WILEY

Bestseller Lists and Product Variety
Author(s): Alan T. Sorensen
Source: *The Journal of Industrial Economics*, Vol. 55, No. 4 (Dec., 2007), pp. 715-738
Published by: Wiley
Stable URL: http://www.jstor.org/stable/4622404
Accessed: 06/03/2014 16:11

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Wiley is collaborating with JSTOR to digitize, preserve and extend access to *The Journal of Industrial Economics.*

http://www.jstor.org

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

THE JOURNAL OF INDUSTRIAL ECONOMICS 0022-1821
Volume LV December 2007 No. 4

BESTSELLER LISTS AND PRODUCT VARIETY*

ALAN T. SORENSEN†

This paper uses detailed weekly data on sales of hardcover fiction books to evaluate the impact of the *New York Times* bestseller list on sales and product variety. In order to circumvent the obvious problem of simultaneity of sales and bestseller status, the analysis exploits time lags and accidental omissions in the construction of the list. The empirical results indicate that appearing on the list leads to a modest increase in sales for the average book, and that the effect is more dramatic for bestsellers by debut authors. The paper discusses how the additional concentration of demand on top-selling books could lead to a reduction in the privately optimal number of books to publish. However, the data suggest the opposite is true: the market expansion effect of bestseller lists appears to dominate any business stealing from non-bestselling titles.

I. INTRODUCTION

THE PERCEIVED IMPORTANCE OF BESTSELLER LISTS is a salient feature of multimedia industries. Weekly sales rankings for books of various genres are published in at least 40 different newspapers across the U.S., and making the list seems to be a benchmark of success for authors. In the movie industry, box office rankings are watched closely by movie studios and widely reported in television and print media. Sales of music CDs are tracked and ranked by Billboard Magazine, whose weekly charts are prominently displayed in most retail music stores.

Ostensibly, the purpose of bestseller lists is to simply report consumers' purchases. However, there are a number of reasons why the conspicuous publication of sales rankings may directly influence consumer behavior (in addition to merely reflecting it). Bestseller status may serve as a signal of quality: for example, bookstore patrons who are unfamiliar with a particular author may nevertheless buy her current bestseller, thinking that its popularity reflects other buyers' (favorable) information about the book's quality.[1] Publicized sales rankings would also directly affect consumer

*I am thankful to the Hoover Institution, where much of this research was conducted, and to Nielsen BookScan for providing the book sales data. The research has benefitted from helpful conversations with Jim King, Phillip Leslie, and Joel Sobel, among many others. Scott Rasmussen provided excellent research assistance. Any errors are mine.

†Author's affiliations: Graduate School of Business, Stanford University and NBER, 518 Memorial Way, Stanford, California, 94305-5015, U.S.A.
e-mail: asorensen@stanford.edu.

[1] There is an extensive literature on quality-signaling when products' qualities are uncertain. See, for example, (Milgrom and Roberts [1986]).

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics. Published by Blackwell Publishing Ltd. 9600 Garsington Road. Oxford OX4 2DQ. UK. and 350 Main Street. Malden. MA 02148. USA.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

behavior in the presence of social effects, with bestseller lists serving as a form of coordinating mechanism.[2] For example, teenagers who want to listen to music that is 'hot' can look to the Billboard charts to find out what is popular, and people may favor movies at the top of the box-office charts because they want to be conversant in popular culture. In the specific case of books, bestseller status also triggers additional promotional activity by retailers.

For the same reasons that bestseller lists may directly affect consumers' purchase decisions, they may also cause sales to be more highly concentrated on the few bestselling products. This, in turn, could influence product variety: if the additional sales accruing to bestsellers as a direct consequence of the publication of the list come at the expense of non-bestselling products, the optimal number of products to offer may decrease (relative to what would have been optimal in the absence of a bestseller list). For example, if publicized box office rankings cause ticket sales to be more concentrated on blockbusters, they may also make it unprofitable to incur the fixed costs of producing a film whose popularity is expected to be only marginal.[3]

This paper examines these issues in the context of the book publishing industry, looking specifically at the impact of the *New York Times* bestseller list on sales of hardcover fiction titles. The empirical analysis addresses two questions. First, does being listed as a *New York Times* bestseller *cause* an increase in sales? Second, does the influence of the bestseller list also affect the number of books that are published, and if so, in which direction? Obvious simultaneity problems make answering the first question a nontrivial empirical exercise; however, subtleties in the construction and timing of the *New York Times* list can be exploited to identify its impact. The results suggest a modest increase in sales for the typical bestseller when it first appears on the list, and the increase appears to reflect an informational effect rather than a promotional effect: the effects are concentrated in the first week a book appears on the list, and they are most substantial for new authors. Regarding the second question, the impact of bestseller lists on product variety is theoretically ambiguous, since it depends on whether market expansion or business-stealing effects dominate. Although the data are less than ideal for addressing this question, I present indirect evidence suggesting the business-stealing effects of bestseller lists are unimportant: if anything, bestseller lists appear to increase sales for both bestsellers *and* non-bestsellers in similar genres.

Although this paper is the first to explore the impact of bestseller lists on product variety, similar questions have been previously addressed in a number of contexts. Early theoretical work emphasized the tradeoff between

[2] See, e.g., (Becker and Murphy [2000], Banerjee [1992], and Vettas [1997]).
[3] This line of reasoning will be clarified in section II.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

quantity and diversity in the presence of scale economies (Dixit and Stiglitz [1977]) and the effects of market structure on product variety (Lancaster [1975]). Empirical studies of product variety have been undertaken for the radio broadcasting industry (Berry and Waldfogel [2001], and Sweeting [2005]), the music industry (Alexander [1997]), and for retail eyeglass sales (Watson [2003]), to name a few examples.

In the following section, I outline a basic theoretical framework for understanding how bestseller lists can influence the number of books that get published in equilibrium. Section 3 provides a brief description of the book industry and the dataset. The empirical analysis (Section 4) proceeds in two parts: first, the data are used to identify and quantify the direct impact of the bestseller list on sales; second, substitution patterns in the data are analyzed to determine the likely direction of the list's impact on product variety. Some broad implications of the results (as well as alternative interpretations) are discussed in the concluding section.

II. A SIMPLE THEORETICAL FRAMEWORK

Consider a highly simplified model in which a single publisher chooses how many manuscripts to publish.[4] There are K manuscripts under consideration. Printing and marketing a manuscript requires a fixed cost, F, in addition to the (constant) marginal printing cost c. Prior to publication, the manuscripts can be ranked in order of expected popularity, with r being the index of the r^{th}-best book among the K alternatives. The market price of a published book, p, is taken as given, and the post-publication price does not adjust to reflect a book's relative popularity.[5] The expected demand for the r^{th} best book is given by $D(K) \cdot \theta(r; K)$, where $D(K)$ can be interpreted as the level of aggregate demand for books (which may depend on how many are offered), and $\theta(r; K)$ is a function determining how aggregate demand is allocated among books depending on their relative popularity (e.g., we could have $\sum_{r=1}^{K} \theta(r; K) \equiv 1$). Only books with positive expected profits will be published, so the number of books will be the maximum K^* such that $(p - c)D(K^*)\theta(K^*; K^*) > F$, as illustrated in Figure 1.

Now consider the potential impact of publishing a bestseller list, so that consumers observe the sales ranks of the top L books. Suppose that the

[4] The model could also apply to movie studios' decisions about how many films to produce, or to record labels' decisions about how many artists to sign.

