

Searching for Physical and Digital Media: The Evolution of Platforms For Finding Books*

Michael R. Baye
Indiana University

Babur De los Santos
Indiana University

Matthijs R. Wildenbeest
Indiana University

May 2013

PRELIMINARY AND INCOMPLETE

Abstract

This chapter provides a data-driven overview of different platforms where consumers can search for books and booksellers, and how the use of these platforms has shifted over time. We highlight a number of challenges and open agenda items related to observed data on consumer search as well as prices of digital and physical books.

Keywords: product search, books, internet, search engines

*Department of Business Economics and Public Policy, Kelley School of Business, Indiana University, Bloomington IN 47405; mbye@indiana.edu, babur@indiana.edu, and mwildenb@indiana.edu. We thank Avi Goldfarb, Shane Greenstein, Catherine Tucker, and Hal Varian for valuable comments, and thank Susan Kayser, I. K. Kim, Joowon Kim, and Zachary Mays for research assistance. Funding for the data and research assistance related to this research was made possible by a grant from Google to Indiana University. The views expressed in this paper are those of the authors and do not necessarily reflect the views of Indiana University or Google.

1 Introduction

New technologies have changed how people search for, acquire, and consume products. The Internet especially has had a profound impact on the way people shop for products; online sales in the U.S. in 2012 were \$213 billion and now account for 8 percent of all retail sales.¹ One of the first online industries that saw a major shift towards online sales is the book industry. According to U.S. Census Bureau statistics, in 2009 e-commerce sales of books and magazines were \$5,214 million—about one-quarter quarter of all U.S. sales of books and magazines for that year.²

The Internet has also fundamentally changed consumers' abilities to identify books that are a good match for their preferences. It allows consumers to find a greater breadth of products and makes the long-tail more accessible; it is now easier to find rare and obscure books. Finding the “right” book takes less effort, thanks in part to newly developed recommendation systems that often operate in real-time. This allows consumers to find books they would otherwise be unaware of, thanks to recommendations based on their previous purchases as well as the purchase patterns of consumers with similar profiles. It also makes it easier to compare the prices different retailers charge for a given book.

More recently, the industry is shifting towards e-books, and this allows consumers to locate, purchase, and read books virtually anywhere. In this chapter we give an overview of recent developments in the book industry and how this has impacted the book industry from both the consumers' perspective as well from the perspective of bookstores and publishers. Doing so, we discuss the economic implications of these developments and point out future areas of research.

Of course, we are not the first to examine the online book industry. Clay, Krishnan, and Wolff (2001) study pricing strategies of online bookstores as well as price dispersion and find that more competition led to lower prices and less price dispersion. Moreover, they find heterogeneity in firm-level behavior to be important. Brynjolfsson, Hu, and Smith (2003) quantify the economic impact of increased variety at online bookstores and find this to have increased consumer welfare

¹Figures from Forrester Research, March 2013.

²Annual bricks-and-mortar bookstore sales (NAICS 451211) were \$15,944 million in 2009 (see the 2011 U.S. Census Bureau Annual Retail Trade Report), while annual book sales from electronic shopping and mail-order houses were \$6,824 million for that year (see Table 1056 of the U.S. Census Bureau, Statistical Abstract of the United States: 2012).

by up to \$1 billion for the year 2000. Chevalier and Goolsbee (2003) show how to use sales rank data to obtain elasticity estimates for Amazon.com and Barnesandnoble.com and find that demand is much more price elastic at Barnes and Noble than Amazon. Chevalier and Mayzlin (2006) study the effect of consumer reviews on sales at Amazon.com and Barnesandnoble.com. Hong and Shum (2006) develop techniques to estimate search costs using only price data and apply their methods to books sold online. Forman, Ghose, and Goldfarb (2009) provide powerful evidence that online and offline markets for books are linked: When a physical book store opens, consumers near the area purchase fewer books online. De los Santos (2012) and De los Santos, Hortaçsu, and Wildenbeest (2012) use data on browsing behavior and book purchases to estimate search costs.

This chapter differs from these and other studies of online book markets in several respects. First, our focus is on consumer search; that is, the mechanisms consumers use to find books. It is apparent that the way consumers search for books has evolved over the past two decades, and our goal is to document some of these changes. Additionally, and consistent with the “Agenda” theme of this volume, we discuss a number of technical challenges associated with identifying book and book-related searches on browser-based platforms (e.g., search engines, price comparison sites, and retail websites such as Amazon.com), as well as closed-systems (e.g., e-readers such as the Kindle and Nook). Finally we provide some preliminary evidence that evolving search platforms are leading searchers to shift away from browser-based platforms (such as search engines) to conduct searches for books.

Our chapter is organized as follows. The next section provides a brief overview of the book industry and highlights some of the more important changes that have occurred over the past 20 years. Section 3 provides evidence on how consumers use different browser-based platforms (such as websites of booksellers, price comparison sites, and search engines) to locate books and booksellers. Section 4 investigates how search intensity changes throughout the life cycle of a book, and how overall searches for book titles on search engines has evolved. Section 5 compares the distributions of prices for digital and physical books. It also examines how these prices evolved before, during, and after the Antitrust Division of the Department of Justice filed suit against Apple and publishers regarding their use of agency contracts. We conclude in Section 6 by briefly

describing several additional agenda items for research programs.

2 Overview of the Book Industry

The book industry has seen many changes throughout the past 20 years. Arguably the most significant change has been the move from selling books in traditional brick-and-mortar bookstores to selling books online. Books were one of the first products that were successfully sold online. One of the world's first and largest online retailers, Amazon.com, started as an online bookstore in July 1995. Barnesandnoble.com was launched in May 1997. Many startups followed, and existing brick-and-mortar bookstores began creating an online presence. Most of these online bookstores (retailers such as 1bookstreet.com, Allbooks, and Fatbrain.com) no longer exist; many went bankrupt during the collapse of the Internet bubble shortly after the new millenium began. Borders, which was one of the largest brick-and-mortar bookstores in the U.S., went bankrupt in 2011. Today, the retail book industry consists of Amazon, Barnes and Noble, and a handful of other sellers that include smaller players as well as mass-merchants (such as Walmart) who have thinner offerings of titles.

A more recent change in the book industry is the shift towards electronic books, or e-books. Although e-books have been around since at least 1971, when Michael Hart made the Declaration of Independence electronically available as part of Project Gutenberg, the release of the Sony LIBRIé e-Book reader in 2004 in Japan (the U.S. followed in 2006 with the release of the Sony PRS-500) marked the beginning of the recent surge in popularity of e-books.³ The Sony LIBRIé was the first commercially available device to use E Ink technology, which dramatically improved the reading experience and required less battery power than existing e-readers at the time. Amazon released its E Ink technology-based Kindle in November 2007. Barnes and Noble followed in November 2009 with the release of the Nook. At the end of June 2012, Amazon led the market for devices used for reading e-books, with an overall market share of 55 percent (with its Kindle and Kindle Fire), followed by Nook (14 percent). Today, Sony's e-book readers have 1 percent of the market.⁴ The definition of an e-reader is increasingly blurred by the presence of devices such as tablets, smart phones, desktops, and laptops with applications for reading e-books. Indeed, after having seen

³See Manley and Holley (2012) for an overview of the history of e-books.

⁴Estimates from Bowker Market Research.

dramatic growth since 2006, the sales of traditional e-readers are expected to decline in 2012, with consumers shifting towards tablets.⁵

The sales of e-books have grown alongside the increasing popularity of e-readers. By July 2010, Amazon was selling more Kindle books than hardcover books, and since April 2011 it is selling more Kindle books than hardcover and paperback books combined (excluding free e-books).⁶ However, as shown in Table 1, the overall increase in e-book sales revenue has grown less dramatically, with e-books accounting for only 5 percent of the market in 2010.

Most e-readers support the open ePub format (Amazon's Kindle, which supports the proprietary AZW format, is a notable exception), which in theory means consumers can buy an e-book at one of the bookstores that supports the ePub format and read it on a different e-reader. However, bookstores use Digital Rights Management (DRM) to protect the content of the e-books, and different bookstores do not necessarily use the same DRM standard. Even though there are ways to get around this (for instance, by stripping the DRM), for most consumers this is difficult to do. This means that in practice the e-book market has many characteristics of a two-sided market, with publishers on one side of the market, readers on the other side, and bookstores and e-readers as intermediaries.⁷

The shift towards e-books is likely to significantly impact the way consumers search for books. Buyers of e-books are less likely to use web browsers to find and purchase books, as it is often more convenient to search within the device that is used to read to e-books. Moreover, the lack of interchangeability due to the aforementioned compatibility issues with respect to DRM will make it less beneficial for consumer to compare prices across bookstores or platforms, resulting in fewer retailers being visited. Unfortunately, we do not observe consumer search patterns within most of the (closed) platforms, and therefore focus on browser-based searches for books in the remainder of this chapter.

These developments—the move towards selling books online as well the transition to e-books—have significantly impacted the sales of books at traditional brick-and-mortar stores. Figure 1 shows

⁵See http://news.cnet.com/8301-1035_3-57558710-94/rip-e-book-readers-rise-of-tablets-drives-e-reader-drop/

⁶See <http://phx.corporate-ir.net/phoenix.zhtml?c=176060&p=irol-newsArticle&ID=1565581>.

⁷See Rysman (2009) for a discussion on the economics of two-sided markets.

that average monthly sales at brick-and-mortar bookstores steadily increased until 2007, when they peaked and have since steadily declined.⁸ This decline in sales of physical books coincides with a recession and Amazon’s introduction of the Kindle e-book reader, which makes it difficult to separately identify the impact of the recession and the entry of digital media. While we view this to be an interesting and open question, the data in Table 2 suggests that the post-2007 decline in sales at traditional bookstores is at least partly due to a shift towards e-books and the online channel more generally: During this period, sales at physical bookstores declined relative to sales at online bookstores.

With this brief background, we now turn to search.

3 Browser-Based Searches for Books and Booksellers

There are numerous ways consumers can use an Internet browser to search for books or a bookseller. Consumers can use a general search engine such as Google, Yahoo, or Bing to search for a specific book title, author name, or other book-related search term. Alternatively, a consumer can use a price comparison sites for books, such as booksprice.com or bookfinder.com. A third option is to go directly to one of the online bookstores, such as Barnesandnoble.com or Amazon, and do a search within its site. This section provides a look at the relative importance of these different platforms for conducting searches, and highlights a number of obstacles that make it difficult to precisely measure the share of book searches that are conducted on different search platforms.

3.1 Search Activity on Book-Oriented Platforms

One way of measuring search activity is to rely on aggregate data on total searches, such as comScore’s qSearch database. The qSearch data is based on the comScore panel, which contains all browsing activity of about two million users in the U.S. It contains monthly data on searches conducted by consumers on the internet.⁹ These searches include searches through traditional search boxes as well as toolbars and widgets. The qSearch database measures actual search volume

⁸Figure 1 uses U.S. Census Bureau data from the Monthly Retail Trade Survey for NAICS 451211 (bookstores).

⁹According to comScore’s qSearch documentation, a search is defined as: (1) a user interaction where the user is presented with a search result page that contain results that match the consumer’s search intent; (2) the search result page allows the user with the ability to refine or change their search parameters; and (3) the search can be initiated from a drop down or clicking a link, as long as first two rules are satisfied.

and search intensity for the main general search engines (Google, Yahoo, Bing, AOL, and ASK) as well as over 200 other online properties such as Facebook, eBay, Wikipedia, and Amazon. Our focus here is on book-oriented platforms that are included in the qSearch database.

