

Essays on Market Effects in the Publishing Industry

**A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY**

Imke Christina Reimers

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy**

Thomas Holmes and Amil Petrin, Advisors

August, 2013

**© Imke Christina Reimers 2013
ALL RIGHTS RESERVED**

Acknowledgements

I was privileged to have an exceptional group of faculty and fellow students working with me throughout my time at the University of Minnesota.

I am deeply indebted to my advisors Thomas Holmes and Amil Petrin for their guidance throughout this project. Their constructive evaluation of my work and their continued support were invaluable. I am also very grateful for my committee member and co-author on the second chapter of my dissertation, Joel Waldfogel, for taking a special interest in my work, for his sincere encouragement and for working with me. I would also like to thank Kyoo-il Kim for his helpful comments as well as for being a member of my dissertation committee.

I benefited greatly from presenting my work on several occasions. I am grateful for many helpful comments from the members of the University of Minnesota Applied Micro Reading Group and Seminar and from the participants of seminars I gave while I was on the job market.

Finally, I would like to thank the support staff of the University of Minnesota Department of Economics for making all of our lives easier.

Dedication

To my family and friends.

Abstract

This thesis is composed of two separate essays about the market effects of different events in the book publishing industry.

In the first essay, I study the effects of a copyright extension on the book publishing industry. The 1998 Sonny Bono Act extended copyright by 20 years so that works created in 1922 and before are in the public domain while works created later remain on copyright. Taking works off copyright tends to promote their availability which benefits consumers. But it also allows entry to dissipate producer surplus, particularly for high-value works. Many copyright-protected books, especially those of low value, have gone out of print, and are effectively unavailable to consumers. Putting these works in the public domain can raise the welfare of both producers and consumers. Evaluating the welfare impact of the copyright extension requires estimation of the differences in consumer and producer surplus across a wide range of books between actual and counterfactual copyright regimes. I assess these differences using a structural model of demand for books, along with an entry model, to simulate the elimination of copyright for books published after 1923. I find that a copyright extension decreases welfare by decreasing variety, causing a decrease in consumer surplus that outweighs an increase in producer surplus. Holding fixed the costs of entry, a move into the public domain increases total surplus from most titles, indicating insufficient entry under copyright protection.

In the second chapter, I present joint work with Joel Waldfogel wherein we analyse the market effects of new technology - namely e-books - on the book publishing industry.

Digitization is transforming the market for books. Lower marginal costs have reduced prices by 10-15 percent in the past four years, and digitization has given creators the ability to circumvent traditional gatekeepers and publish their work directly. The number of self-published works has grown by almost 300 percent since 2006 and now exceeds the number of traditionally published works. Given the inherent difficulty in predicting the ex post appeal of creative products at the time of investment, a growth in available new products can substantially expand the appeal of available products. While e-book data are not systematically available, we are able to document that falling

prices have increased consumer surplus by \$2-3 billion per year. Using bestseller lists in conjunction with title-level data on physical sales and our best estimates of e-book sales, we document that many self-published books have substantial ex post appeal to consumers. Works that began their commercial lives through self-publishing began to appear on bestseller lists in 2011 and by late 2012 such works accounted for a tenth of both bestseller listings and estimated unit sales. In romantic fiction, self-published works account for almost a third. These changes challenge the role of gatekeepers while benefiting consumers.

Contents

Acknowledgements	i
Dedication	ii
Abstract	iii
List of Tables	viii
List of Figures	x
1 The Welfare Effects of Lengthened Copyright Protection for Books	1
1.1 Introduction	1
1.2 Data and Industry Background	5
1.2.1 Data	5
1.2.2 Preliminary Analysis of the Effect on Industry-Level Supply and Demand	8
1.3 Model	13
1.3.1 Demand	14
1.3.2 Supply and Equilibrium	16
1.4 Estimation and Results	20
1.4.1 Demand	21
1.4.2 Supply	27
1.5 Policy Analysis	33
1.5.1 Moving Titles into the Public Domain	34
1.5.2 Decomposition of the Welfare Effect: A Switch to Free Entry	37