[5] This assumption would be absurd in most contexts, but in this case it is at least descriptive of a curious practice in multimedia markets. Prices of books, movies, and CD's almost never reflect the popularity of the individual products. (Clerides [2002]) provides direct evidence and a thorough discussion of pricing issues in the book industry.) Typically, 'price points' for books and CDs are determined before they are marketed, and subsequent adjustments are extremely infrequent. Movie ticket prices are even more rigid: in the summer of 2002, for example, it cost the same amount to see 'Chicago' (which won the Oscar for best picture and was a box-office success) as 'Boat Trip' (which flopped at the box office and was universally ridiculed by critics).

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

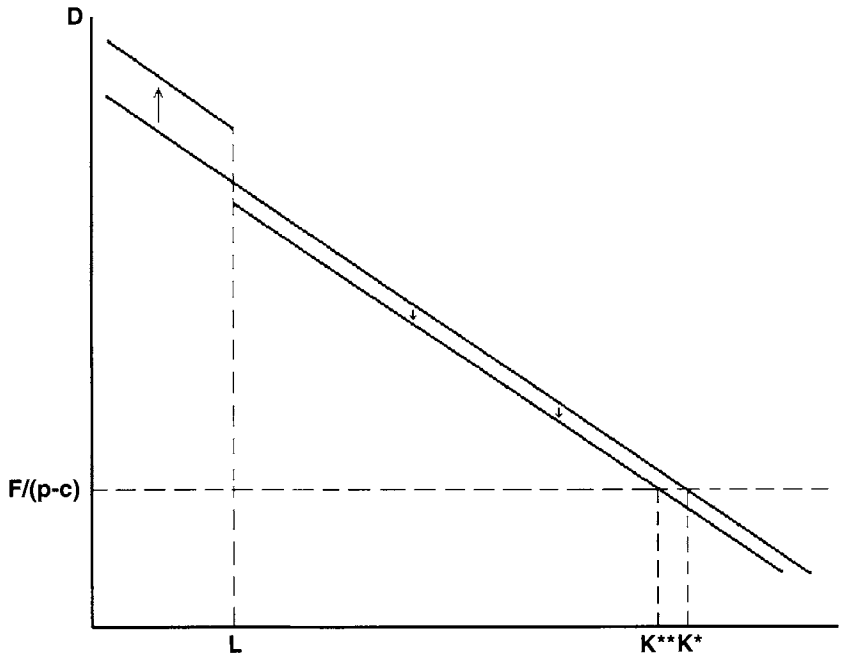


Figure 1
The publish/no-publish margin

aggregate consumer response to a bestseller list leads to an increased concentration of sales on bestsellers: letting $\theta(r; L,K)$ denote the expected market share of the r^{th}-ranked book among K alternatives when a bestseller list of length L is published, and $\theta(r; \emptyset,K)$ denote the market share when no list is published, we assume that $\theta(r; L,K) > (<)\theta(r; \emptyset,K)$ if $r \leq (>)L$. If bestseller lists have any direct impact on consumer behavior, the effect would almost certainly have this feature. The most obvious mechanism for this effect is informational: if consumers are uncertain about books' qualities, and they believe that at least some past purchasers had meaningful information about the books they purchased, then bestseller status would be a signal of quality. Alternatively, social effects may lead to higher demand for bestsellers: consumers may want to read what everyone else is reading in the interest of keeping up with what is popular—e.g., they don't want to be left out of the conversation when they go to the cocktail party. In the market for hardcover fiction books, an additional mechanism pushes sales toward bestsellers: retailers routinely discount bestsellers and position them prominently in their stores.

If the overall level of demand $D(K)$ is independent of any bestseller-list effects, then the list unambiguously reduces the number of books that can be

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

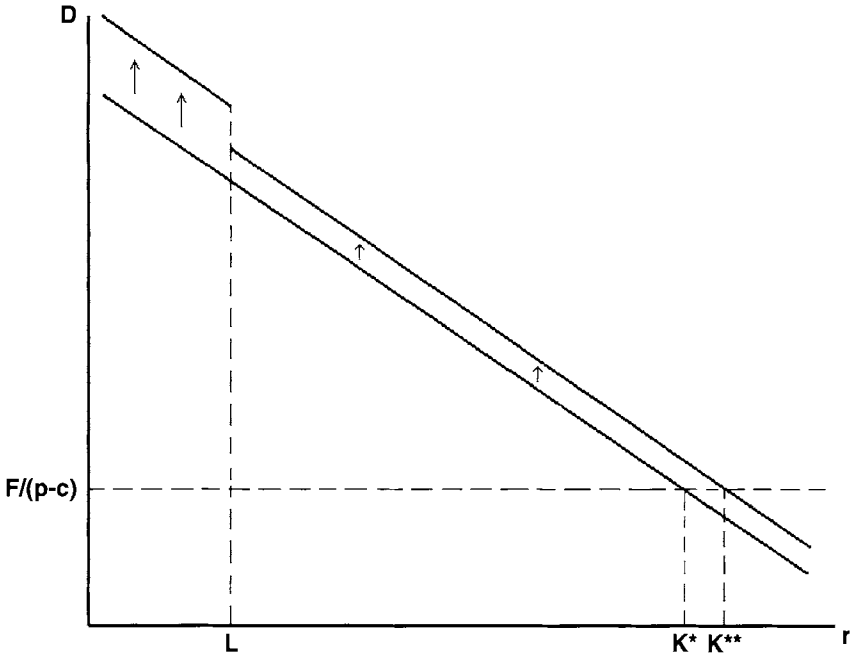


Figure 2
List may increase overall level of demand

profitably published. This is illustrated in Figure 1: the increased sales for the top L books come at the expense of the non-listed books, shifting the publish/no-publish margin to the left (from K^* to K^{**}). More realistically, however, the publication of a bestseller list could increase overall demand in addition to changing the allocation of demand across titles—i.e., the additional promotion and information about bestsellers could attract consumers that otherwise would not have purchased any book at all. In this case, the impact of bestseller lists on the publish/no-publish margin is ambiguous. (Figure 2 illustrates a case in which more books would get published in the presence of a list, even though the list leads to a relatively higher concentration of sales among bestsellers.)

This framework, while obviously oversimplified, illustrates the principal ideas underlying the empirical analyses to follow. Ideally, we want to examine sales data to see if indeed $\theta(r; L,K) > \theta(r; \emptyset,K)$ for $r \leq L$—that is, to see if bestseller lists *cause* an increase in demand for bestsellers relative to non-bestsellers. In order to say anything about whether bestseller lists affect the number of books that get published, we must then ask a much more subtle question of the data: how are sales of relatively unpopular (non-bestselling) books affected by the publication of bestseller lists? That is, how

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

does $\theta(r; L,K)$ compare to $\theta(r; \emptyset,K)$ for r very close to K?[6] Unfortunately (but not surprisingly), the available data are inadequate for answering this question directly. Instead, we will look for indirect evidence of substitution between bestselling and non-bestselling titles, which would suggest the potential for (and the likely direction of) product variety effects.[7]

III. BACKGROUND AND DATA

III (i). *The Book Industry*

In the U.S., the vast majority of books are produced by a small number of large publishing houses like Random House and Harper Collins.[8] The odds against a manuscript's being accepted by one of these publishing houses are long, especially in the case of fiction. Thirty per cent or fewer of available manuscripts in any given year are in print, and although ninety per cent of published books are nonfiction, seventy per cent of the manuscripts submitted to traditional publishers are fiction (Suzanne [1996]).[9] Most successful manuscripts are brokered to publishers by literary agents. These agents are typically reluctant to take on first-time authors, and their fees tend to be steep (around 15 per cent of authors' royalties). However, using an agent greatly increases the author's chances of success: unsolicited manuscripts (manuscripts received 'over the transom') are estimated to have fifteen thousand to one odds against acceptance (Greco [1997]). Manuscripts are sometimes sold to publishers by auction, but this method is the exception rather than the rule and is used primarily by established, brand-name authors.