As a starting point, and to highlight some of the challenges that exist in measuring browser-based searches for books, Table 3 provides a snapshot (February, 2012) of the total number of searches on leading platforms that a consumer might use to search for a book. Importantly, though, these data include *all* searches on these platforms, not just searches for books. Thus, it is not surprising that the lion's share of searches during that month were conducted on the main general search engines. The level of search activity is even more skewed towards search engines when one measures searches using search visits.¹⁰ Amazon processed the majority of searches among the bookstores, having almost ten times as many searches as the number two bookstore, Barnes and Noble. In terms of total searches, a smaller number of total searches were conducted on price comparison sites.

Table 3 highlights some of the challenges involved in measuring the evolution of searches for books and booksellers: These data overstate the importance of search engines for conducting book searches, owing to the fact that the aggregate data include searches for weather, restaurants, and information unrelated to books. While we will address this issue in more detail in the next section, we first sidestep the issue by using the qSearch data to examine the evolution of total searches on the book-oriented platforms in Table 3 (Amazon, Google Books, Abebooks, Barnes and Noble, and Book Finder).

Figure 2 shows the evolution of searches between 2011 and 2012 for these book-oriented qSearch properties. The graph shows that, especially for Amazon and Barnes and Noble, there is seasonality in searches during the holiday shopping season, with searches moving up sharply near the end of the calendar year. Searches at Barnes and Noble appear to have peaked near the end of 2011, while searches at Amazon have generally increased, with the exception of the most recent few months.

This raises another drawback of using qSearch data: It does not indicate how many of the

¹⁰According to comScore's qSearch documentation, a search visit is a sessions in which a user conducted one or more searches. If searches are conducted at different points during the day, with more than 30 minutes of search inactivity at the site, it will count as multiple search visits.

searches at these book-oriented platforms are, in fact, book related. While this is not so much an issue for sites that specialize in books (such as Barnesandnoble.com, Bookfinder.com, and Google Book Search), it is clearly an issue for Amazon. While Amazon began its life as an online bookstore, it has evolved into a mass merchant where 53 percent of its 2009 revenues derived from products outside the books category. And today, Amazon is itself a platform that services thousands of Amazon Marketplace sellers.

3.2 Book-Related Searches on Search Engines

We now turn to the issue of measuring the number of book-related searches on search engines. For this we make use of comScore's Search Planner data, which is based on the browsing activity of two million users in the U.S. that are part of the comScore panel. As part of this data we observe the most popular search terms used at the main general search engines to reach a specific site (site profile). To illustrate, Table 4 shows the top 25 search terms on Google that resulted in traffic to Barnesandnoble.com.¹¹ We examined the entire list of terms resulting in traffic to Amazon and Barnes and Noble (2,053 and 180 search terms, respectively), and grouped them into 5 categories: Site Name, Non-Book, E-reader, and Other Bookstore Name.

The results are displayed in Figures 3 and 4, and provide a breakdown of the observed search terms consumers entered at the main search engines that resulted in clicks to Amazon and Barnes and Noble, respectively. Figure 3 shows that clicks on book-related search terms leading to Amazon are less than a third of the number of clicks on non-book related search terms. Figure 4 shows that this is the opposite for Barnes and Noble: book-related search terms generate far more traffic to Barnes and Noble than non-book related search terms. However, a striking observation from both charts is that the overwhelming majority of searches at the main search engines leading users to Amazon or Barnes and Noble were navigational searches, i.e., search terms that only include the name of the retailer, including misspellings. This means that most of the book-related traffic Amazon and Barnes and Noble receive from the main search engines is from consumers using search engines instead of the address bar to navigate. As shown in Figure 5, when one excludes

¹¹In September 2011 Barnes and Noble acquired Borders' intellectual property, including Borders' domain name. Borders.com now redirects to barnesandnoble.com, explaining the relatively large number of Borders' related search terms in Table 4.

navigational searches and accounts for the fact that not all visits to Amazon are book-related (see also the next section), the number of observed book-related searches on Google is about half the total number of searches on Amazon and Barnes and Noble.¹²

A drawback of the Search Planner data is that we only observe a search term if its clicks exceed a certain threshold. Even though this means that we do observe the most popular search terms generating traffic to a specific bookstore, we do not observe the more obscure searches taking place at the search engines. For instance, for Amazon we only observe the search terms corresponding to 24 percent of total click volume. Brynjolfsson, Hu, and Simester (2011) have shown that internet markets have increased the importance of niche markets in terms of sales. This long tail phenomenon suggests that part of the unobserved searches may include searches for book titles and other more obscure book-related search terms, in which case Figures 3 and 4 understate the true importance of book-related search terms. Nevertheless, the importance of the long tail for book sales and searches remains an open question.

3.3 How Searchers Reach Booksellers

To better understand how searchers actually reach the websites of different online booksellers, we analyze incoming traffic to five of the leading online bookstores in the U.S. (Amazon.com, Barnesandnoble.com, Booksamillion.com, Ecampus.com, and Powells.com). In addition to observing the most popular search terms for these retailers, comScore's Search Planner data allows us to observe which sites get clicks for specific searches done at the major search engines (term profile) as well as the source of incoming traffic for each site (source loss).

The results summarized in Figure 6 show the number of visitors to these online bookstores in February 2012, categorized by the most important referring sites or platforms. The majority of visitors used one of the general search engines to get to a bookstore's site, and as shown earlier in Figures 3 and 4, most of these searches were navigational searches. The second most important source of incoming traffic to the leading bookstores are so-called direct logins, i.e., visits that are not

¹²The book search visits for Amazon and Barnes and Noble are direct logins as well as visits from other bookstores to Amazon or Barnes and Noble (see also Figure 6). Since not all visits to Amazon are book related, the book search visits for Amazon are weighted by percentage of media revenue in 2009 (which is 47 percent). The book-related search visits on Google are obtained by taking only book related search terms that generate traffic to either Amazon or Barnes and Noble (see Figures 3 and 4).

preceded by a visit to another website. Slightly over 13 million visitors visited another bookstore first.¹³ Finally, about 7.5 million visits are preceded by a visit to one of the price comparison sites. The chart also shows that Amazon received the vast majority of visits to bookstores. The difference between Amazon and Barnes and Noble, which is second in terms of overall number of visitors, is relatively large for all referring sites and platforms except for the other bookstores category. This suggests that visits to Barnes and Noble are more often accompanied by a visit to another bookstore than a visit to Amazon.¹⁴ This is also shown in Figure 7, which shows the number of visitors from referring sites or platforms as a percentage of the total visits for each of the five bookstores. Notice that referrals from other bookstores are relatively important for Barnes and Noble, and also for Ecampus and Books-A-Million.

4 The Dynamics of Book-Related Searches

This section provides a look at how searches for books on search engines evolve during a window around the release date of the book, and also explores how the shifts toward other search platforms has impacted book-related searches at the main search engines.

4.1 Combining Data from comScore and Google Trends

Since our comScore database only contains information about searches for specific book-related keywords on a monthly basis for the 2010 to 2012 period, our strategy is to combine these data with data from Google Trends to create a dataset that runs from 2004 through 2013.¹⁵

For a given search term, Google Trends provides a measure of the number of searches for that term between 2004 and 2013. The measure is relative: the maximum number of searches within the period is normalized to 100. To illustrate, Figure 8 displays search volume reported by Google Trends for the term “Amazon.”¹⁶ The bars represent monthly search “volume,” whereas the red

¹³The other bookstores category include Abebooks.com, Alibris.com, Amazon.com, Audible.com, Barnesandnoble.com, Biggerbooks.com, Bkstr.com, Bncollege.com, Bookbyte.com, Booksamillion.com, Campusbookrentals.com, Christianbook.com, Ebooks.com, Ecampus.com, Half.com, Kobobooks.com, Powells.com, and Textbooks.com.

¹⁴Note that Borders is not part of the other bookstores, so this does not reflect users being redirected from the borders.com domain. See also footnote 11.

¹⁵Choi and Varian (2012) described how Google Trends data can help to predict economic time series of interest. Google Trend data has been used in epidemiology to detect influenza epidemics (Ginsberg et al., 2009) and in economics to predict the unemployment rate (D’Amuri and Marcucci, 2010) and inflation (Guzman, 2011).

¹⁶We limit the search volume to the United States only.

curve gives the seasonally adjusted trend. The figure shows search volume has been rising for the term “Amazon,” with the growth in search volume accelerating after July 2010. Figure 9 displays the relative search volume over time for “Amazon” and “Barnes and Noble” in a single graph. The number of search queries for Barnes and Noble was relatively stable up to 2010, after which searches went up quickly. However, consistent with Figure 2, searches at Barnes and Noble peaked near the end of 2011, whereas the search volume for Amazon kept growing.

Since the Google Trends measure of search is relative, the relative position of the two lines in Figure 9 is uninformative: It would be erroneous to infer that Barnes and Noble had more navigational searches than Amazon for the 2004-2012 period. Our strategy for overcoming this limitation is to use information about the actual number of searches for a given search terms in comScore to convert the Google Trends index into historical data on the numbers of searches for that search term.

Figure 10 provides one example of the utility of combining Google Trends and comScore data. This graph shows the evolution of searches for e-readers and related terms. By combining the data, we not only obtain a time series dating back to 2004, but the resulting numbers of searches are cardinal so the different graphs may be compared. Figure 10 illustrates that searches for “Kindle” took off in January, 2008, and the number of searches for the Kindle has exceeded that for the Nook every year since. Notice that the recent decline in searches for these two e-readers is consistent with our earlier remarks regarding the projected decline in post-2012 e-reader sales.

4.2 Searches for Specific Titles

We are now in a position to take a deeper dive into searches for books on search engines, and to examine how search patterns have evolved over time. We use the Search Planner data for a sample of 735 books to study how the number of clicks on Google changes throughout a book’s life cycle. Our sample includes titles released in every year since 2003, but roughly half of the sample is comprised of books released in 2012. The search queries we use to obtain the clicks on Google are the exact titles of the book in our sample. In addition to clicks, we observe book characteristics such as ISBN, list price, publisher, data of publication, format (hardcover or paperback), number of pages, edition (e.g., first or reprint), dimension, and weight. Moreover, for 2012 we have daily

observations on a book’s price as well as availability at each of the leading online bookstores.¹⁷ For Amazon we also observe a book’s customer rating, number of customer reviews, and sales rank. Most books in our sample are first editions; the vast majority are New York Times Bestsellers that came out during 2011 and 2012.

The Search Planner data has two limitations. A first limitation is that we only observe clicks for the period between February 2010 and February 2013. Secondly, we only observe clicks if the clicks exceed a certain threshold. To deal with these two matters, we supplement the Search Planner data with Google Trends data, as discussed above. For each book in our sample we use Google Trends to capture the volume of search queries that users enter into Google when searching for the title of a book.¹⁸ Since Google Trends only gives a relative measure of search volume, we convert the index to actual levels of search volume by setting the maximum value of the index during the period between February 2010 and February 2013 equal to the corresponding clicks according to Search Planner.

Figure 11 gives an example for the book “The End of Illness” by David B. Agus, which was released on January 17, 2012. Searches on Google for this title start relatively high, but within a few months drop to relatively low levels. Also shown in this graph is Amazon’s sales rank for this title, which shows a reverse pattern.

