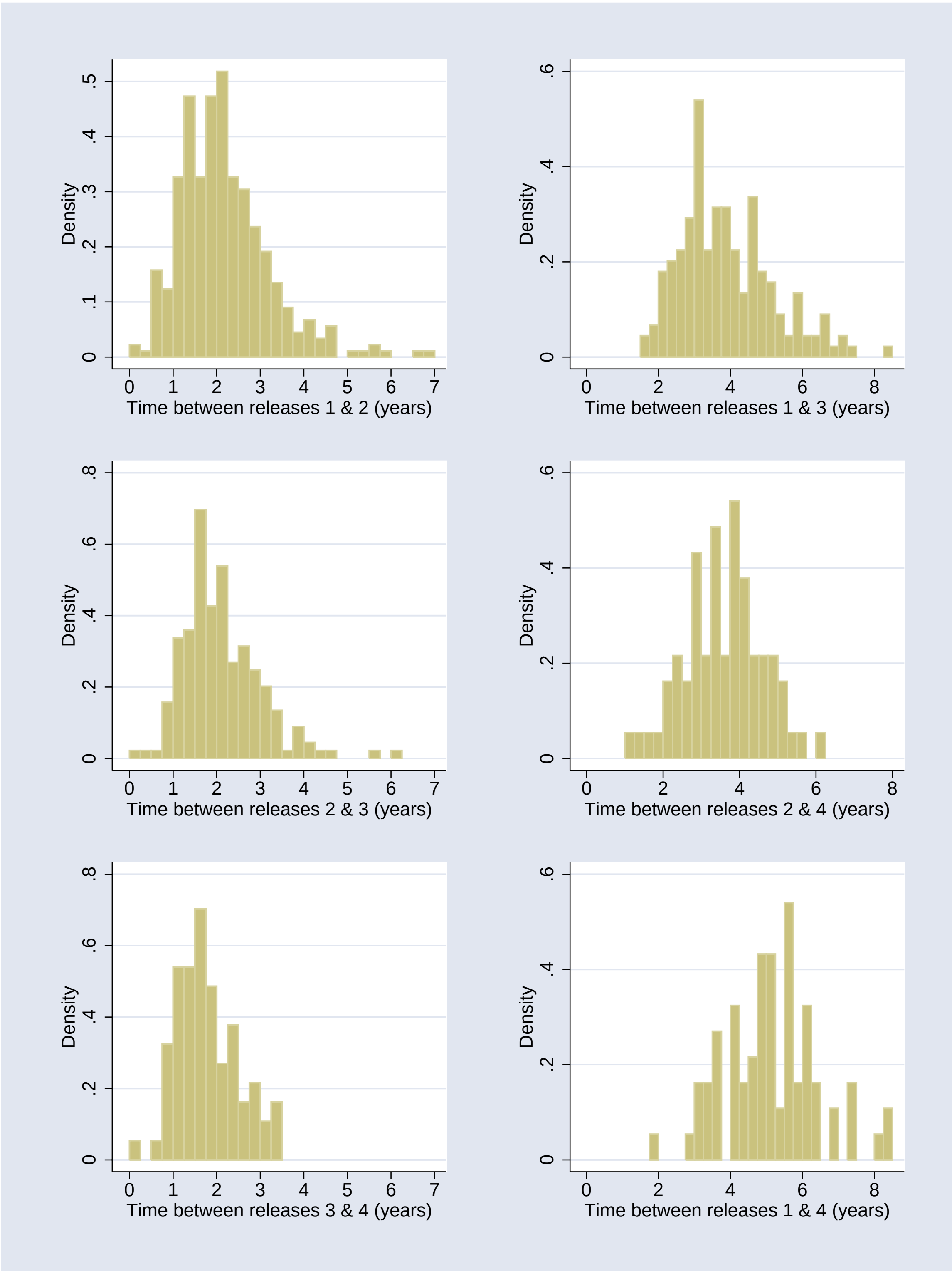

Figure 4: Time patterns of backward externalities