1.5.3	Decomposition of the Welfare Effect: Project Gutenberg	40
1.6	Conclusion	41
2	Storming the Gatekeepers:	
	Digital Disintermediation in the Market for Books	43
2.1	Introduction	43
2.2	Background	46
2.2.1	The Traditional Publishing Industry	46
2.2.2	The Growth of E-Readers	46
2.2.3	Electronic Books and Lower Marginal Costs	48
2.2.4	The Appearance of Self-Publishing Channels	48
2.2.5	Product Discovery Institutions	48
2.3	Theoretical Framework	50
2.4	Data	52
2.4.1	Aggregate Data Sources	53
2.4.2	USA Today Bestseller List Data	54
2.4.3	Publishers Weekly Top 1000 E-books for 2012	54
2.4.4	Title-level Information on Physical Books	55
2.4.5	Title-level Information on Electronic Books	55
2.5	Descriptive Results	56
2.5.1	Evolution of Physical and E-book Quantities	56
2.5.2	Evolution of Prices	59
2.5.3	The Number of New Books	62
2.5.4	Sales Concentration	63
2.5.5	Are Self-Published Works Consequential	64
2.6	Digitization and Welfare	66
2.6.1	Surplus for Market Participants: Basic Approach	66
2.6.2	Differentiated Products Approach to Welfare	68
2.7	The Coarsening of Product Offerings and Consumption	73
2.8	Conclusion	76
3	References	77

Appendix A. Appendix to chapter 1	81
A.1 Obtaining Demand Data	81
A.2 Discontinuity in Availability and Prices	82
A.3 Market Shares and Elasticities	83
A.4 Imposing Marginal Costs by Format and Copyright Status	85
A.5 A Move into the Public Domain - Selected Titles	86
 Appendix B. Appendix to chapter 2	 88
B.1 Titles that were originally self-published	88
B.2 Alternative specifications for demand estimation	91

List of Tables

1.1	Composition of Editions	6
1.2	Summary Statistics for Quantities and Prices (annual numbers, by title)	8
1.3	Regression Discontinuity - In-print Versions per Title	11
1.4	Regression Discontinuity - Price per Edition	12
1.5	Demand Estimation	22
1.6	Quality Rankings - Top 10	24
1.7	Average Annual Demand per Edition, All Titles (no Gutenberg)	26
1.8	Marginal and Fixed Costs	29
1.9	Profits per Edition (in \$ per year)	32
1.10	Mean Welfare Changes by Creative Quality (in \$ per title over one year)	36
1.11	Mean Welfare Changes by In-Print Status (in \$ per title over one year)	36
1.12	Mean Mankiw Effect by Creative Quality (in \$ per title over one year) .	38
1.13	Mean Welfare Changes Without Zero-Price Options (in \$ per title over one year)	40
2.1	Major Book Review Outlets	49
2.2	Trade Book Sales by Book Format and Type, 2008-2011 (millions of units)	57
2.3	New Book Prices	60
2.4	New Titles (Baker & Taylor)	62
2.5	Total and Self-Published Titles	63
2.6	Increased Consumer Surplus from Price Reduction	67
2.7	Amazon Prices in 2013 - by Format and Publisher	69
2.8	Quantities Sold in 2012 - by Format and Publisher	69
2.9	Demand Estimates: Mean Price Elasticity is Set to -5	71

2.10	Compensating Variation, Million \$, Total in 2012 (Compared to no E-books and no Self-Publishing), Price Elasticity = -5	73
A.1	Marginal and Fixed Costs	85
A.2	Low-Quality Titles - All figures in \$	87
B.1	Originally Self-Published Works and their Sales, 2010-2012 ytd	89
B.2	Demand estimates: Mean price elasticity is set to -1	91
B.3	Demand estimates: Mean price elasticity is set to -2	91
B.4	Compensating variation, million \$, total in 2012 (compared to no ebooks or self-publishing), price elasticity = -1	92
B.5	Compensating variation, million \$, total in 2012 (compared to no e-books or self-publishing)	93

List of Figures

1.1	Number of ISBNs per Title: Copyright Effect = -28.053 versions	10
1.2	Average price per hardcover edition: copyright effect = insignificant . . .	13
1.3	Nested Logit Structure	15
1.4	Editions per title as a function of quality	25
1.5	Shares as a % of Original Shares after adding a Paperback Edition	27
1.6	Fixed Cost Bounds by Creative Quality - Hardcover	31
1.7	Profits by Creative Quality	39
2.1	Percent of US Adults with Tablets and E-readers	47
2.2	Growth of 'Mainly E-book' Entries in USA Today List	59
2.3	US City Average CPI for Recreational Books	61
2.4	Price Distribution for Top 100 Amazon Kindle Bestsellers	61
2.5	Number of Distinct Titles on USA Today List per Year	64
2.6	Share of Bestseller Listings Originally Self-Published: All Genres	65
2.7	Share of Bestseller Listings Originally Self-Published: Romance Genre . . .	66
2.8	Stephen King's Cumulative Page Output Since 1977	68
2.9	Number of NYT Notable Titles on USA Today List	75
2.10	NYT Notable Titles' Share of USA Today Estimated Sales	75
A.1	Sales Ranks, Arrowsmith (Mass Market Paperback)	82
A.2	RDD: High-Popularity Works	83
A.3	RDD: Low-Popularity Works	83
A.4	Fixed Cost Bounds by Creative Quality	86