The decision to extend a contract to an author is made only after review of the manuscript's quality and salability by several stages of editors. If a manuscript survives the review process, the publisher offers the author a contract granting royalty payments in exchange for exclusive marketing rights as long as the publisher keeps the book in print. Royalties average seven percent of the wholesale price on hardcover books by new authors,

[6] Note that the illustration in Figure 1 makes it appear that sales of the K^{th}-ranked book and the $L+1^{st}$-ranked book are equally affected, which need not be the case. For instance, it is quite plausible that the impact is a declining function of a book's rank. Moreover, only the effect on books at the margin is relevant for determining the number of books that get published.

[7] Because this model does not explicitly specify consumers' preferences, it says nothing about the welfare effects of a reduction in product variety. This paper will be deliberately agnostic on this point, since the welfare effects could plausibly go either way: on the one hand, fewer books could mean foregone surplus from titles that would have appealed to readers with diverse tastes; on the other hand, fewer books could mean that less time is wasted reading bad literature.

[8] For the first quarter of 2003, the top six publishing conglomerates accounted for over 80 per cent of unit sales in adult fiction.

[9] These numbers imply that 10 per cent of fiction manuscripts get published.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

and may increase once the book achieves a certain level of sales (Suzanne [1996]). Upon the delivery of a completed manuscript a large portion of expected royalties are given to the author in the form of an advance; many authors never receive additional payments because advances often exceed royalties earned from actual sales. Publishers retain a large share of royalties in escrow accounts to compensate for returned books; booksellers may return unsold books to the publishers for full price, and therefore return rates exceeding fifty per cent are common (Greco [1997]).

In spite of the difficulty that authors seem to face in getting their manuscripts published, the book industry generates an astonishing flow of new books each year. Across all categories (fiction and nonfiction) and all formats (hardcover, trade paper, and mass-market paper), over 100,000 titles were published in the year 2000 alone. In adult fiction, the number of new books published (called 'title output' within the industry) has increased dramatically over the past decade. The industry's trade publication, *Bowker Annual*, reports that title output for hardcover fiction more than doubled from 1,962 in 1990 to 4,250 in 2000. In contrast, the rate of increase was much more gradual prior to 1990. In fact, Bowker reports that the number of fiction titles in 1890 was over 1,100, so title output had less than doubled in the 100 years prior to 1990 (Bogart [2001]).

The dramatic increases in title output in the 1990's were roughly concomitant with an increase in the concentration of sales among bestsellers. From the mid-1980's to the mid-1990's, the share of total book sales represented by the top 30 sellers nearly doubled (Epstein [2001]). In 1994, over 70 percent of total fiction sales were accounted for by a mere five authors: John Grisham, Tom Clancy, Danielle Steel, Michael Crichton, and Stephen King (Greco [1997]).[10]

Publishers and authors employ a number of marketing strategies in their attempts to achieve the kind of success enjoyed by these top-selling authors. Publishers' marketing budgets may range between ten and twenty-five per cent of net sales (Cole [1999]), and authors are expected to appear publicly in promotion tours. Book reviews are highly sought after but difficult to obtain: tens of thousands of books are published each year in the United States, but the *New York Times* (for example) reviews only one per day. Publishers also may pay retail stores for shelf space or inclusion in promotional materials. Book marketers concentrate their efforts on creating a successful launch; retail stores may remove low-selling new releases from their shelves after as little as one week (Greco [1997]).

[10] Rosen [1981] provides a classic explanation for the presence of such 'superstars.' A critical piece of the argument is that the convexity of returns results from the imperfect substitutability of the products offered—i.e., reading several unremarkable books may not be a good substitute for reading a single great one.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

In spite of all the resources spent on marketing, one of the best kinds of publicity—appearing on a bestseller list—cannot be bought.[11] Bestseller lists have long played an important role in the book industry. Regular publication of bestseller lists began in 1895, when a literary magazine called *The Bookman* started printing a monthly list of the top six best-selling books. The *New York Times Book Review* began publishing its bestseller list as a regular feature in 1942. Although many other prominent lists now exist,[12] the *New York Times* list is generally considered the most influential in the industry (Korda [2001]).

III (ii). *Data*

The main dataset to be analyzed consists of weekly national sales for over 1,200 hardcover fiction titles that were released in 2001 or 2002. The sales data were provided by Nielsen BookScan, a market research firm that tracks book sales using scanner data from an almost-comprehensive panel of retail booksellers.[13] Additional information about the individual titles (such as the publication date, subject, and author information) was obtained from a variety of sources, including *Amazon.com* and a volunteer website called *Overbooked.org*. Table 1 reports summary information for the books in the data, broken into three subsamples.

The overall sample represents a relatively large fraction of the universe of hardcover fiction titles released in this time period, though it is likely to be somewhat skewed toward popular books.[14] Books that never sold more than 50 copies in a single week (nationwide) were dropped from the sample, since their weekly sales numbers appeared to be mostly noise. Also, some books are excluded from the empirical analyses if their release dates were difficult to determine.[15] In such cases, the number of books is reduced to 799. As will

[11] Occasionally an author has tested this proposition by purchasing numerous copies of his own book, in hopes of pushing it onto a bestseller list. None of these attempts has ever been truly successful; the typical outcome has been considerable embarrassment for the perpetrator once the scheme was uncovered.

[12] Most major U.S. newspapers publish their own local list (sometimes in addition to the *New York Times* list), and a number of national lists compete with the *New York Times* list for attention (e.g., *Publishers Weekly*, *Wall Street Journal*, *USA Today*, and *BookSense*).

[13] BookScan collects data through cooperative arrangements with virtually all the major bookstore chains, most major discount stores (like Costco), and most of the major online retailers (like *Amazon.com*). They claim to track at least 80 per cent of total sales.

[14] In constructing the set of candidate books to track, we had to first locate a book (and its ISBN number) in order to consider it. Obscure, slow-selling books are (by definition) harder to locate.

[15] Three sources of information were used in determining the exact week of release. For most titles, *Amazon.com* lists the exact day of release. If not, BookScan reports the month of release, and 'eyeballing' the data usually reveals an obvious release date. If for any reason the release date was not obvious—e.g., the *Amazon.com* and BookScan release dates didn't match, and/or the release date wasn't obvious from looking at the sales data—the title was excluded from the sample whenever the results might be sensitive to the accuracy of the release date.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

TABLE I
SUMMARY STATISTICS

All books with 26+ weeks of data (n = 1,217):					
	Min	Median	Max	Mean	Std. Dev.
List price	10.95	24.95	68.85	24.28	2.63
Release week*	1	57	83	49.64	24.18
New author	0	0	1	.15	.36
Sales, first 6 months	87	3,960	1,443,345	35,183	111,444
Max. one-week sales	50	486	360,133	6,903	24,453
Books with reliable release dates (n = 799):					
	Min	Median	Max	Mean	Std. Dev.
List price	10.95	24.95	60	24.44	2.38
Release week*	1	56	83	48.70	24.52
New author	0	0	1	.14	.35
Sales, first 6 months	87	8,063	1,443,345	51,396	134,445
Max. one-week sales	50	972	360,133	10,227	29,626
New York Times bestsellers (n = 205):					
	Min	Median	Max	Mean	Std. Dev.
List price	10.95	25	29.95	25.05	2.25
Release week*	1	40	83	41.33	23.89
New author	0	0	1	.05	.22
Sales, first 6 months	20,963	93,507	1,443,345	174,987	222,530
Max. one-week sales	2,108	18,438	360,133	36,047	50,058
Week first appeared	4	4	34	4.5	2.4
Weeks on list	1	4	30	5.8	5.3

*For release week, the first week of 2001 is week 1, and the first week of 2002 is week 53.

be shown in the next section, for nearly all books the vast majority of sales occur in the first 4–6 months, and subsequent sales are relatively uninformative. In light of this, only the first 26 weeks of sales are included in the sample for any given book.