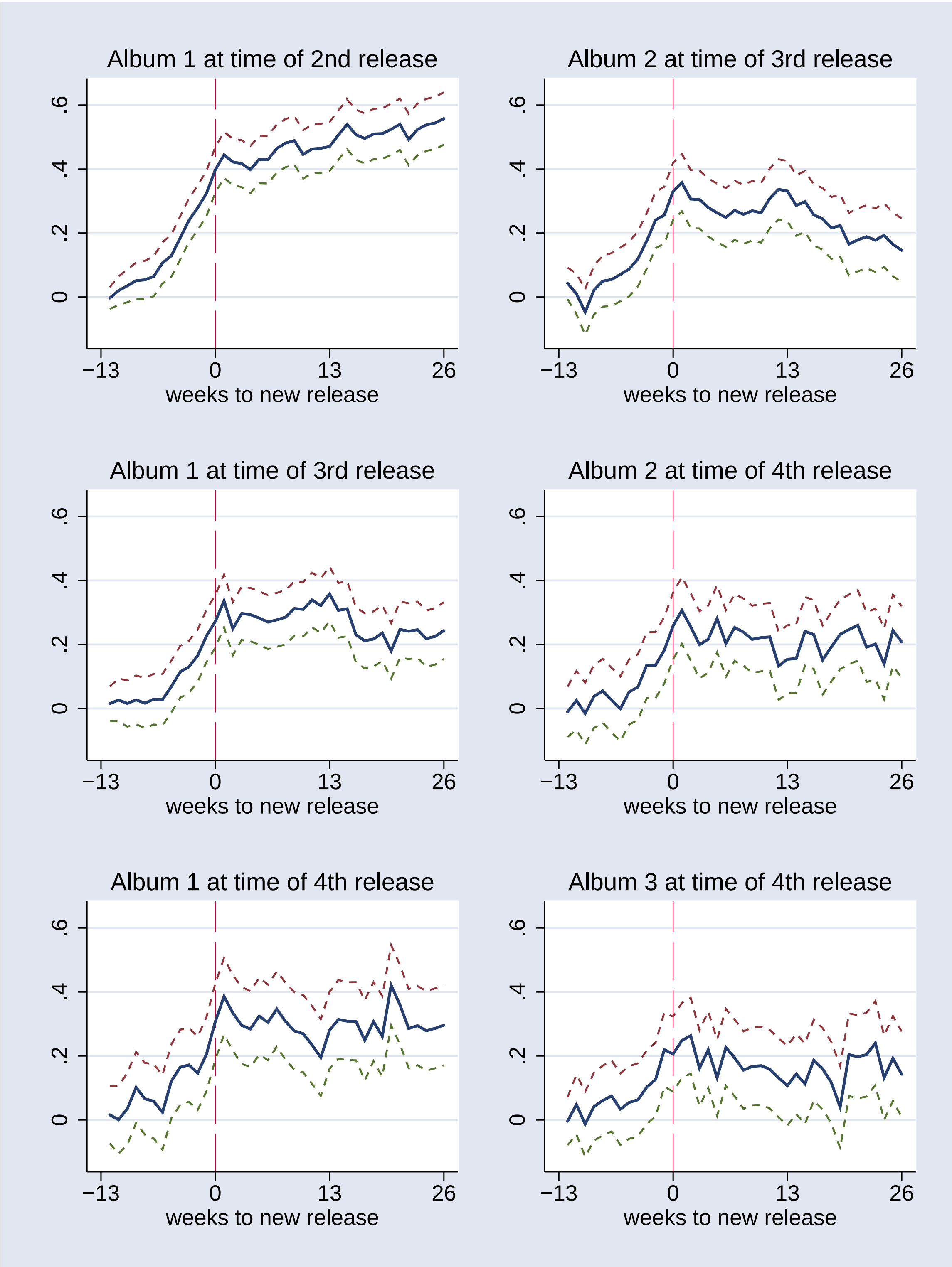

Figure 5: Time patterns of backward externalities: first-differences model