Chapter 1

The Welfare Effects of Lengthened Copyright Protection for Books

1.1 Introduction

Throughout recent history there has been ample discussion and research among policy makers and scholars about the length and terms of a copyright. The copyright term has been extended 11 times since 1909. Most recently, the 1998 Sonny Bono Copyright Term Extension Act¹ extended the copyright term from 75 years (or "life of author + 50 years" if the work was created after 1978) to 95 years (or "life of author + 70 years" if the work was created after 1978). As a result, there is a cliff at 1923 where copyright drops off to zero: works that were created after 1922 remain under copyright protection while works created earlier are in the public domain. What is the welfare effect of this copyright extension? An extension cannot affect creative incentives of authors working 75 years earlier. The impacts instead operate through the works' availability, prices and edition qualities.

While a copyright - a legal grant of monopoly power to intellectual property - is

¹Also known as the Mickey Mouse Protection Act as it keeps Walt Disney's Mickey Mouse character under copyright protection until 2023.

understood to provide incentives for creation (see, for example, Williams (2010)), copyright monopolies have potentially harmful impacts on consumers. The distribution of protected works is controlled by the copyright holding monopolists. Potential publishers of editions (edition managers) of protected titles face an additional cost of obtaining the license to publish. Also, many copyright-protected titles have gone out of print and have owners who cannot easily be identified. These "orphan" works are effectively unavailable to consumers, at least in the new book market. Thus, eliminating their copyright protection could benefit both producers and consumers by making publishing a title easier. Producers can freely enter into producing editions of public domain titles. Moreover, new technology enhances the welfare impact of being in the public domain: institutions such as the Gutenberg Project and Google Books make digital copies of public domain titles available for free online².

Even for works that already exist, however, the removal of copyright is not unambiguously beneficial *ex ante*. The literature on the welfare properties of free entry (e.g. Mankiw and Whinston (1986), Dixit and Stiglitz (1977)) indicates that while entry can benefit consumers through lower prices and increased variety, it can also dissipate producer profits if fixed costs are present. The overall welfare impact of removing copyright protection thus becomes an empirical question that requires quantification of both the cost to producers and the benefits to consumers.

The point of departure for this paper is a growing literature that examines the impact of copyright protection via comparisons of works that are similar except in their copyright status by comparing works published before and after a certain "cliff" date. Heald (2007) documents that coming off copyright substantially raises the probability that a title will be in print. My descriptive results match his findings. Li, MacGarvie, and Moser (2012) do something similar with respect to the price effects of the Copyright Act of 1814. They find that prices are higher if a title is on copyright. The descriptive results are informative and suggestive of welfare impacts, but a full welfare assessment requires measurement of consumer and producer surplus under existing and counterfactual regimes, and this requires a structural model of supply and demand.

This paper thus adds several dimensions to the existing literature. First, it accounts

²Google's efforts to put the orphan works online attracted litigation from the American Association of Publishers and the Authors Guild. See, for example, <http://www.googlebooksettlement.com> for more information on the issue.

for differences in market structure and the possibility of excess entry. Second, it allows for differences in the creative quality of each work: the title's (the creative work's) appeal. For instance, a title like *Gone With the Wind*, a "timeless classic" written in 1936, has a higher creative quality than most other books in the studied time frame. It will be produced regardless of its IP status. There will also be titles with a very low creative quality. Consumption of these works will be negligible whether they are protected by copyright or not. For titles whose quality lies in between these two types, their copyright status should significantly affect their availability and consumption. Third, the paper quantifies the effect on the publishing industry by simulating a counterfactual market situation when individual protected titles move into the public domain.

The work utilizes the discontinuity at the cliff date by looking at fiction bestseller books that were written around 1923. The data set contains each title's in-print status as well as edition specific sales data of a 12 month period in 2011 and 2012. It follows the discrete choice demand literature in estimating a repeated nested logit model to recover substitution patterns between hardcover, paperback, and e-books, as well as to obtain estimates of a work's creative quality (the quality of the title itself). Given the estimates from the demand model, my entry model allows me to obtain marginal and fixed costs of publishing an edition on the title-format level conditional on the title's copyright status and market structure. My analysis here builds on Berry and Waldfogel (1996) as well as Mazzeo (2002) by introducing product differentiation into a system that includes a structural profit function and allows me to estimate entry decisions. I use the exogenous *creative* quality of a title together with the number of existing editions as a determinant of each title-format combination's cost structure with a moment inequalities approach that builds on Ishii (2005) and Pakes, Porter, Ho, and Ishii (2011). I use this framework to determine the relative effects of free entry (the public domain) on consumers and publishers. I find that the movement of a title into the public domain increases total surplus by increasing consumer surplus by a larger amount than it dissipates producer surplus. Even holding fixed costs constant across regimes, I find that the Mankiw and Whinston (1986) problem of excess entry does not exist among most titles although it does hold for some titles.