The pattern for this particular title is fairly representative of other titles as well. In particular, we ran a linear regression of the total number of searches for each book title in our sample book title searches on a constant and a set of dummies that represent months since release. The regression results for various specifications are shown in Table 5. For the specification in column (1), which includes book fixed effects to control for heterogeneity among books, most of the month-since-release coefficients are significantly different from zero at conventional levels. Figure 12, which plots the predicted values of each of the dummies using the parameter estimates for specification (1), shows that the patterns of search we observed around the release window for “The End of Illness” (see Figure 11) is fairly typical for newly released books: Search activity tends to be high

¹⁷These are Amazon.com, Barnesandnoble.com, Booksamillion.com, Walmart.com, Powells.com, and Ecampus.com. See Section 5 for a more detailed description of the price data.

¹⁸We use exact phrases (exact words in quotes), e.g., “*Gone girl*” instead of *Gone girl*. We limit the search volume to the United States only.

in the first few months, but then quickly drops to significantly lower levels.

Column (2) of Table 5 is a specification that includes book characteristics rather than book fixed effects. Although this reduces the fit of the model and the significance of some of the months-since-release dummies, the magnitude of the coefficients is similar to that in the first specification. Most book characteristic variables have parameters that are significantly different from zero. There is less search for first edition books than for books in later editions. This may stem from the fact that later editions are only published if the book is popular. There are more searches for paperbacks than hardcover books, and again this may reflect the fact that paperbacks are targeted to the masses and therefore result in more searches.

The final column of Table 5 adds a linear and quadratic time trend to control for possible changes in search behavior between 2004 and 2013. Both coefficients are highly significant, suggesting that (apart from changes in search around release windows), the number of book searches at search engines has evolved. Figure 13 plots the number of searches predicted by the model for each year in our sample. This plot shows that book searches on Google increased between 2003 and 2007, but since then has been declined. This predicted pattern is largely consistent with the trend in average monthly retail sales for bricks-and-mortar bookstores, as shown in Figure 1, and may reflect the same structural changes in the book industry that have led to a decline in the sales of physical books at regular bookstores. For instance, the decline coincides with the increasing popularity of e-readers shown earlier in Figure 10, the general shift towards e-books, and the rising prominence of Amazon as a platform for searching for books and other products.

5 Book Prices

Another open question is how changes in search technologies and behavior have impacted the prices of physical and digital books. A priori it is not clear how prices should develop over time. In the short run, intermediaries may use penetration pricing in an attempt to induce consumers to switch from buying printed books to e-books. Once the market is more mature, and a sufficient number of consumers is locked into a specific platform, firms may find it optimal to increase prices.

We analyze book prices using a large dataset of daily prices collected using a scraper written

in Java. The dataset contains daily price information throughout 2012 for physical books as well as e-books sold at Amazon, Barnes and Noble, and Books-A-Million, physical books sold at Ecampus.com, Powell's Books, and Walmart, and e-books sold at Sony's e-bookstore, Kobo, and Google. In addition to prices we observe availability and several book characteristics such as publisher, format (hardcover or paperback) edition, number of pages, weight, and dimension.

Figure 14 gives kernel estimates of the price densities for hardcover books, paperbacks, and e-books. The prices used are average prices across books within format and across bookstores, so the variation reflects variation over time. E-books are on average less expensive than hardcover books, but slightly more expensive than paperbacks.

The relationship between publishers and retailers complicates the analysis of prices for e-books. Traditionally, books were sold using a wholesale model: the publishers sell a book to a retailer at a wholesale price, which is typically half the list price for the book. Retailers are then free to sell the book at whatever price they want. Initially this model was used by all players for selling e-books as well. However, to foster e-book sales, Amazon began heavily discounting popular e-books, selling them for \$9.99—a price that is below the wholesale price of many physical books. Publishers, fearing that e-books would cannibalize sales of printed books and that consumers would get used to low e-book prices, were not thrilled by low e-book prices. When Apple launched its ibookstore in April 2010, five of the “big six” publishers (HarperCollins, Hachette, Macmillan, Penguin, and Simon & Schuster) adopted Apple's newly developed agency model. Under the agency model the retailer gets a fixed percentage of the book's sales price (typically 30 percent). Moreover, under the agency model, the retail price is no longer set by the retailer but set by the publisher. Other online bookstores reached similar agreements with the publishers. As a reaction to these developments, the Department of Justice sued Apple and the five publishers on April 12, 2012, alleging that they conspired to fix the prices of e-books.

Table 6 provides descriptive statistics for the prices of books in our sample. The table makes a distinction between e-books that are sold under the agency model and those that are sold under the wholesale model. Average prices are generally higher for the agency model e-books, and show less variation.

Figure 15 plots the average prices of paperbacks, e-books sold under the agency model, and e-books sold under the wholesale model. Even though the DOJ reached a settlement with three of the publishers in April, 2012 (HarperCollins, Hachette, and Macmillan), and as part of the settlements e-book retailers are allowed to discount e-book titles, prices for e-books sold under the wholesale model increased shortly after the lawsuit was announced. The steep decline in e-book prices sold under the wholesale model a few months later coincides with the final approval of the settlement between the DOJ and the three publishers, with HarperCollins books starting to be discounted on Amazon soon after (Hachette and Simon & Schuster discounting began in December 2012). Figure 16 gives the average coefficient of variation over time for the three book formats, and shows that there was a decline in price dispersion following the approval. The two remaining publishers (Macmillan and Penguin) settled in early 2013. An open question is the underlying reasons for these pricing patterns.

6 Concluding Remarks

Our initial look at the evolution of platforms where consumers search for, purchase, and read books highlights a number of issues that should be on the agenda of those interested in these markets. There are a variety of different platforms where consumers can search for books, including online bookstores (such as Barnesandnoble.com), mass merchants (Amazon), and price comparison sites (Bookfinder.com). While data suggest that consumers are increasingly conducting searches on “closed systems” (such as e-readers) and retailer platforms (such as Amazon), measurement of these trends is difficult. In contrast to these platforms, data on the types of searches conducted on search engines is readily available.

While Amazon and Barnes and Noble get significant numbers of visits from those conducting searches on Google and Bing, most of the observed searches are navigational searches—situations where users query a search engine as a “shortcut” for navigating to the actual site of the bookseller where actual searches are performed. Navigational searches to these top booksellers has steadily increased going back to 2004. The number of browser-based searches on book-related platforms such as Amazon, Barnes and Noble, Book Finder, and Google Books are sizable, but it is an

open question whether browser-based searches on these platforms will continue to grow. Indeed, our preliminary analysis of searches for book titles on Google suggests a declining trend in book searches since 2007. While these patterns are consistent with shifts toward non-browser searches on closed platforms, including tablet/mobile apps, additional research is required to reach any definitive conclusions.

Another open agenda item is the importance of the long tail for product searches in general, and more specifically for book searches. The search terms corresponding to long-tail searches at the major search engines are not observed because of low search volume; we do not observe niche searches, which presumably take place in the long tail. Additional research will be needed to analyze the long tail.

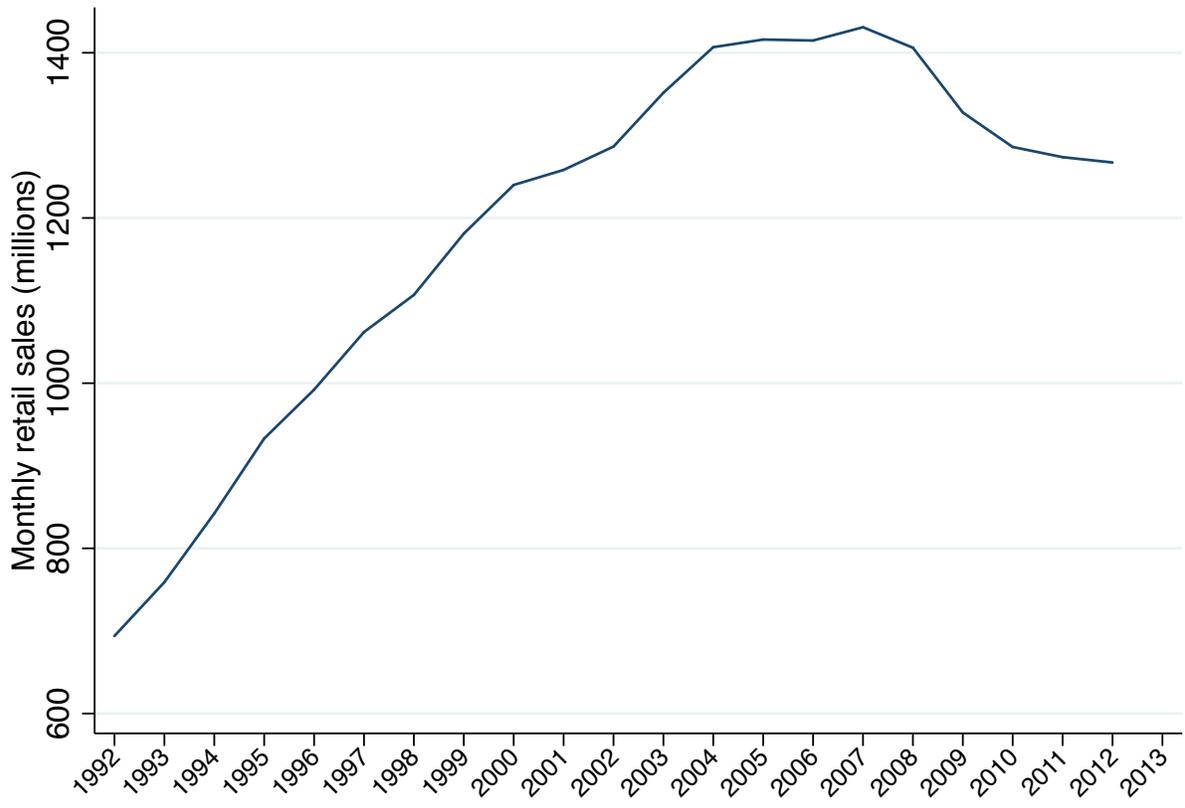
Finally, an interesting new area of research has emerged relating to e-books. We have identified several new developments with respect to pricing strategies of publishers and booksellers, and have provided some preliminary findings on how some of these developments have affected book prices. More research will be needed on how different selling formats (agency and wholesale model) affect the book industry, from both a theoretical and empirical point of view.¹⁹

¹⁹Abhishek, Jerath, and Zhang (2013) develop a theoretical model that is geared towards a comparison of the different selling formats in the book industry.