Data on bestseller-list status come directly from the *New York Times*. The analysis focuses on the *New York Times* list because it is a nationally published list (and the sales data are national), and because it is almost universally regarded as the most influential list in the industry. A majority of retail booksellers (including online bookstores) have special sections devoted to *New York Times* bestsellers and offer price discounts on these titles, and authors are sometimes offered bonuses for every week their book appears on the *New York Times* list.

To construct its list, the *New York Times* surveys nearly 4,000 bookstores each week, in addition to a number of book wholesalers who serve additional types of booksellers (like supermarkets, newsstands, etc.) The reported sales figures from these respondents are then extrapolated to a nationally representative set of sales rankings using statistical weights. Because the *New York Times* list is constructed using sampling methods, it often makes 'mistakes'—i.e., books that should have made the list in a given week

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

(because their sales exceeded the sales of the book listed at rank 15) are sometimes omitted, and the ordering of listed books sometimes doesn't reflect the 'true' ranking of sales (as indicated by the BookScan data). Also, assembling the list takes time, so the printed bestseller list reflects rankings from three weeks prior. Both of these features of the *New York Times* list—the mistakes and the time lags—are critical in the empirical analysis of its impact on sales.

IV. EMPIRICAL ANALYSIS AND RESULTS

IV (i). *Skewness in Book Sales*

The most striking pattern in the data is that book sales are remarkably skewed in two important ways. First, the distribution of sales across books is heavily skewed. Figure 3 plots total sales in the first six months against sales rank for the top 100 books in the sample. Even when looking only at the top decile of books, the skewness of the distribution is striking. Of the 1,217 books for which at least 26 weeks were observed,[16] the top 12 (1 per cent) account for 25 per cent of total six-month sales, and the top 43 (3.5 per cent) account for 50 per cent. The 205 books that made it to the *New York Times* bestseller list account for 84 per cent of total sales in the sample. The most popular book in the sample, *Skipping Christmas* by John Grisham, sold more copies in its first three weeks than did the bottom 368 books in their first six months combined.

Second, book sales tend to be skewed with respect to time for any given title: that is, sales tend to be heavily concentrated in the first few weeks after a book's release. Of the 1,217 books in the sample for which 26 or more weeks are observed, 898 (73.8 per cent) hit their sales peak sometime in the first four weeks. The median 'peak week' is week 2. Somewhat surprisingly, this pattern also seems to hold for debut authors, for which one might expect gradual diffusion of information and therefore an S-shaped sales path. For new authors, 112 of 182 books (61.5 per cent) peak in the first 4 weeks, and the median peak week is 4.

This second form of skewness is important to keep in mind when interpreting the models and results in the following sections. The steady decay of a book's sales over time is the dominant pattern: with the exception of seasonal effects (e.g., Christmas), any other changes in a book's sales tend to be second-order relative to this decay trend.[17] Essentially, the sales paths

[16] Books for which the release date was questionable are *not* excluded here, since getting the release date right isn't critical for this exercise.

[17] Decay-like sales patterns are not unique to the book industry; for example, Einav [2003] reports similar patterns for movie box office sales and incorporates exponential decay directly into his empirical model.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

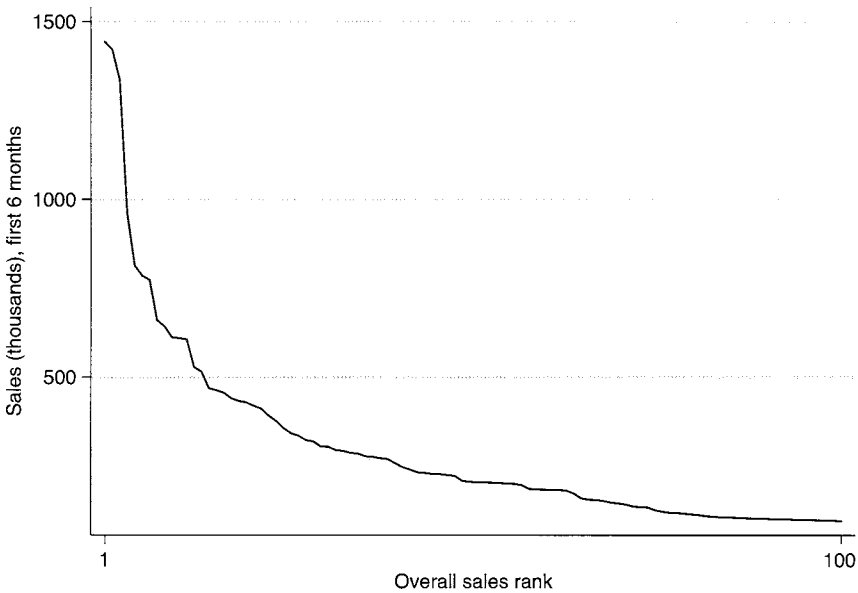


Figure 3
Skewness of book sales

typically resemble exponential decay patterns. Eyeballing the time path of sales for all the books in the sample, one rarely observes a book's sales 'take off' after it hits the bestseller list; if anything, making the list appears to temporarily slow the pace of decline.

IV (ii). *Do Bestseller Lists Directly Affect Sales?*

Theory clearly suggests that bestseller lists may do more than simply reflect consumer behavior: the lists may directly *influence* consumer behavior, so that a book's appearance on the bestseller list has an independent effect on its sales. However, measuring such an effect is a difficult empirical problem: the set of books that receive the 'treatment' of being listed as bestsellers is clearly not random, and a naive empirical approach would likely confuse the direction of causality. There is obviously a correlation between the level of sales and bestseller status (by the very definition of a bestseller list), but we cannot infer from this correlation that being listed as a bestseller *causes* higher sales.

However, given the available data and the subtleties in the construction of the *New York Times* list, there are at least two ways we can attempt to identify the list's direct influence. One strategy is to exploit the so-called 'mistakes' that are sometimes made in the list. As mentioned previously, the

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

TABLE 2
CHARACTERISTICS OF LISTED BOOKS VS. OMITTED BOOKS

Genre	Per cent of 'mistakes' listing this genre ($N = 75$)	Per cent of bestsellers listing this genre* ($N = 44$)	t-stat (p-value)
Literature/Fiction	51	64	−1.39 (0.17)
Mystery/Thriller	51	36	1.53 (0.13)
Romance	21	11	1.47 (0.14)
Science Fiction	12	7	0.96 (0.34)
Religion	1	0	1.00 (0.32)
Horror	3	2	0.13 (0.89)
Average list price:	24.98	24.54	1.09 (0.28)

*Only books that were ranked 13–15 when they first appeared on the bestseller list are included in the bestsellers group. Numbers are based on books' genres as listed on Amazon.com. Percentages add to more than 100 because books typically list more than one genre. Prices are publishers' list prices, from BookScan.

process used to generate the *New York Times* list is inexact. Although the list is by and large quite accurate when compared with the 'true' sales numbers available from BookScan, it is not uncommon for a bestselling book to be missed—i.e., a book may not appear on the list even though its sales exceeded the sales of listed books. In principle, these mistakes provide a means of identifying the effect of appearing on the list, by serving as an appropriate control group. (Comparing listed books to unlisted books is a bad experiment, since whether a book is listed is a nearly deterministic function of the dependent variable; but comparing listed books to books that *should* have been listed is, in principle, a valid experiment—as long as the 'mistakes' are random occurrences.)