More specifically, I find that removing a copyright increases consumer surplus from

a title by between \$200 (less popular works) and \$50,000 (more popular works) annually. About one third of this change in consumer surplus stems from the availability of digitized editions that are offered free of charge through channels such as Google Books or Project Gutenberg. While the increase in consumer surplus is small (on the order of 0.01 cents per potential reader), it is still larger than the dissipation in producer surplus that results from publishing additional editions of a title. The decreases in profits to incumbents range from \$100 to \$37,000, depending on the title's creative quality. They are the result of two counteracting forces. On one hand a move into the public domain decreases the costs of introducing an edition, potentially increasing profits. But on the other hand a move into the public domain implies a free-entry system, which dissipates profits from each edition. Isolating the effect of the switch to a free-entry regime shows that the Mankiw effect of excess entry holds only for a few of the titles in my dataset (titles that already have many editions and face a low demand), while an overwhelming majority of the titles would benefit from a move into the public domain.

For titles that are currently out of print, a move into the public domain makes the title available again - at least through channels such as Project Gutenberg or Google Books. In this case, both producer surplus and consumer surplus increase unambiguously, thus increasing welfare.

My dataset is biased towards titles that are in print regardless of their IP status as I consider only titles that were popular upon original publication. However, most titles that were written in the time frame that I observe have since moved out of print. The quantitative effect of a copyright term extension on the industry will have to distinguish between the two types of titles. My results indicate that a one-year copyright extension would decrease total surplus from the publishing industry by \$10 million to \$20 million each year.

The remainder of the paper proceeds as follows. Section 2 discusses the industry and dataset. I also present a descriptive analysis that shows the effect of a title's IP status on its availability and price. In section 3, I present a model of discrete choice and entry. Section 4 gives a quick overview of the estimation strategy and presents the results. Section 5 presents a policy experiment in which I re-solve the model after changing the copyright term. I conclude in section 6 with a short discussion of policy implications.

1.2 Data and Industry Background

1.2.1 Data

To examine the effect of the Sonny Bono Copyright Term Extension Act (CTEA) of 1998 and the subsequent extension of a monopolistic market structure on the variety of products of a particular title, I consider a set of book titles that can a priori be regarded as similar: the annual ten bestselling fiction titles of the years 1910 to 1936³. While this is neither an exhaustive list of high quality titles from the time period nor of the most popular books given today's demand, it is likely to include several titles that still face positive demand today, as well as titles that may have become orphan works - works where the copyright holder cannot be found and nobody feels responsible for distributing these titles.

Upon publication, an edition of a book title is assigned an international standard book number (ISBN). An ISBN uniquely identifies a book's title, its publisher, edition, date of publication, format and suggested retail price. Titles are offered in different formats. In my analysis, I focus on hardcover, paperback and e-book editions. Each title-format combination can be offered in several ISBNs, or editions⁴. I observe, on average, approximately four ISBNs per title and month. These ISBNs are grouped into title-format pairs. The choice set consists of 1026 ISBNs, which add to 747 possible title-format combinations⁵. I observe the quantity demanded and several measures of availability of each of the ISBNs and title-format combinations on a monthly basis from July 2011 to July 2012.

I collect each ISBN's availability and price, as well as its format (hardcover, paperback, e-book) and available conditions (new and used) as found on Amazon's website, at the Project Gutenberg website and in the Bowker Books-in-Print directory. I also observe information on the size of today's publishing industry and wholesale prices through the Book Industry Study Group BookStats report of 2011, as well as through publications of Publishers Weekly periodicals. While Amazon is the biggest internet retailer of physical books, Project Gutenberg is a reasonably well-known website that

³These can be found on several websites, but Korda (2001), was used as a reliable reference

⁴I use the terms *ISBN* and *edition* interchangeably throughout the paper.

⁵That is, 249 titles, each in three formats. Some of the titles made the top ten best seller lists in multiple years.

offers e-book versions of public domain works for downloads free of charge ⁶.