References

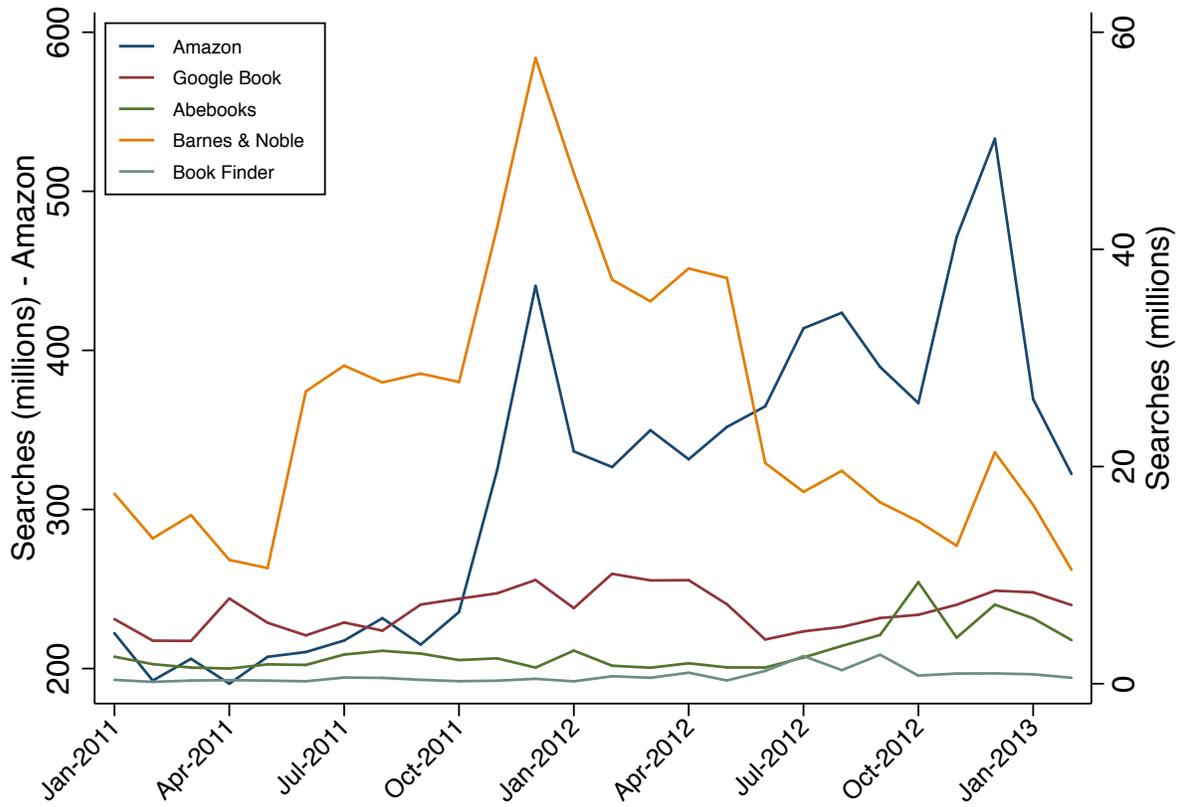
- [1] Abhishek, Vibhanshu, Kinshuk Jerath, and Z. John Zhang (2013). “Agency Selling or Reselling? Channel Structures in Electronic Retailing.” Working Paper.
- [2] Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Duncan Simester (2011). “Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales.” *Management Science* 57 (8), pp. 1373-1386.
- [3] Brynjolfsson, Erik, Yu (Jeffrey) Hu, and Michael D. Smith (2003). “Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers.” *Management Science* 49 (11), pp. 1580-1596.
- [4] Chevalier, Judith and Austan Goolsbee (2003). “Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com.” *Quantitative Marketing and Economics* 1 (2), pp. 203-222.
- [5] Chevalier, Judith A. and Dina Mayzlin (2006). “The Effect of Word of Mouth on Sales: Online Book Reviews.” *Journal of Marketing Research* 43 (3), pp. 345-354.
- [6] Choi, Hyunyoung and Hal Varian (2012). “Predicting the Present with Google Trends.” *The Economic Record*, 88 (s1), pp. 2-9.
- [7] Clay, Karen, Ramayya Krishnan, and Eric Wolff (2001). “Prices and Price Dispersion on the Web: Evidence from the Online Book Industry.” *The Journal of Industrial Economics* 49 (4), pp. 521-539.
- [8] D’Amuri, Francesco and Juri Marcucci (2010). “Google it! Forecasting the US Unemployment Rate with Google Job Search Index.” Working Paper.
- [9] De los Santos (2012). “Consumer Search on the Internet.” Working Paper.
- [10] De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest (2012). “Testing Model of Consumer Search Using Data on Web Browsing and Purchasing Behavior.” *American Economic Review* 102 (6), pp. 2955-2980.

- [11] Forman, Chris, Anindya Ghose, and Avi Goldfarb (2009). “Competition between local and electronic markets: How the benefit of buying online depends on where you live.” *Management Science* 55 (1), pp. 47-57.
- [12] Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant (2009). “Detecting Influenza Epidemics Using Search Engine Query Data,” *Nature* 457 (7232), pp. 1012-1014.
- [13] Guzmán, Giselle (2011). “Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations.” *Journal of Economic & Social Measurement* 36 (3), pp. 119-167.
- [14] Hong, Han and Matthew Shum (2006). “Using Price Distributions to Estimate Search Costs,” *RAND Journal of Economics* 37 (2), pp. 257-275.
- [15] Manley, Laura and Robert P. Holley (2012). “History of the Ebook: The Changing Face of Books.” *Technical Services Quarterly* 29 (4), pp. 292-311.
- [16] Rysman, Marc (2009). “The Economics of Two-Sided Markets.” *Journal of Economic Perspectives* 23 (3), pp. 125-143.



Source: U.S. Census Bureau data from the Monthly Retail Trade Survey for NAICS 451211 (bookstores).

Figure 1: Monthly Retail Sales at Brick-and-Mortar Bookstores



Note: Amazon searches scaled on left, all others on right.