During the years 2001–2002, there were 182 instances in which a hardcover fiction book was not listed as a *New York Times* bestseller when in fact it should have been,[18] representing 109 different books. (In several cases, there were multiple weeks in which a book should have been on the list but was not.) The majority of these (roughly 70%) were narrow misses: had the books been listed, they would have been ranked 13–15 on the list. In order to construct a fair comparison, I focus on two sets of books: those that were listed at rank 13, 14, or 15 when they first appeared on the *New York Times* list ($n = 44$), and those that should have appeared at 13, 14, or 15 when they were mistakenly omitted ($n = 75$). Table 2 summarizes some observable characteristics of the books in the two groups. If the omissions were not random, but rather an attempt by the *New York Times* at 'editorializing' the list, we might expect to see a different subject composition among the omitted books. The distribution of subjects and list prices appears to be mostly similar between the two groups, lending some confidence that the

[18] To be precise, I will say a book should have been listed if, for the week that was relevant for generating the list, the book's sales (according to BookScan) exceeded the sales of the book that was listed at #15.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

TABLE 3
COMPARISON OF SALES CHANGES: LISTED BOOKS VS. OMITTED BOOKS

	Mean	Std. Dev.
% Δ sales, first week appearing on bestseller list (n = 44)	−.076	.702
% Δ sales, week in which mistakenly omitted (n = 75)	−.227	.157

Difference = .151. One-tailed t-test for H_0: difference is zero: $t = 1.404$, $p = .084$.

omissions are indeed random mistakes. The only notable differences are that the genre 'Literature & Fiction' is more likely (and 'Romance' or 'Mystery & Thrillers' less likely) among books making the list than among omitted books, but these differences are not statistically significant. Moreover, when considering the overall composition of the list (not just positions 13–15), the proportions of Romance and Mystery novels match almost perfectly with the proportions among the omitted books, so it seems clear the omissions were not the result of any bias against particular genres.

Table 3 reports a comparison of sales for the two groups. For books that were published for the first time on the *New York Times* list, sales declined by an average of 7.6 per cent relative to the previous week.[19] For mistakenly omitted books, sales declined by an average of 22.7 per cent. Taken at face value, the difference implies that in the first week, being listed leads to 19 per cent more sales than would have otherwise occurred. However, the difference is statistically imprecise: for significance levels less than .08, we would fail to reject the null hypothesis that the list has no effect using a one-tailed test. The comparison reported in Table 3 does not control for any covariates, but doing so has very little effect on the estimate. Using simple linear regressions to control for seasonal effects and time-since-release effects yields estimates of the same approximate magnitude.

Given the availability of panel data on book sales, a second strategy for identifying the effect of bestseller lists is to use all the available data (not just 'mistake' books) to measure the week-by-week changes in sales associated with changes in bestseller status. Whereas the results in Table 3 are based on comparisons *across* books (with mistakenly omitted books serving as the control group for the listed books), this second approach is based on within-book variation over time (so the control is effectively the same book's sales trajectory *prior* to its appearance on the list). Observing sales over several weeks for each book makes it possible to control for book fixed effects, thus absorbing the obviously endogenous differences in sales levels for bestseller vs. non-bestseller titles. Moreover, the time lag involved in the *New York Times* list means that we have at least three 'pre-treatment' observations on

[19] Recall that the dominant pattern in sales over time is a steady decay: even for books appearing on the bestseller list, it is rare for sales to increase from one week to the next.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

each bestseller before it hits the list. (Due to the time lag, the soonest a book can appear on the list is at the beginning of its fourth week on bookstore shelves.) In principle, therefore, the data can be used to estimate a model that controls for book-specific differences in the level of sales *and* book-specific differences in sales trends observed prior to appearance on the bestseller list.[20]

An empirical model of book sales over time must accommodate the dominant decay trend in sales over time (as described in section IV(i)) and allow for appearances on the bestseller list to directly affect sales. Moreover, goodness of fit is even more important than usual here, since the counterfactual we want from the model (how many units a book would have sold if it hadn't appeared on the bestseller list in a certain week) is essentially a forecast. Given these considerations, one simple approach is to model book sales as an autoregressive process in which the autoregression parameter is a function of covariates:

$$(1) \qquad sales_{it} = \lambda_{it} sales_{it-1} + \varepsilon_{it},$$

$$\lambda_{it} = X_{it}'\lambda$$

with X_{it} a set of covariates for book i in week t (including bestseller status) and ε_{it} a demand shock that is independent (but not necessarily identically distributed) across i and t. The autoregressive structure is appealing in part because it tracks sales quite well, so that the prediction in any period will never be off by much. Moreover, this specification focuses the estimation on changes in sales for a given book rather than differences in the level of sales across books, and allows for book-specific differences in the rate of decay in sales (via book-specific constants in λ_{it}). In other words, the model can accommodate unobserved, book-specific heterogeneity in both the level and trajectory of sales. In order for any remaining endogeneity to be important it must take a peculiar form. In particular, the *timing* of a book's appearance on the bestseller list would have to correspond to weeks of idiosyncratically high demand in order to bias our estimate of the list's impact. Given that the current list reflects sales from three weeks prior, such endogenous timing seems very unlikely. Although publishers are surely strategic about when they release books, it seems safe to presume they don't systematically release books exactly three weeks in advance of anticipated positive demand shocks.[21]

[20] Though implemented somewhat differently, the empirical strategies employed here are similar in spirit to the analysis of Reinstein and Snyder [2000], who exploit quirks in the timing of movie critics' reviews to identify their impact on box office sales.

[21] One potential caveat is that publishers may strategically avoid releasing books at the same time as major blockbusters, whose release dates are announced well in advance. (This is discussed in more detail in section 4.3 below.) However, even in this case it is not clear why publishers would time their releases such that the positive shock comes exactly three weeks after release.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

TABLE 4
REGRESSION ESTIMATES: BOOK SALES AND LIST STATUS

	I	II	III	IV
NYT-listed	.043		.030	
	(.015)		(.019)	
NYT, week 1		.087		
		(.025)		
NYT, week 2+		.011		
		(.020)		
NYT, 1st week off		−.002		
		(.033)		
NYT-listed × rank			.004	
			(.004)	
NYT-listed × [1–5]				.042
				(.016)
NYT-listed × [6–10]				.040
				(.027)
NYT-listed × [11–15]				.068
				(.053)
Oprah	6.356	6.286	6.344	6.346
	(.111)	(.111)	(.114)	(.112)
GMA	.108	.100	.118	.109
	(.085)	(.083)	(.086)	(.085)
Weeks out	−.064	−.055	−.066	−.065
	(.022)	(.021)	(.024)	(.023)
R^2	.977	.978	.977	.977
# books	799	799	799	799
# observations	19,024	19,024	19,024	19,024

Robust standard errors in parentheses. All explanatory variables are interacted with lagged sales, as explained in the text, and each specification includes week fixed effects and book fixed effects. The number of observations does not equal the number of books times 25 because some weeks are missing.

Table 4 reports estimates of this model for four separate specifications, using weeks 2–26 for each book in the sample. All four specifications include a full set of week dummies to control for seasonal variation in book sales (there is a large increase in sales in mid to late December, for example), as well as a full set of book dummies in λ_{it}. Column I reports the estimated coefficient on an indicator that equals one for every week in which the book appeared on the *New York Times* bestseller list. The point estimates are statistically significant at the 5 per cent level, and imply that sales decline about 4 percentage points more slowly when a book is listed as a bestseller.

Given the many theoretical reasons why appearances on a bestseller list should have a direct impact on sales, it is not surprising to find a statistically significant effect. The more interesting question is perhaps *why* the list has an impact—i.e., what is the relevant mechanism by which list appearances boost sales? One class of explanations is based on the idea that bestseller status is informative: for example, bestseller appearances could signal high quality, or signal what other people are reading (and thereby facilitate social coordination). Another possibility is that sales increases result from promotional activities undertaken for bestsellers (e.g., prominent placement on bookstore shelves). A rather mundane explanation would be that the

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

increases are merely price effects, resulting from the automatic discount that many stores offer on current bestsellers.