The Bowker directory is an exhaustive list of all ISBNs that have been issued since 1948. It includes the issue date, the current in-print status, suggested retail price, format and publisher of each ISBN. I also collect the in-print status and the lowest price offered by publishers in the Bowker directory, and I determine the degree of the works' availability at Project Gutenberg. Table 1.1 summarizes the composition of observed editions in a sample month (October 2011). Note that not all ISBNs are offered in each condition, while some are offered as new and as used copies.

Table 1.1: Composition of Editions

Variable	Mean	Std. Dev.
Hardcover	0.3580	0.4797
Paperback	0.4938	0.5003
E-Book	0.1481	0.3555
New	0.5593	0.4968
Used	0.7877	0.4092
Age	30.738	32.778
Major publisher	0.3383	0.4734

Summary stats for 810 editions, October 2011

It is easier to find each of the books in the dataset in used condition than it is to obtain them new, with almost 80% of the editions being available as used editions and only 55% being available in new condition. I observe quantities demanded for all conditions of each ISBN, although I am not able to tell which condition an individual consumer obtained unless a title is available in only one of the two conditions. Moreover, e-books can be found almost exclusively among public domain titles as titles that are protected by copyright are more likely to be available only as physical editions. The composition of formats in my dataset reflects the true composition of available formats as observed through Bowker's Books-in-Print directory well. The average age of an edition is 30 years. It varies between 0 years and over 100 years. Of the editions in the dataset, only 34% are available through a major publisher. This also reflects the fact that only a small fraction of title-format combinations have multiple editions published

⁶comparable to Google Books

by the same publisher. Looking at imprints of publishers, less than 2% of the title-format combinations (11 out of 747) have multi-edition imprints. When considering the major publishing companies, the number rises to 3.5% (26 out of 747 title-format combinations). This indicates that publishers (edition managers) compete with each other after the entry decisions have been made.

I collect monthly data on the number of books sold through Amazon. I observe hourly sales rankings of each ISBN, as collected from April 2011 to August 2012. Since the quantity demanded of the titles in my dataset is quite low, an improvement in the ISBN's ranking over the previous hour's ranking is approximately equivalent to the sale of one book⁷. I also obtain the monthly number of books that are downloaded at Project Gutenberg as listed on each title's website. The monthly download count allows me to obtain a distribution of title qualities over the industry as well. Moreover I obtain annual library lending figures for the titles between 2006 and 2010 in the British library system. Since authors are compensated for the consumption of books through the library system in Britain, these numbers are available there, while they are not available in the United States. However, the number of titles that have been checked out through the library system is very small⁸, so that I collect these data over a longer time frame (namely, four years) in order to get more telling information.

Prices on the Amazon website are collected on a monthly basis, while the Bowker Books-in-Print directory lists each ISBN's one-time suggested retail price. I track each of those. My demand estimation uses the monthly Amazon price as it is generally uncommon that a book in my dataset is sold at its list price. Table 1.2 shows quantities demanded through each of the three channels, as well as the two measures of the editions' prices on the annual title level. While there is an average of 144 sales of a title through Amazon (at a positive price), an average title on Gutenberg is downloaded (free of charge) almost 1000 times each month. This suggests a rather large price elasticity as the mean percentage change in quantity per unit change in price translates into a mean pseudo price elasticity of -9.68%.

⁷see appendix A.1 for a more detailed description of the methodology

⁸This is either due to differences in tastes across the two countries, or it indicates that libraries are not as big as, e.g., Amazon.

Table 1.2: Summary Statistics for Quantities and Prices (annual numbers, by title)

Variable	Mean	Std. Dev.	Min	Max
Bowker Editions per Title	18.185	20.985	0	146
Amazon Title Sales	101.428	364.930	0	3289
Gutenberg Title Downloads	931.083	1419.18	114	11417
PLR 4-Year Check-Outs	133.207	645.905	0	6695
Amazon Price (in 2011 \$)	21.896	11.101	5.49	98.00
List Price (in 2011 \$)	33.279	35.021	1.14	100.00

Lastly, the BookStats wholesale prices give me a first estimate of marginal costs of a title that is protected by copyright. While the wholesale prices for hardcovers and paperbacks give sensible approximations of marginal costs at \$8.72 and \$4.79, respectively, the wholesale price of \$7.72 for e-books is implausibly high⁹. The marginal cost of distributing an electronic edition is expected to be close to zero, apart from royalties paid to the copyright holder.

1.2.2 Preliminary Analysis of the Effect on Industry-Level Supply and Demand

This section shows the size and significance of the effect of a copyright on the number of editions per title that are in print (that is, available in the new books market) and on the average prices per edition. While a copyright significantly affects the availability of editions, its effect on prices is significant only for e-books.