Figure 2: Total Searches at Book-Oriented sites

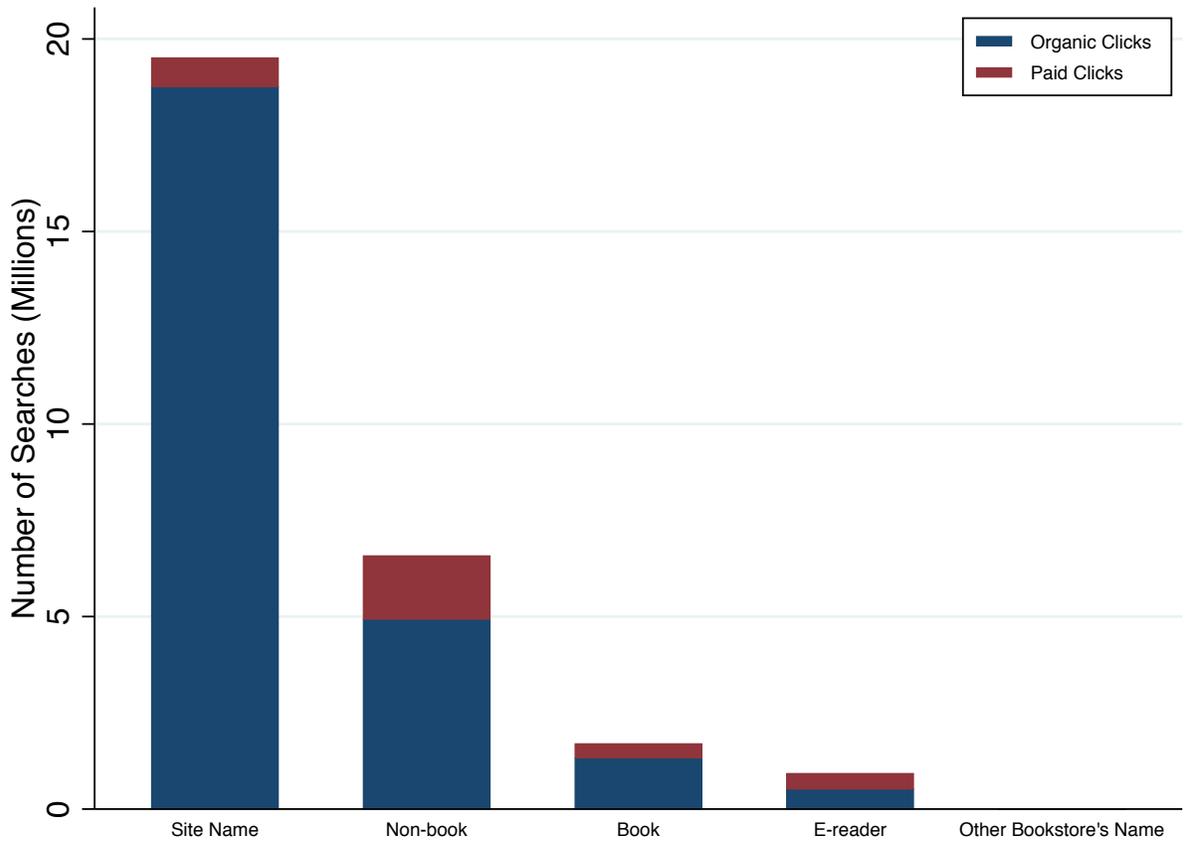


Figure 3: Search Terms in Main Search Engines Leading Users to Amazon

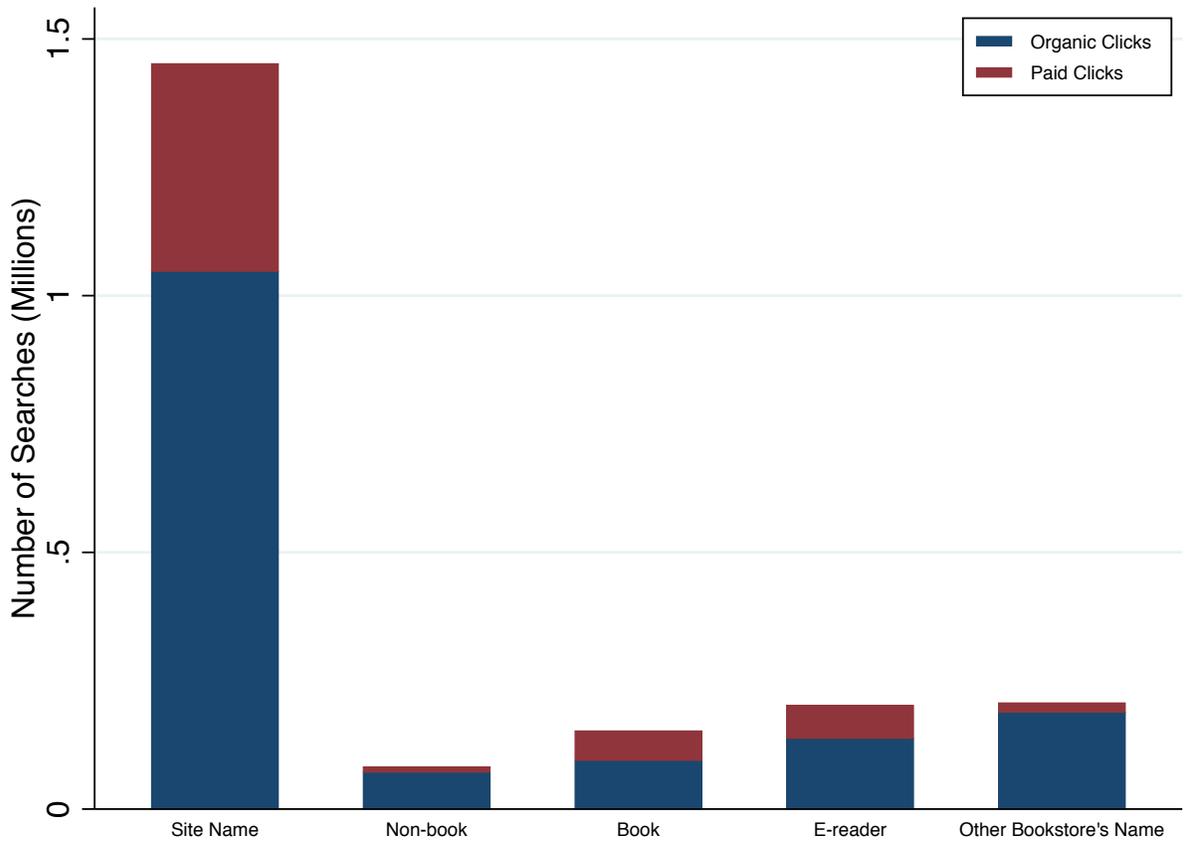


Figure 4: Search Terms in Main Search Engines Leading Users to Barnes and Noble

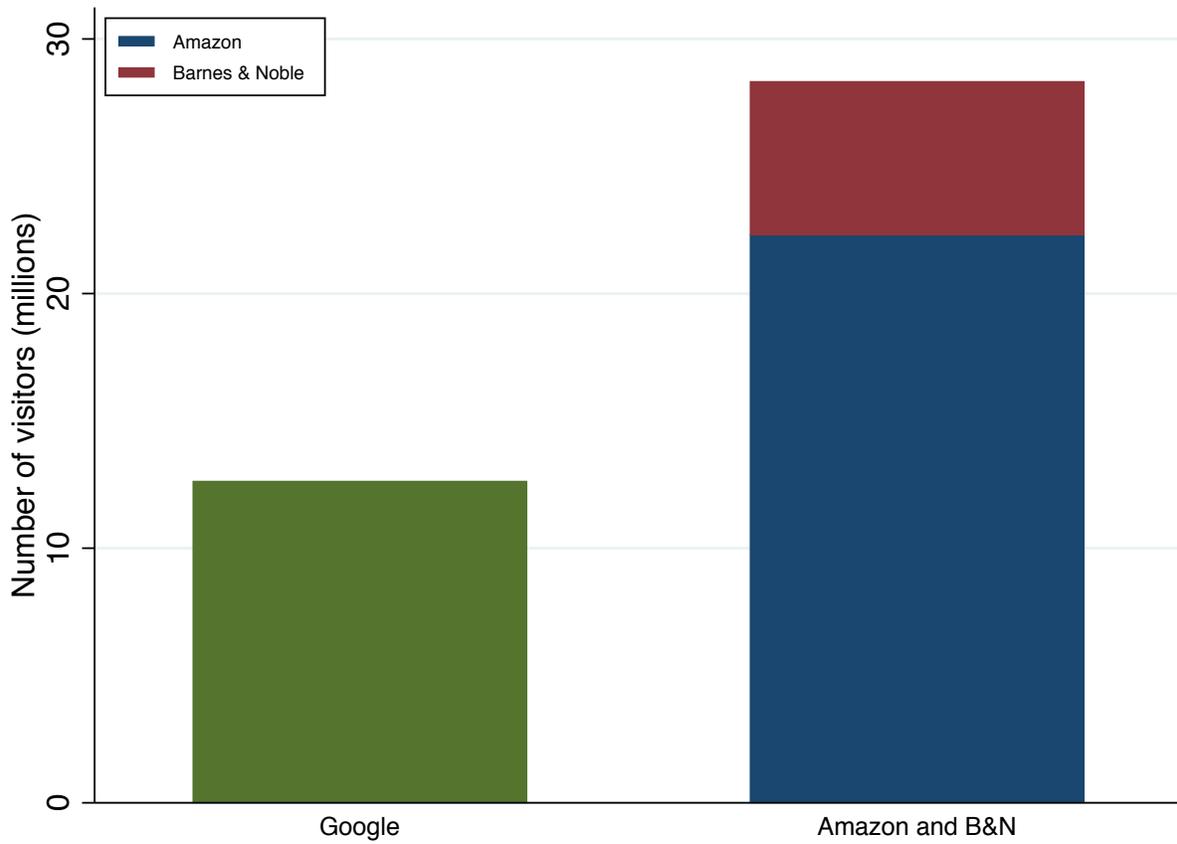


Figure 5: Book-Related Search Visits on Google, Amazon, and Barnes and Noble

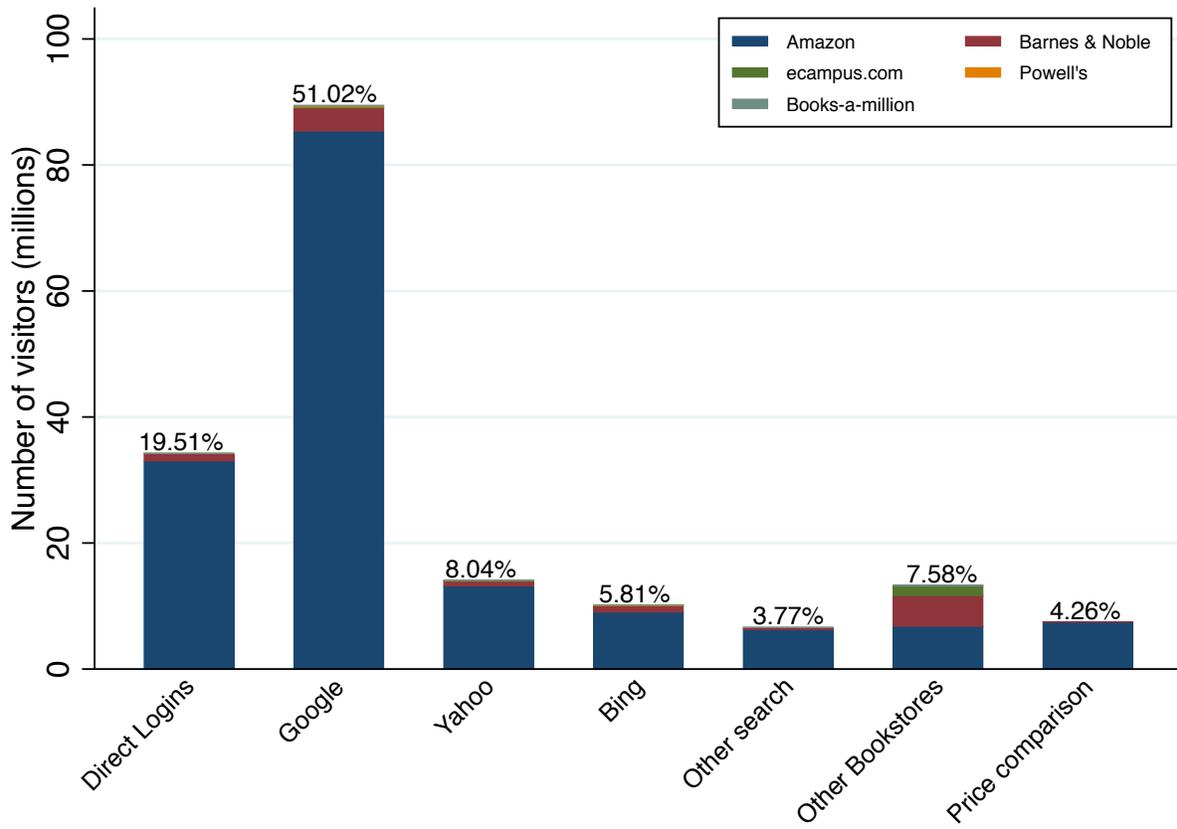


Figure 6: Total Incoming Visits to Bookstores from Search Engines and Other Platforms

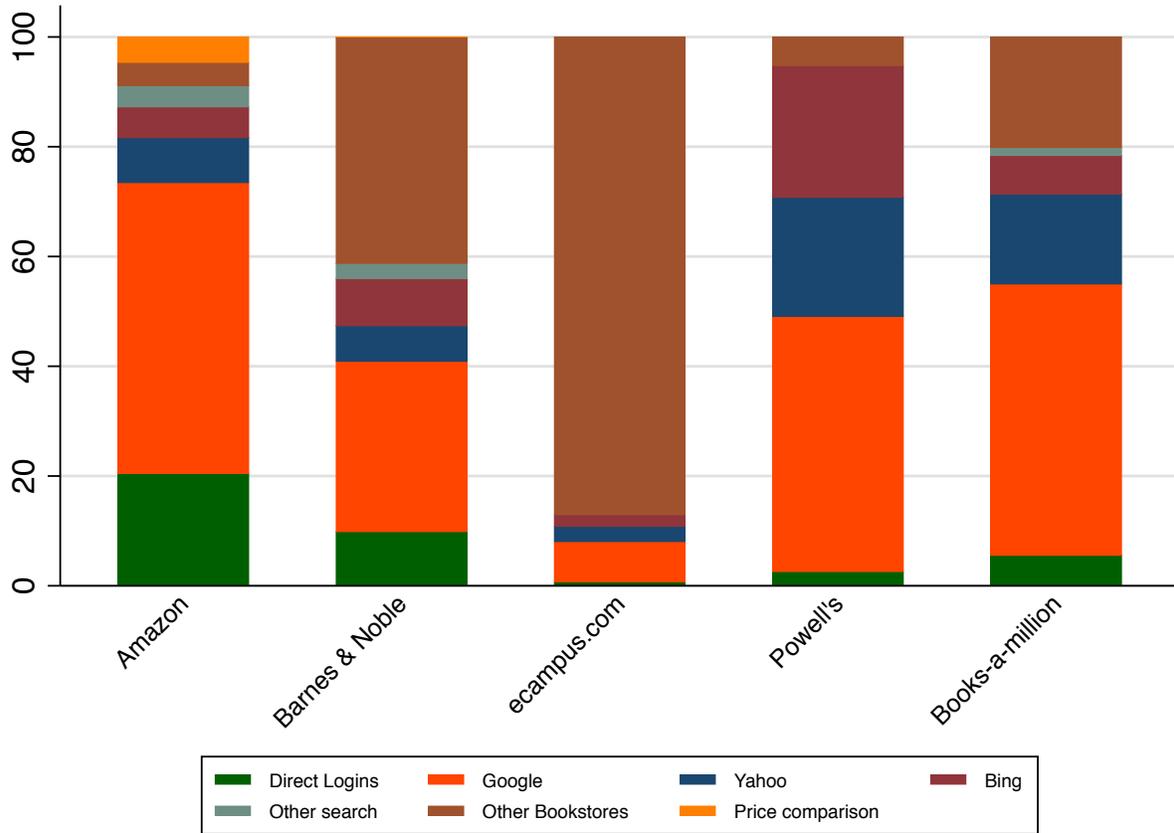


Figure 7: Percentage of Incoming Visits to Bookstores from Search Engines and Other Platforms

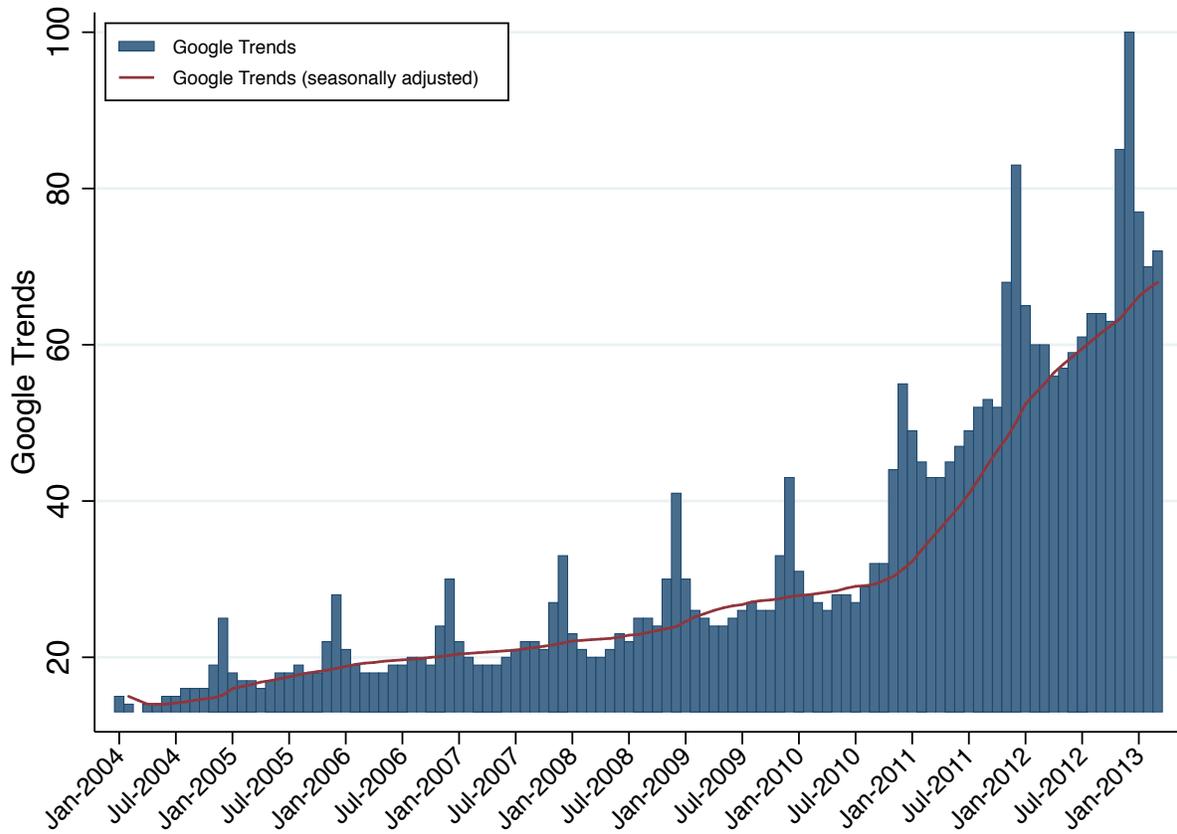


Figure 8: Google Trends for the Term “Amazon”