The specifications reported in columns II–IV of Table 4 attempt to shed light on the source of the bestseller list's impact, and specifically try to distinguish whether information or promotion is driving the effects. Column II separates the effect of a list appearance between the first week and subsequent weeks on the list. If the sales increases are driven by information, we should expect them to be concentrated in week one, when the information is new to consumers. On the other hand, if sales increases mostly reflect promotional activities, then the effects should be apparent as long as the book remains on the list. The estimates are more consistent with the information-shock explanation: they indicate an 8.7 percentage-point change in the week the book first appears (statistically significant at the 5 per cent level), with any effects in subsequent weeks (combined) being statistically indistinguishable from zero. Moreover, a book's removal from the list (which, according to the bookstore managers with whom I spoke, abruptly terminates any special in-store promotional activity for the book) has little impact on its sales: the coefficient on the dummy for 'first week off' is neither statistically nor economically significant. Given this result, it is very unlikely that the measured impact of bestseller status represents a pure price effect.

Just as informational effects should be concentrated in the week of a book's first appearance on the list, they may also be concentrated among books at the top of the list. There is obviously a big difference in the buzz generated by a number-one bestseller than by a book barely breaking through to the bottom of the list, and social effects in consumption are presumably strongest for the books at the very top. By contrast, the effects of in-store promotions should be equally important for all books on the list, because all of them are put on a special rack, and/or all of them are discounted. Columns III and IV of Table 4 report estimates from regressions in which the impact of a list appearance is allowed to depend on the book's position on the list. Based on these estimates, there is no evidence that the list's impact is larger for books at the top of the list: the interaction of the list dummy with list rank is statistically indistinguishable from zero in specification III, and the coefficients on list dummies for separate groups (books 1–5 on the list, books 6–10, and books 11–15) are statistically indistinguishable in specification IV ($F_{2,18114} = 0.13$, p-value $= 0.88$). If anything, the point estimates suggest larger effects for books at the *bottom* of the bestseller list.

In interpreting these results, it is important to remember that the information content of an appearance on the list is quite low for some authors—everyone knows that Grisham's newest novel will make the list, so there is no information shock when it first appears—and the estimates in Table 4 report the *average* effect over all books that made the list. Indeed,

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

it could be that for most books, appearing on the list has virtually no effect on sales; only appearances that are surprises make a difference. The results reported in columns III and IV of the table, suggesting that appearances at the top of the list have no more impact on sales than appearances at the bottom of the list, may simply reflect that most top-ranked bestsellers are books that were expected to be there.

In order to examine this issue more carefully, the model was re-estimated allowing for book-specific heterogeneity in the effect of being listed. That is, the model was estimated with

$$\lambda_{it} = \beta_i NEWBS_{it} + X'_{it}\lambda, \tag{2}$$

where $NEWBS_{it}$ is an indicator equal to one if book i appeared on the bestseller list for the first time in week t. The most striking feature of the estimated β_i's is how they relate to the authors' histories. Figure 4 shows kernel densities[22] for the estimated β_i's separately for well-established authors vs. newer authors.[23] For established authors, the distribution is centered at -0.02 and is approximately symmetric. For authors who were less well-established, the distribution has a positive mode and a long right tail, which is consistent with the idea that bestseller status has an appreciable impact only for some books. Perhaps the most interesting result relates to the small sample of eleven *debut* authors who made the list. For these authors, the estimated β_i's are consistently large and positive: of the eleven, nine are in the top quartile, and seven are in the top decile. The average value of $\hat{\beta}_i$among these new authors is 0.35 (compared with 0.05 for the remainder of the authors). Although the sample size is obviously small here, the patterns are clearly consistent with the hypothesis that appearing on the bestseller list only has an impact when the appearance is informative.

The heterogeneity evident in Figure 4 is important to keep in mind when interpreting the quantitative results from Table 4. Taken at face value, the estimated coefficients imply a modest *average* effect of list appearances on sales, even though the specified autoregressive model allows the effect to persist. For example, comparing a book that appeared only once on the bestseller list (in its fourth week from release) to a book that started with the same initial sales but never appeared on the list, and assuming a constant λ

[22] Two outliers were omitted in the estimation of the density. One, an extreme negative value, was for a book that was announced as a television book club pick on the same day it first appeared on the bestseller list, so the data have difficulty distinguishing the two effects. The other, an extreme positive value, was for a book with a 'mystery author.' *The Diary of Ellen Rimbauer* was nominally written by a 'new' (but fictitious) author named Joyce Reardon, and it was rumored that the book was actually written by Steven King. Sales spiked considerably when the book first hit the bestseller list, so its estimated β_i is very large.

[23] The groups of authors were split based on the number of past adult fiction titles each author had published. The top third (having published 35 or more titles) were designated the 'established' authors, and the remaining two thirds were called the 'newer' authors.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

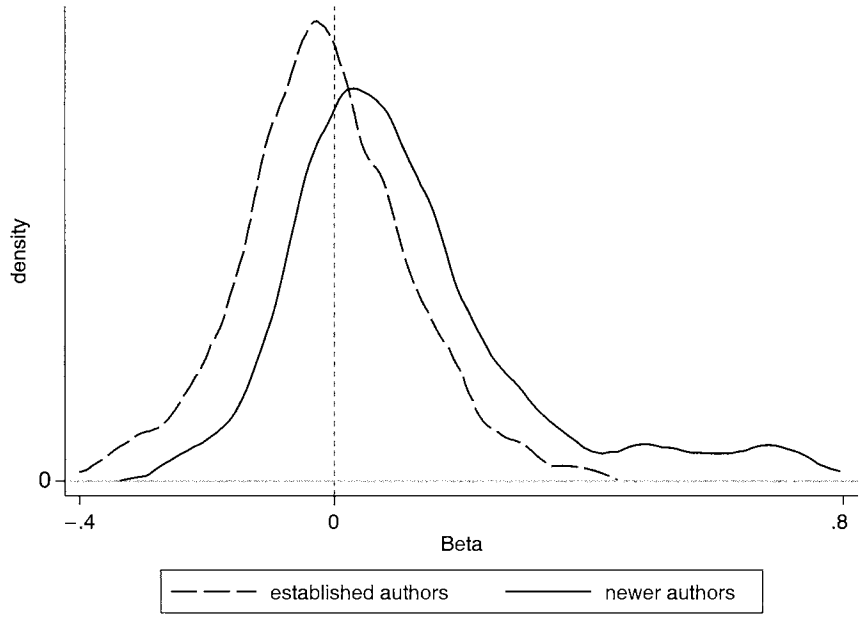


Figure 4
Heterogeneity in the impact of bestseller list on sales

equal to 0.7, the 8 percentage-point increase in week 4 translates to a 13.7 per cent difference in expected sales over the first 52 weeks.[24] However, the average effects seem to be masking the real story: for established bestselling authors, the list has no discernible impact on sales; but for new authors, appearing on the list has a relatively dramatic impact. (A one-time increase in λ of 0.35 in week 4 would lead to a 57 per cent increase in sales over the following 52 weeks.)