The Copyright Term Extension Act of 1998 offers an experiment that documents the differences between the two market structures (competition and monopoly). This "treatment" is a deterministic function of the year of publication with a sharp jump at 1923. Any titles that have been published after 1922 have been treated by the copyright extension and hence their availability and variety are determined by a copyright holder, while titles from before 1923 are not treated and individual editions face competition in a free-entry system. I tighten the bound around 1923 to include only titles that were published between 1915 and 1930. We can assume that the titles I consider are similar *a priori*, conditional on their copyright status. All titles are fairly "old" (published

⁹See, for example, the developments of the collusion case on e-books.

between 82 and 97 years ago) and have been fairly successful in the United States upon their original publication. All titles are fiction novels that can be considered among the higher quality titles of their time. Moreover, the 1998 CTEA was unexpected at the time the titles were created. These facts suggest that the conditional mean of any exogenous variable we examine is left and right continuous at the year in which the treatment occurs (1923). Following, among others, Imbens and Lemieux (2008), the equation

$$y_j = X_j' \beta + \alpha 1_{\{IP=1\}} + k(\text{Year}_j) + \epsilon_j$$

where X_j includes title specific characteristics and $k(\text{Year}_j)$ is a function of the year of original publication provides information on the immediate effect of copyright on indicators of supply through the coefficient α . In particular, I examine the number of in-print editions in each format while controlling for a work's creative quality by including British library checkouts and an indicator variable that equals one if the author has won a Pulitzer Prize.

The RDD setup shows the negative effects of copyright on availability. The total number of ISBNs in print (i.e., available through some publisher in new condition) per title is decreased by 28 versions when a title remains under copyright protection. Figure 1.1 illustrates this trend clearly. It plots the average number of editions per title for the bestsellers I observe in each vintage. It also shows a fitted line with a polynomial function of publication year as the explanatory variable.

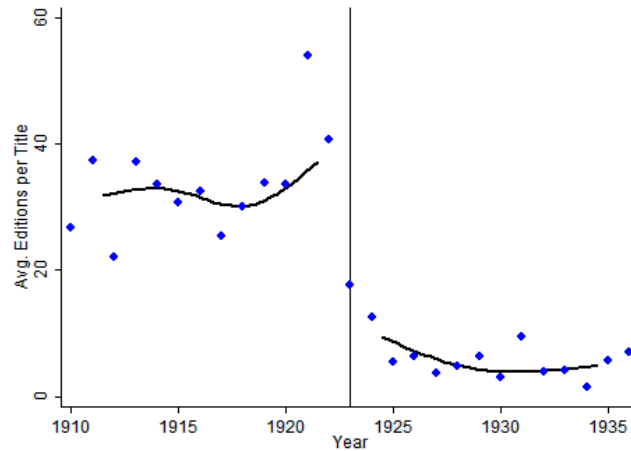


Figure 1.1: Number of ISBNs per Title: Copyright Effect = -28.053 versions

The effect can be analyzed by format. Table 1.3 shows results of a regression of the variety of each format on a title's quality and availability through different channels. Paperbacks carry around 16.6 of the "lost" versions, while hardcovers make up for around 6.5 versions, and the availability of e-books decreases by the remaining 4.9 versions per title. Treating the quantity demanded through British libraries (public lending rights, or PLR) as an indicator of quality, its positive effect is expected, although the magnitude of the effect is small. The substitution patterns between formats play an important role in determining the exact effect on each of the formats. These substitution patterns - and the effect of a title's creative quality - are taken into account and analyzed in the full model. The full model also quantifies the costs associated with obtaining the right to publish a book in a monopoly as opposed to a competitive market structure.

Table 1.3: Regression Discontinuity - In-print Versions per Title

	(1)	(2)	(3)
	Hardcover	Paperback	E-Book
Copyright	-6.517*** (1.572)	-16.60*** (3.344)	-4.936** (2.489)
PLR Demand	.0028*** (.0010)	.0058*** (.0010)	.0043*** (.0006)
Pulitzer Prize	2.439*** (.7316)	4.804*** (1.597)	3.238*** (.9917)
Constant	9.085*** (1.140)	18.30*** (2.219)	1.224 (.8613)

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

An RDD analysis of prices, on the other hand, shows significant differences based on a title's copyright status only among e-books. Table 1.4 shows regressions of price on an edition's format and copyright status. Average prices of physical editions in my dataset seem to be set independently of a title's copyright status when the title is rather obscure to begin with after controlling for an edition's quality¹⁰. This indicates that the price setting decision is likely to be made in similar fashion across copyright regimes. The formal supply model takes this into account.