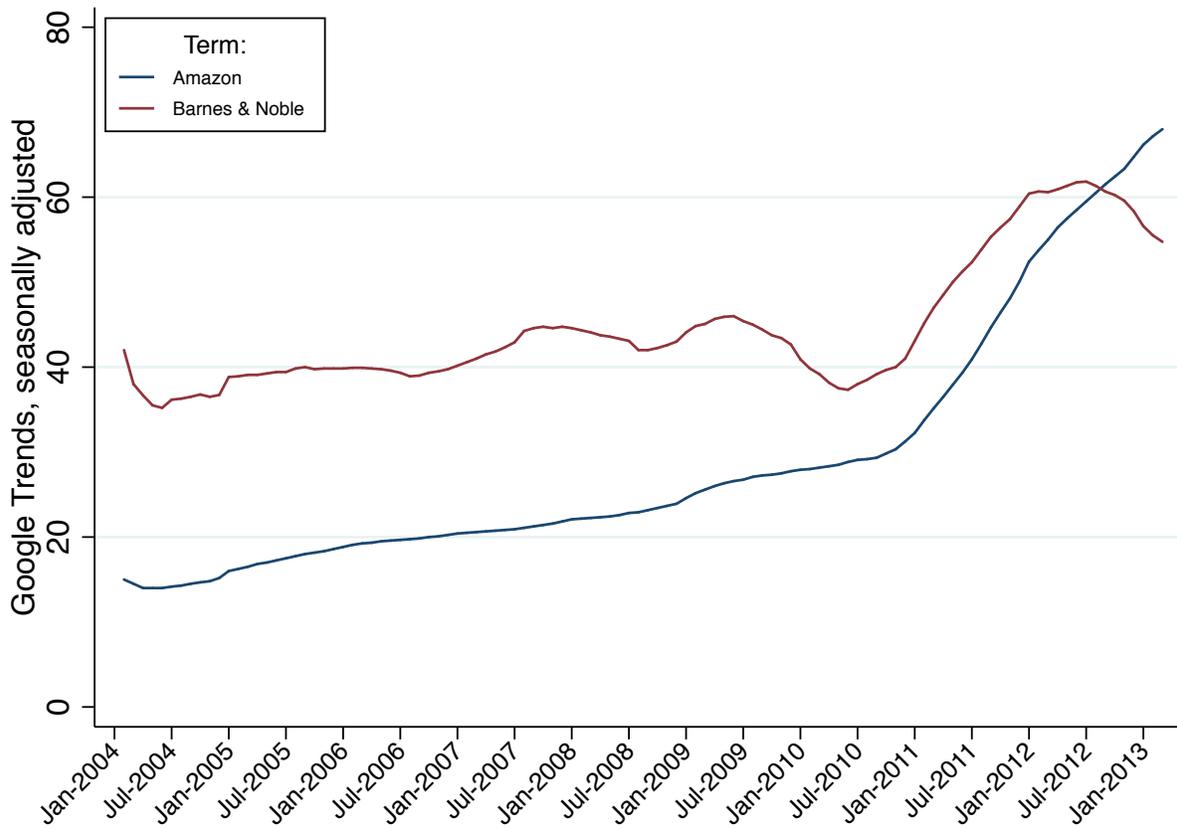


Figure 9: Google Trends for the Terms “Amazon” and “Barnes and Noble”

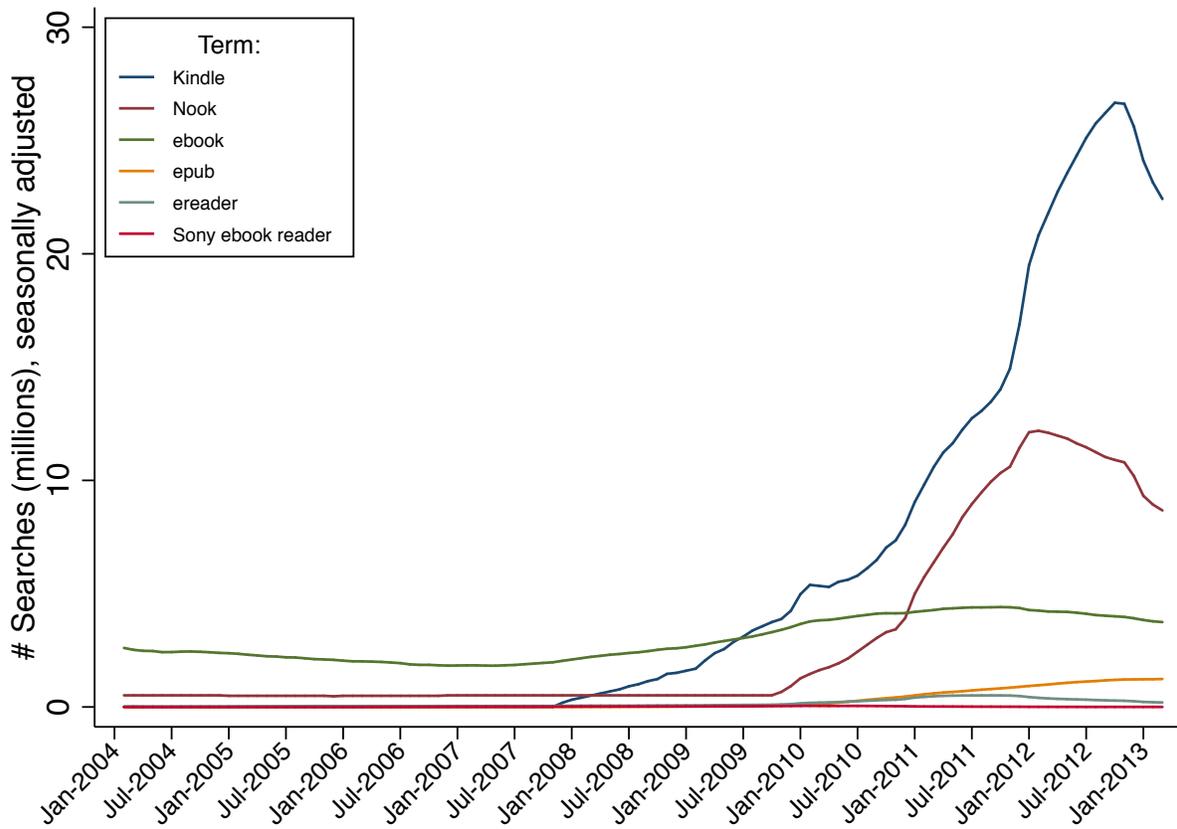


Figure 10: E-reader related search on Google, 2004-2013

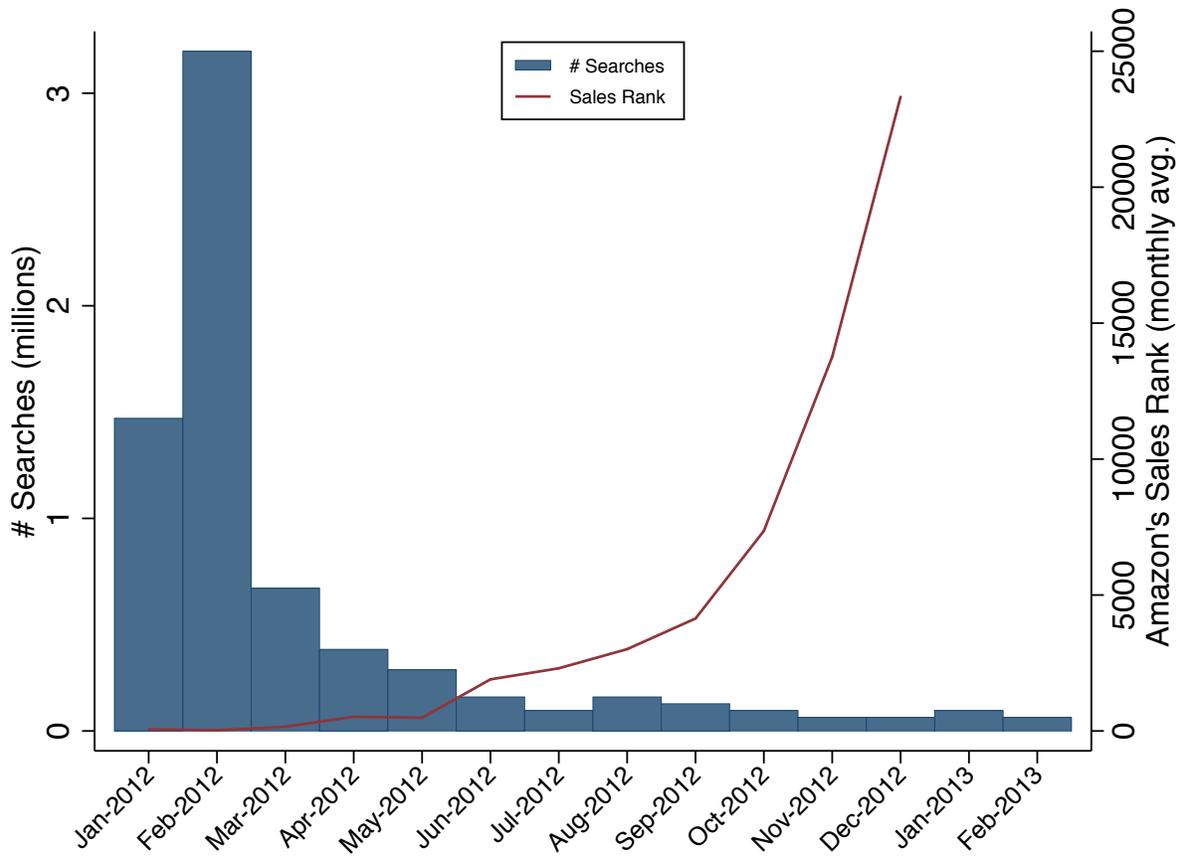


Figure 11: Google Searches for “The End of Illness”