Note that even at their largest, the effects of bestseller list appearances are small when compared to the effects of the announcement variables included in the regressions reported in Table 4. The 'Oprah' indicator is equal to one if the book was announced as a selection for Oprah Winfrey's book club in that week, and the 'GMA' indicator is equal to one if the book was announced as the pick for the 'Good Morning America' show's book club. The estimated effects of these announcements are very large—they dwarf any effect of the bestseller list. The influence of the announcements could derive from various

[24] If S_0 is initial sales, then total sales over T weeks is $S_0 \sum_{t=0}^{T-1} \lambda^t$. If λ increases by an amount Δ in week r (but then reverts to its previous value and is otherwise constant), then the percentage increase in sales is equal to $\Delta \sum_{t=r-1}^{T-1} \lambda^{t-1}$.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

sources: the announcements could simply alert a relatively large number of consumers to the existence of the book; they could serve as a quality signal; or they could act as a coordination mechanism by which a large number of consumers agree to read the same book. (The latter mechanism is the apparent objective of the television book clubs.)

Finally, although the specifications reported in Table 4 should adequately control for unobserved heterogeneity related to books' varying levels of sales, it is nevertheless useful to construct a reality check for the key estimated coefficients. For example, instead of regressing sales on an interaction of lagged sales and an indicator for 'first week on the list,' we can interact lagged sales with an indicator for 'first week almost on the list,' defined as any unlisted book with sales greater than 90 per cent of the 15^{th}-ranked bestseller. If the coefficients reported in Table 4 are merely an artifact of latent heterogeneity in sales dynamics that is correlated with differences in the level of sales, then the coefficient on this variable would be positive and significant. In fact, running this regression yields a coefficient estimate of 0.028 with a standard error of .018, lending some additional credibility to the results reported in the table.

IV (iii). *Do Bestseller Lists Affect Product Variety?*

The results of the previous section indicate that some books appearing on the *New York Times* bestseller list enjoy a consequent increase in sales. But to what extent are those extra sales 'stolen' from non-bestselling titles? As was explained in section II, whether (and in which direction) a bestseller list influences product variety depends on whether the list has any impact on books near the publish/no-publish margin. If the consumer who buys a book because he saw it on the bestseller list would otherwise have bought a book near the margin of profitability, then one can argue that the list reduces the privately optimal number of books by further concentrating demand on bestsellers. However, it is also possible that the consumer would have otherwise bought no book at all—in that case, sales are more concentrated on bestsellers, but this comes at no expense to non-bestselling titles, and the number of books published is unaffected. Moreover, it is also possible that bestseller lists increase demand for *all* books, for example by simply bringing more consumers into the bookstore. Bestsellers and non-bestsellers could, in principle, be complementary goods if consumers buy multiple books when they visit the bookstore.

The data available are clearly insufficient to provide a conclusive answer to this question. The ideal experiment might be one in which a large set of consumers makes purchases in the presence of a bestseller list, and another set of consumers makes purchases without having any exposure to the bestseller list (either through the media or at the retail outlet itself).

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

TABLE 5
SUMMARY OF LISTED GENRES

Genre	Per cent of books listing this genre	Per cent of bestsellers listing this genre
Literature/Fiction	68	44
Mystery/Thriller	39	56
Romance	11	19
Science Fiction	6	9
Religion	2	3
Horror	2	3

Based on books' genres as listed on *Amazon.com*. Percentages add to more than 100 because books typically list more than one genre.

The ubiquity of the *New York Times* list makes it virtually impossible to find any group of book purchasers that resembles such a control group.

What the available data *can* potentially reveal is indirect evidence of substitution between bestsellers and non-bestsellers. Even this is difficult, however, since the data contain no price variation that would enable estimation of cross-price elasticities. The only useable variation is time variation in the subject composition of the bestseller list. If substitution between non-bestsellers and bestsellers is important, then—presuming that books in the same genre are closer substitutes than books in different genres—sales of non-bestselling books should decline when the bestseller list is comprised of books in similar genres. For example, sales of a non-bestselling detective novel would decrease in a week when three detective novels simultaneously hit the bestseller list.

In order to capture these kinds of substitution patterns, a variable summarizing each book's similarity to the current set of bestsellers was constructed by comparing the book's genre(s) (as listed by Amazon.com) to the genres of all books on the current bestseller list. Specifically, the pairwise similarity between books A and B is defined as

$$(3)\ sim(A,B)=\frac{2\times(\#\text{ of genres shared by books A and B})}{(\text{number of genres listed for A})+(\text{number of genres listed for B})}.$$

This measure is equal to one if the two books' genres are identical, and zero if there is no overlap at all.[25] Book A's 'average subject similarity' to the current bestseller list is then computed as $1/15\sum_{r=1}^{15} sim(A,r)$, where r indexes the current bestsellers by rank. Table 5 shows the relative frequencies of the genres in the sample for all books and for bestsellers only. As is clear from the table, mysteries and thrillers are the most common bestsellers, and

[25] Amazon.com often lists multiple genres for the same book: e.g., 'Contemporary fiction' and 'Romance.' So $sim(A,B)$ will be less than one unless books A and B list exactly the same genres.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

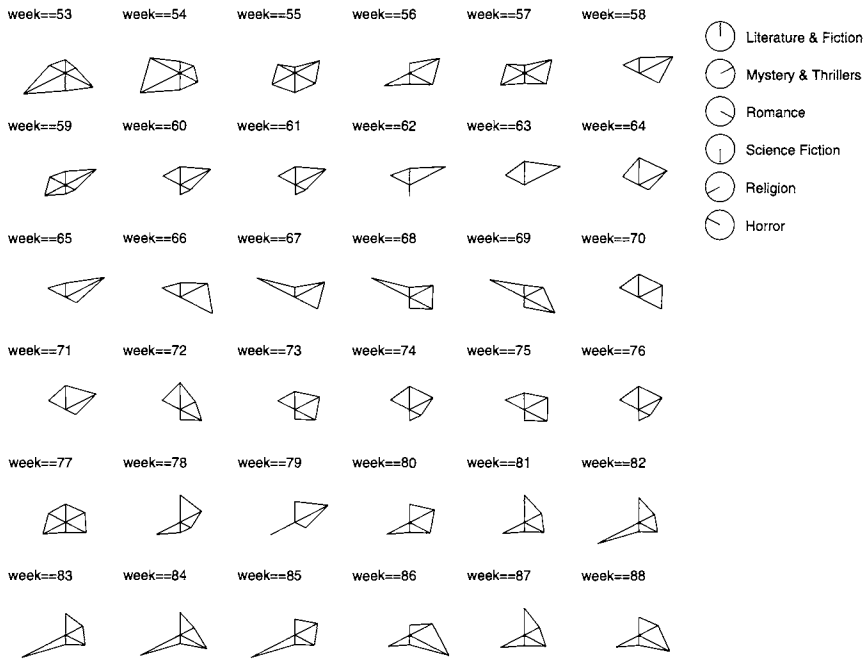


Figure 5
Variation in bestseller list composition over time

romance novels are represented more heavily among bestsellers than among books overall. Importantly, there is substantial variation in the composition of the bestseller list over time. Figure 5 shows a series of star graphs to illustrate variation in list composition for a sample 36-week period. To the extent that there is meaningful substitution between non-bestsellers and bestsellers, we should observe that sales respond to changes in the average similarity variable induced by variation in the list composition.