¹⁰Note that I use Amazon prices instead of the suggested retail prices. Especially with the increased visibility of prices through, e.g., e-commerce books are hardly ever sold at their suggested retail prices.

Table 1.4: Regression Discontinuity - Price per Edition

	(1)	(2)	(3)	(4)
	Hardcover	Paperback	E-Book	All
Copyright	4.433 (7.839)	-0.401 (1.682)	3.724*** (0.170)	1.436 (2.134)
PLR Demand	0.018*** (.0031)	-0.0020*** (.00065)	-0.00010** (.00005)	0.0036*** (.0008)
Pulitzer Prize	-2.882 (6.316)	-3.462** (1.475)	-0.291*** (.0856)	0.508 (1.757)
Editions per Title	-0.389*** (.132)	-0.0920*** (.0269)	0.0107*** (.0016)	-0.166*** (.0334)
Constant	31.998*** (6.389)	18.08*** (1.27)	0.288*** (.070)	34.371*** (2.09)
Paperback				-17.225*** (1.732)
E-Book				-31.747*** (1.996)

Standard errors in parentheses
 *** p<0.01, ** p<0.05, * p<0.1

This pattern in prices is also shown in figure 1.2, which illustrates the lack of discontinuity in prices of hardcover editions around the cliff date of 1923. Further analysis on the effect of the extension on availability and prices is shown in appendix A.2. While the copyright extension has a large effect on the availability and variety of books, the effect on prices is minor. My analysis will therefore focus mainly on availability while modelling price changes as a consequence of entry decisions.

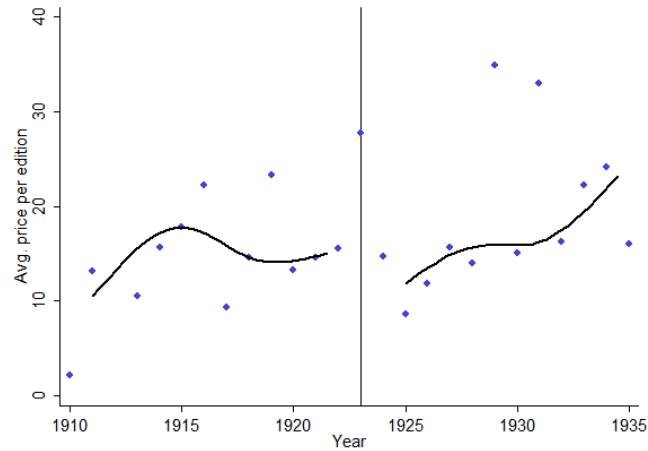


Figure 1.2: Average price per hardcover edition: copyright effect = insignificant

While the extension of a copyright term seems to have a negative effect on a title’s availability along all products in the product line, the effect on demand is more difficult to quantify. Both distributors of physical books (Amazon and PLR) sell more copyrighted titles than public domain titles, but it is important to determine how much of that difference is accounted for by added competition due to the increased ease of downloading a book through (e.g.) Project Gutenberg. My demand model identifies the titles’ creative qualities to account for parts of the difference in addition to explicitly modeling substitution patterns across titles and formats.

1.3 Model

The model has a consumer side and an entry component. A nested logit demand model is used to obtain creative qualities, price elasticities and substitution patterns. Copyright holders and edition managers¹¹ make entry and pricing decisions depending on these demand dimensions as well as the title’s copyright status. If a title is in the public domain the edition manager makes his entry decision in a free-entry system. If the title is protected by copyright, the entry decision is controlled by the copyright holder who maximizes the joint profits from all editions of a title. After the entry decisions have

¹¹Editions (ISBNs) are unique in their title, format, size, date of publication, and several other aspects (e.g. study notes or an introduction).

been made, edition managers compete in prices to maximize their own profits.

The estimated supply model allows me to recover marginal costs by format and copyright status, and to estimate fixed costs by format, copyright status *and* the title's creative quality.

I view the copyright holder as an entity that is separate from the edition managers. It is possible that an edition manager holds the copyright. In this case the smaller number of entrants and the higher price are a result of double marginalization rather than an increase in fixed costs. This would change the fixed cost estimates but the welfare analysis remains mostly unchanged. The counterfactual section comes back to this point.