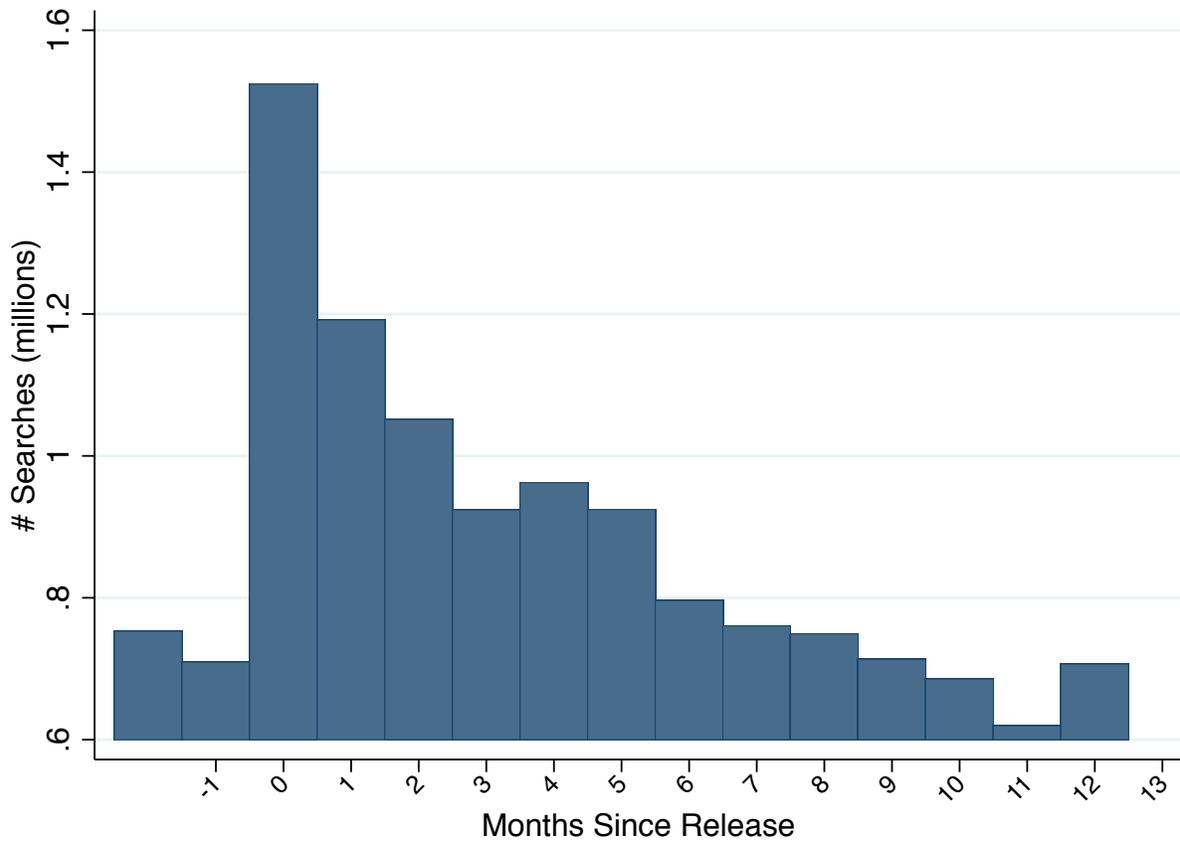


Figure 12: Predicted Searches for Book Titles on Google

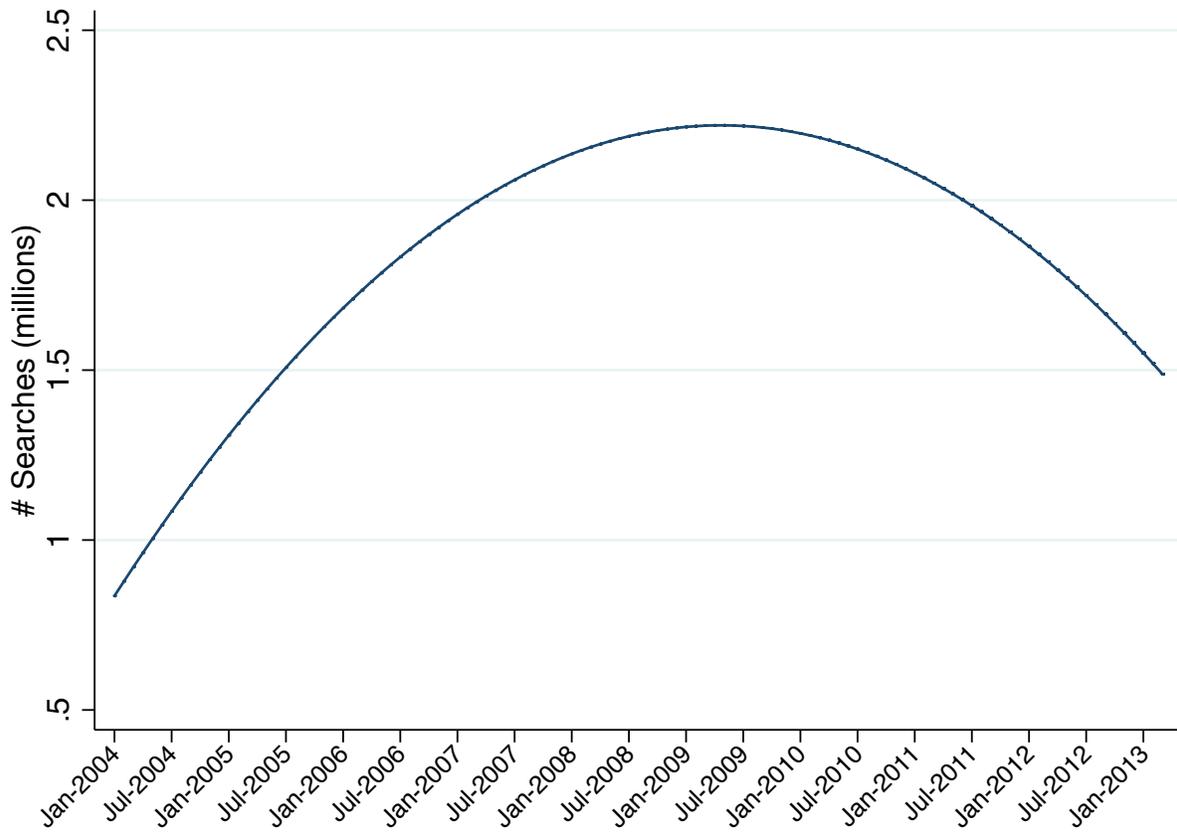
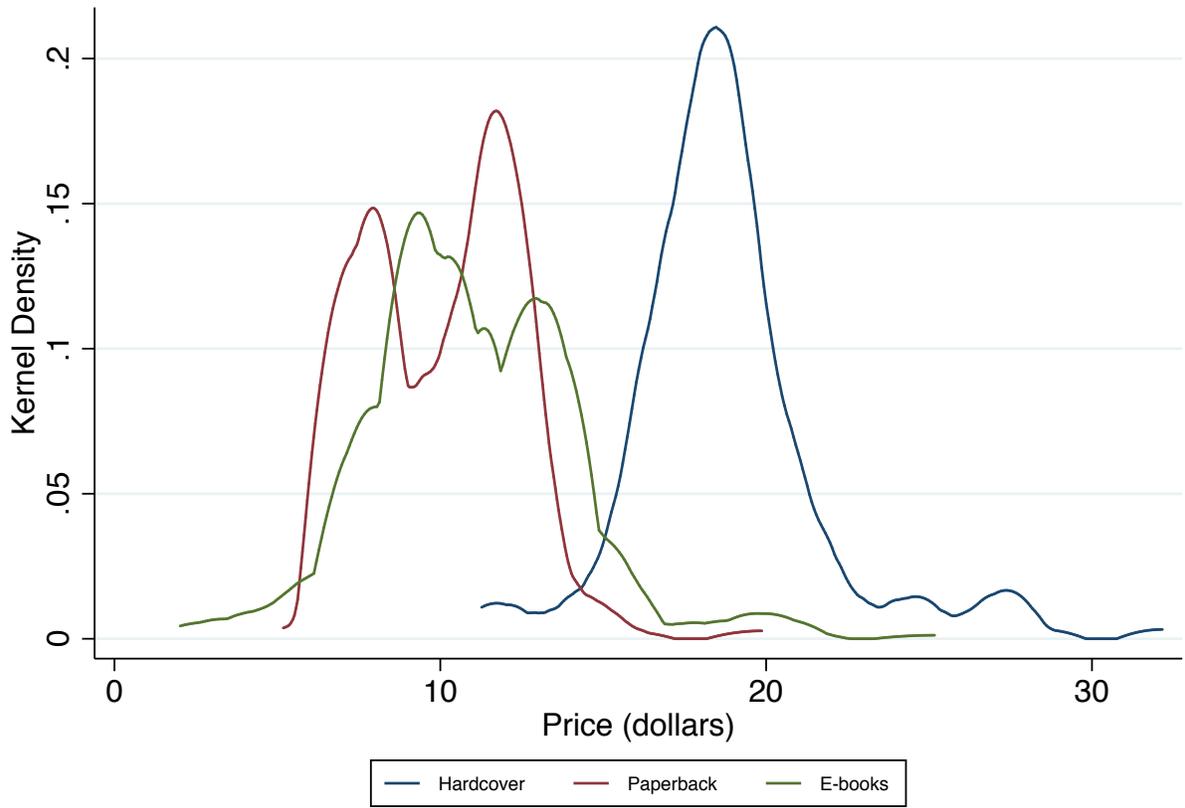


Figure 13: Predicted Searches for Book Titles on Google, 2004-2013



Truncated at 50.