Table 6 reports estimates of a model analogous to the autoregressive model of the previous section, but with similarity measures included in the λ_{it}.[26] Instead of a business-stealing effect, the estimated coefficients seem to suggest a complementarity between bestsellers and non-bestsellers. Weeks in which the genre-composition of the bestseller list is close to a non-bestselling

[26] Because the purpose is to evaluate the potential response of non-bestsellers to changes in their similarity with bestsellers, only books that never made the bestseller list are used in the estimation. The table reports specifications with and without subject fixed effects; without these fixed effects, the coefficients on the similarity measures could partly reflect average differences in the prevalance of certain genres on the bestseller list, which would obviously be correlated with sales in that genre.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

TABLE 6
REGRESSION ESTIMATES: BOOK SALES AND SIMILARITY TO CURRENT BESTSELLERS

	I	II	III	IV
Avg. Similarity:				
All bestsellers	.044	—	.228	—
	(.033)		(.080)	
New on list	—	.232	—	.400
		(.100)		(.102)
Weeks out	.005	.005	.003	.002
	(.002)	(.002)	(.002)	(.002)
Week dummies?	yes	yes	yes	yes
Subject dummies?	no	no	yes	yes
R^2	.945	.945	.951	.951
# books	595	595	595	595
# observations	14,115	14,115	14,115	14,115

Robust standard errors in parentheses. All explanatory variables are interacted with lagged sales, as explained in the text. Avg. Similarity measures how similar a book is to the current set of bestsellers, based on the books' listed genres. Only books that never made the bestseller list are included in the estimation. The number of observations does not equal the number of books times 25 because some weeks are missing.

book's genre tend to be *good* weeks for the non-bestseller. Consistent with the findings in Table 4, the effect seems to be most pronounced for books appearing on the list for the first time: when the similarity measure is defined only for the subset of bestsellers whose first appearance was in the given week, the positive effect of the similarity is stronger. This pattern could plausibly reflect multiple-book purchases: for example, weeks in which a new romance novel hit the bestseller list draw more than the usual fraction of romance enthusiasts into bookstores, and they may buy several romance novels when visiting the store.

Although the estimated coefficients on the similarity measures are statistically significant, they should be interpreted carefully. One caveat to keep in mind is that publishers choose their release dates strategically. For example, there is some evidence that publishers of non-bestselling books tend to avoid release dates that coincide with the release of a blockbuster title in the same genre. The regressions in Table 6 may be biased against finding business-stealing, because publishers choose their books' release dates to actively avoid business-stealing. That is, if book A were expected to steal business from book B, then book B's release date would be strategically distant from book A's release, so that potential instances of business-stealing are less likely to be observed in the data. Another thing to keep in mind when interpreting the coefficients is that they represent (at best) indirect evidence of the substitution patterns between bestselling and non-bestselling titles, and the results may not generalize easily to other categories of books. For example, non-fiction titles on similar topics are perhaps less likely to exhibit complementarities. A consumer looking to buy a book on investments, for instance, seems more likely to choose just one (instead

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

of buying many) than would a consumer looking for a suspense or romance novel.[27]

V. DISCUSSION & CONCLUSIONS

Based on the sales data analyzed here, it seems clear that appearing on the *New York Times* bestseller list has a direct impact on a book's sales. Estimates based on two alternative identification strategies show that sales increase when a book appears on the list. However, the magnitude of the effect is modest, and appears to reflect spikes in the sales of books that were surprises on the list (e.g., books by new authors).

Given the amount of attention paid to the *New York Times* list (and the desperate schemes occasionally employed by authors to secure a position on the list), the estimates of its impact may seem surprisingly small. However, the present analysis ignores two effects that could be quite significant to authors and publishers. First, while appearing on the bestseller list may have only a modest immediate impact on the book's sales, it may dramatically increase the popularity of future books by the same author. Second, paperback sales may be influenced by whether the hardcover edition was a bestseller. (Indeed, paperback versions of books that were hardcover bestsellers typically announce that fact prominently on the front cover.) Although these effects can't be measured with the available data, anecdotal evidence suggests they may be important.

Although there are a variety of reasons why the bestseller list might directly influence sales, the patterns in the data are most clearly consistent with an information-based explanation. Retail-level promotions cannot rationalize two of the study's central findings: whereas retailers' promotions persist for the duration of a book's term on the list, the estimates suggest that the impact of appearing on the list is transitory, with the bulk of the effect realized in the first week. Moreover, the impact on sales is most pronounced among relatively unknown authors (new authors in particular), a pattern that favors information over promotion as an explanation for the effect.

The impact of the list on sales (and the consequent increase in the concentration of demand on bestsellers) raises the interesting counterfactual question: 'Would more books be published if it weren't for bestseller lists?' Whether (and in which direction) the bestseller list shifts the publish/no-publish margin depends on the nature of substitution between bestselling and non-bestselling titles—i.e., are the extra sales of bestsellers ones that

[27] In unreported analyses using sales data for nonfiction titles, coefficients on similarity measures like those in Table 6 indeed suggested a pattern of substitution rather than complementarity. Nonfiction titles have the advantage of being able to categorize subjects much more finely than in fiction, but it was difficult to include nonfiction titles in the broader analysis because the coverage of nonfiction bestsellers in my sample was sparse.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions

would have otherwise gone to less popular books? The results here offer indirect evidence that for hardcover fiction, bestsellers and non-bestsellers within the same genre may in fact be complements: weeks in which books of a particular genre first appear on the bestseller list tend to be strong-selling weeks for non-bestsellers of the same genre. Although too indirect to be conclusive, this result suggests that market expansion effects dominate any business-stealing associated with bestseller lists, so that bestseller lists may in fact increase the number of books published in equilibrium.

REFERENCES

Alexander, P., 1997, 'Product Variety and Market Structure: A New Measure and a Simple Test,' *Journal of Economic Behavior and Organization*, 32(2), 207–214.

Banerjee, A., 1992, 'A Simple Model of Herd Behavior,' *Quarterly Journal of Economics*, 107(3), 797–817.

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

Berry, S. and Waldfogel, J., 2001, 'Do Mergers Increase Product Variety? Evidence from Radio Broadcasting,' *Quarterly Journal of Economics*, 116(3), 1009–1025.

Bogart, D. (ed.), 2001, *The Bowker Annual*, 46th edition.

Clerides, S., 2002, 'Book Value: Intertemporal Pricing and Quality Discrimination,' *International Journal of Industrial Organization*, 20(10), 1385–1408.

Cole, D., 1999, *The Complete Guide to Book Marketing*, (Allworth Press, Chicago).

Dixit, A. and Stiglitz, J., 1977, 'Monopolistic Competition and Optimum Product Diversity,' *American Economic Review*, 67(3), 297–308.

Einav, L., 2003, 'Gross Seasonality and Underlying Seasonality: Evidence from the U.S. Motion Picture Industry,' working paper, Stanford University.

Epstein, J., 2001, *Book Business: Publishing Past, Present, and Future*, (W.W. Norton).

Greco, A., 1997, *The Book Publishing Industry*, (Allyn & Bacon).

Korda, M., 2001, *Making the List: A Cultural History of the American Bestseller, 1900–1999*, (Barnes & Noble Books, New York).

Lancaster, K., 1975, 'Socially Optimal Product Differentiation,' *American Economic Review*, 65(4), 567–585.

Milgrom, P. and Roberts, J., 1986, 'Prices and Advertising Signals of Product Quality,' *Journal of Political Economy*, 94, 796–821.

Reinstein, D. and Snyder, C., 2000, 'The Influence of Expert Reviews on Consumer Demand for Experience Goods,' working paper, George Washington University.

Rosen, S., 1981, 'The Economics of Superstars,' *American Economic Review*, 71(5), 845–858.

Suzanne, C., 1996, *This Business of Books*, (Wambtac).

Sweeting, A., 2005, 'Too Much Rock & Roll? Station Ownership, Programming and Listenership in the Music Radio Industry,' working paper, Northwestern University.

Vettas, N., 1997, 'On the Informational Role of Quantities: Durable Goods and Consumers' Word of Mouth Communication,' *International Economic Review*, 38, 915–944.

Watson, R., 2003, 'Product Variety and Competition in the Retail Market for Eyeglasses,' working paper, Northwestern University.

© 2007 The Author. Journal compilation © 2007 Blackwell Publishing Ltd. and the Editorial Board of The Journal of Industrial Economics.

This content downloaded from 130.212.18.200 on Thu, 6 Mar 2014 16:11:23 PM
All use subject to JSTOR Terms and Conditions