Figure 14: Price Distribution for Hardcover, Paperback and Digital Books

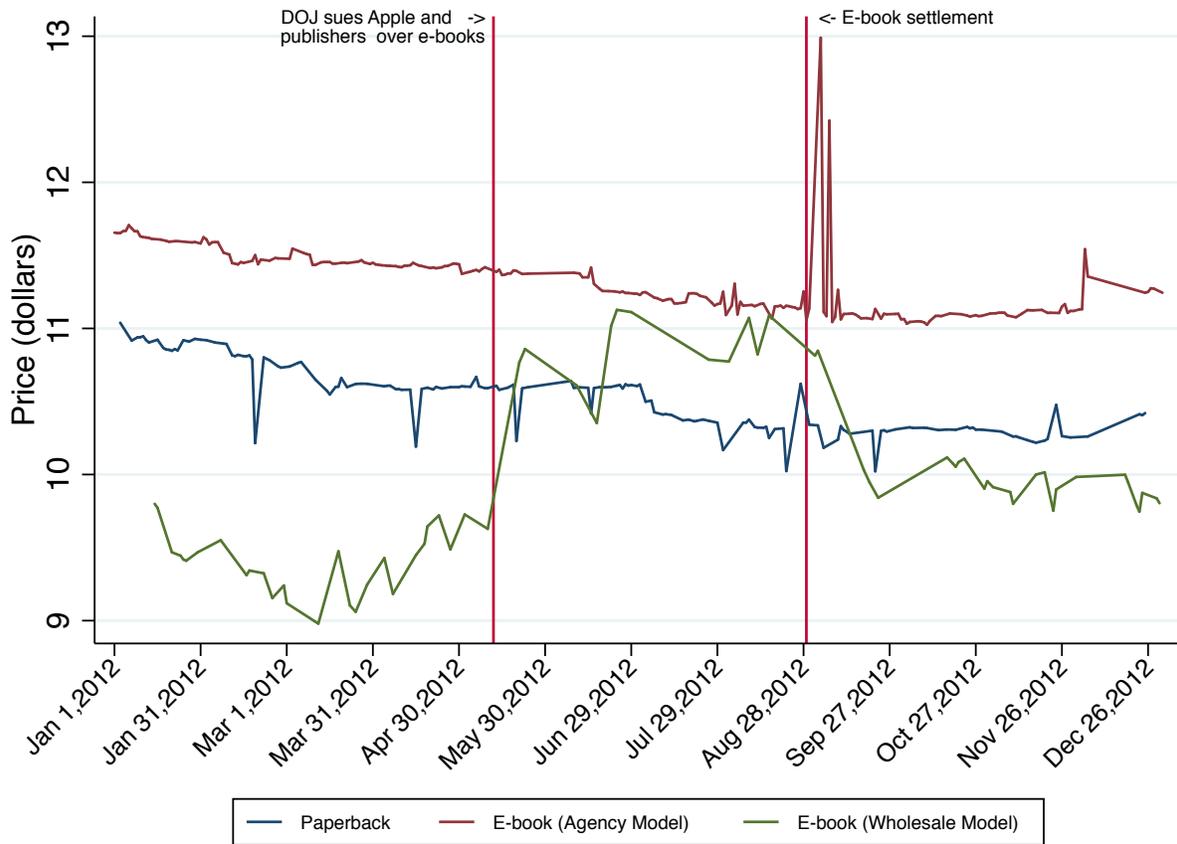


Figure 15: Average Book Prices, by Format

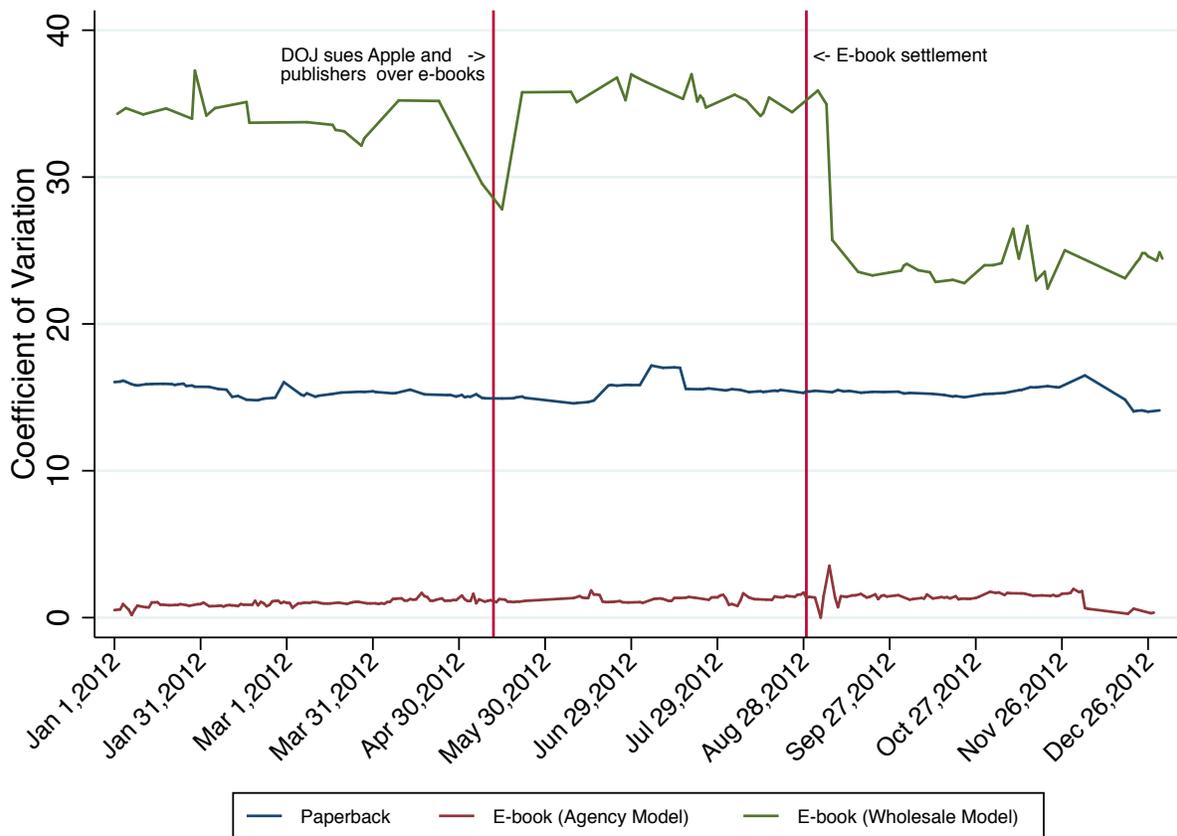


Figure 16: Average Coefficient of Variation, by Format

Table 1: U.S. Print and E-Book Sales 2006-2010

Year	Trade print		E-books		Total
	\$ in mln.	%	\$ in mln.	%	\$ in mln.
2006	11,123	99.5	54	0.5	11,177
2007	11,384	99.4	67	0.6	11,451
2008	10,831	99.0	113	1.0	10,944
2009	10,711	97.2	313	2.8	11,024
2010 (est.)	10,757	94.5	632	5.5	11,389

Source: Mintel/American Association of Publishers. E-book sales excludes educational and professional books.

Table 2: Revenue Leading Bookstores

	Revenue (\$ in mln.)					Change (%)
	2005	2006	2007	2008	2009	
<i>Bookstore Chains</i>						
Barnes and Noble	5,103	5,261	5,411	5,122	5,596	9.7
Borders Group	4,031	4,063	3,775	3,242	2,791	-30.8
Books-A-Million	504	521	536	515	509	1.0
<i>Online Bookstores</i>						
Amazon.com	3,046	3,582	4,630	5,350	5,964	95.8
Barnesandnoble.com	440	433	477	466	525	19.4
Borders.com	-	-	-	46	60	-
Booksamillion.com	28	26	27	25	24	-13.8

Source: Simba information, from company reports. Figures for Amazon.com are for U.S. books, music and video/DVD segment only. Up to 2008 Borders.com was powered by Amazon.com.

Table 3: Total U.S. Searches on Platforms, February 2012

	searches (×1,000)	searches (%)	search visits (×1,000)	search visits (%)	searches per search visit
<i>Search engines</i>					
Google Web Search	10,932,714	38.99	3,988,116	54.86	2.74
Bing Web Search	2,271,281	8.10	814,432	11.20	2.79
Yahoo! Web Search	1,983,978	7.08	877,060	12.06	2.26
<i>Auction sites</i>					
eBay	753,652	2.69	228,606	3.14	3.30
<i>Bookstores</i>					
Amazon	326,658	1.16	150,643	2.07	2.17
Barnes and Noble	37,205	0.13	10,620	0.15	3.50
Abebooks.com	1,659	0.01	561	0.01	2.96
Kobobooks.com	619	0.00	148	0.00	4.18
<i>Price comparison sites</i>					
Google shopping	82,654	0.29	39,930	0.55	2.07
Bing shopping	22,403	0.08	9,863	0.14	2.27
Pronto	10,183	0.04	7,825	0.11	1.30
Bookfinder.com	692	0.00	220	0.00	3.15
Pricegrabber	52	0.00	49	0.00	1.06
Shopzilla	13	0.00	9	0.00	1.33
<i>Other book sites</i>					
Google Book Search	10,124	0.04	4,719	0.06	2.15
Total Internet	28,040,241	100.00	7,269,658	100.00	3.86

Source: comScore qSearch, February 2012.

Table 4: Top 25 Google Search Terms for Barnesandnoble.com

Rank	Search Phrase	# Organic Clicks
1	barnes and noble	376,678
2	borders bookstore	82,689
3	borders book store	52,006
4	***	27,699
5	barnes and noble locations	21,666
6	name name	20,675
7	barnes and nobles	19,800
8	amazon	19,748
9	barnesandnoble	17,785
10	nook	13,483
11	facebook	10,680
12	barnes \%26 noble	9,900
13	nook tablet	9,775
14	barnes	8,623
15	borders.com	7,465
16	google	7,163
17	bn.com	7,118
18	books	6,923
19	borders	6,382
20	the vow book	6,199
21	name fire	5,143
22	bn	4,819
23	nook color update	4,564
24	gmail	4,383
25	www.bn.com	4,179
	All search phrases	2,886,126

Source: comScore Search Planner data from February 2012. Search phrases are ranked by the total number of organic clicks on Google.

Table 5: Regression Results

	(1)	(2)	(3)
<i>Months before release</i>			
2 months	-0.696 (0.188)***	-0.688 (0.281)**	-0.765 (0.282)***
1 month	-0.734 (0.188)***	-0.732 (0.281)***	-0.771 (0.281)***
<i>Months after release</i>			
1 month	-0.299 (0.188)	-0.297 (0.281)	-0.257 (0.281)
2 months	-0.426 (0.188)**	-0.424 (0.281)	-0.342 (0.282)
3 months	-0.540 (0.188)***	-0.537 (0.281)*	-0.413 (0.282)
4 months	-0.503 (0.188)***	-0.499 (0.281)*	-0.329 (0.284)
5 months	-0.540 (0.191)***	-0.515 (0.285)*	-0.311 (0.288)
6 months	-0.654 (0.195)***	-0.628 (0.291)**	-0.391 (0.296)
7 months	-0.681 (0.199)***	-0.581 (0.297)*	-0.318 (0.303)
8 months	-0.692 (0.202)***	-0.580 (0.301)*	-0.280 (0.308)
9 months	-0.725 (0.205)***	-0.612 (0.306)**	-0.277 (0.315)
10 months	-0.747 (0.209)***	-0.613 (0.312)**	-0.248 (0.323)
11 months	-0.798 (0.216)***	-0.600 (0.321)*	-0.212 (0.333)
12 months	-0.719 (0.221)***	-0.510 (0.328)	-0.095 (0.342)
First Edition Indicator		-0.813 (0.200)***	-0.820 (0.202)***
<i>Format</i>			
Mass Market Paperback		1.228 (0.249)***	1.467 (0.255)***
Paperback		0.757 (0.208)***	0.736 (0.208)***
Other		0.256 (0.365)	0.660 (0.377)*
Pages		-0.005 (0.000)***	-0.005 (0.001)***
Weight		0.097 (0.007)***	0.099 (0.007)***
Time Index			0.094 (0.024)***
Time Index squared			-0.001 (0.000)***
Constant	1.398 (0.133)***	1.049 (0.297)***	-1.344 (0.913)
<hr/>			
Book Indicators	Yes	No	No
Publisher Indicators	No	No	Yes
Observations	10,615	10,585	10,585
R-squared	0.664	0.213	0.214

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Standard errors in parentheses.

Table 6: Descriptive Statistics Book Prices

Bookstore	Hardcovers			Paperbacks			E-books (agency)			E-Books (wholesale)		
	Mean	Std. Dev.	CV(%)	Mean	Std. Dev.	CV(%)	Mean	Std. Dev.	CV(%)	Mean	Std. Dev.	CV(%)
<i>Both Formats</i>												
Amazon	16.3	5.7	34.7	9.9	6.3	63.7	11.2	2.7	23.8	8.6	7.5	86.6
Barnes and Noble	16.8	6.0	36.0	10.5	7.1	67.8	11.3	2.6	23.2	10.2	5.0	48.8
Books-a-million	16.7	6.6	39.8	10.2	9.7	94.6	11.4	2.5	22.1	9.9	4.2	41.7
<i>Physical Books Only</i>												
Ecampus.com	21.9	4.9	22.4	10.5	9.0	85.6						
Powell's	27.0	7.1	26.3	12.9	4.2	32.6						
Walmart	16.2	3.6	22.2	9.0	7.3	80.9						
<i>E-Books Only</i>												
Sony							11.4	2.5	22.2	10.0	3.8	37.7
Kobo							11.2	2.5	22.0	11.0	8.6	78.2
Google							11.1	2.4	21.9	10.2	9.0	88.6