1.3.1 Demand

Each creative work (or title) is indexed by $w \in W$, and each edition is indexed by j . Each edition j is published in a format denoted by $k \in \{H, P, E\}$, where H = hardcover, P = paperback and E = e-book. For each title that is in the public domain, there is one additional "format" through the Gutenberg website. A consumer's utility is a function of the creative work's quality as well as the edition's age and condition, and her income relative to the edition's price. The market consists of Americans who decide whether or not to buy a title each month. The "outside good" therefore includes any books that are not included in this data set as well as not reading at all. Given the size of the US population, the market size is 300 million.

The substitution patterns across the titles in the dataset and the outside good are likely to differ from those across editions *within* each title and format. In the demand estimation, I therefore allow for correlations of consumer tastes to differ on two levels: the correlation of tastes for editions within a title-format combination can differ from the correlation of tastes across formats within each title, and these correlations can differ from the correlation of tastes for editions across titles. As editions within each title-format combination differ in their publication date, condition and price, editions are identified in each title-format combination. Figure 1.3 illustrates the nesting structure.

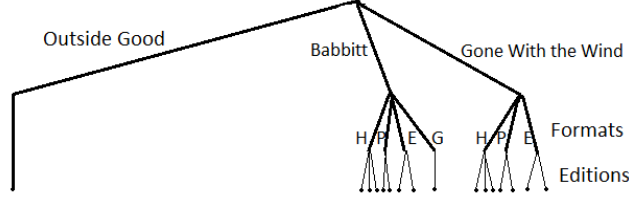


Figure 1.3: Nested Logit Structure

As in Berry (1994), if the unobserved components follow extreme value distributions, and if the mean utility level for the outside good is set to zero, then the difference in log-probabilities between ISBN (edition) j and the outside good can be expressed as

$$\ln(s_{jt}) - \ln(s_{0t}) = \delta_{jt} + \sigma_1 \ln s_{j|kw,t} + \sigma_2 \ln s_{k|w,t} + \epsilon_{ijt},$$

where

$$\delta_{jt} = \beta_0 + \beta_{age}age_{jt} + \beta_{new}new_{jt} - \alpha p_{jt} + \phi_w + \xi_{jt}.$$

In this model, p_{jt} is the edition's price as indicated by Amazon (it is set to zero if the edition is offered through Project Gutenberg), and the title fixed effects ϕ_w show the title's *creative* quality. This creative quality is exogenously given and cannot be influenced by an edition manager (as opposed to a *production* quality that is implicitly included in the ξ_{jt} term). The $s_{j|kw,t}$ is the conditional probability that a consumer chooses ISBN j given that she has selected to buy an edition of format k and title w , and $s_{k|w,t}$ is the probability of choosing format k given the choice of title w .

The parameters σ_1 and σ_2 should lie between zero and one. If the nested logit estimation obtains $\sigma_2 = 0$, the correlation of preferences across formats within the fiction titles in the data set is the same as the correlation of tastes for editions across titles, and consumers are as likely to switch between titles as they are to switch between formats within the titles in my data set. If $\sigma_2 = 1$, there is a high correlation of preferences across the formats within a title, and consumers are more likely to switch between formats within each title than to switch to another title or the outside good. Similarly, σ_1 determines the level of taste correlation between editions within a title-format combination as opposed to editions of different formats or titles. In each case a high $\sigma_i, i \in \{1, 2\}$, suggests that an additional edition would not increase the consumer

base much as most of the sales of a new edition will be attributed to sales cannibalization of other editions.

The nested logit demand estimates provide information on price elasticities of demand and on the level of market expansion versus sales cannibalization between editions of a title. In each month t market shares of edition j in the nested logit model are given by

$$s_{jt} = s_{j|kw,t} \cdot s_{k|w,t} \cdot s_{w,t}.$$

This expands into a tractable equation for an edition's market share as a function of its quality and the quality of its competitors¹²:

$$s_j = \frac{\exp\{\delta_j/(1 - \sigma_1)\}}{D_k^{(\sigma_1 - \sigma_2)/(1 - \sigma_2)} D_w^{\sigma_2} (1 + \sum_{w' \in W} D_{w'}^{1 - \sigma_2})} \quad (1.1)$$

where $D_k = \sum_{j \in J_{wk}} \exp\{\delta_j/(1 - \sigma_1)\}$ and $D_w = \sum_{k \in J_w} \left(D_k^{(1 - \sigma_1)/(1 - \sigma_2)} \right)$, and the month subscripts are dropped for notational convenience. An additional edition with mean quality δ_{wk} will affect each existing edition's market share. If the σ_2 term (the coefficient on the share of the broader nest) is smaller than the σ_1 term (the coefficient on the share of an edition within the narrow nest), market shares of an edition j within title w are a decreasing function of the number of editions for that title, while the market share of the *title* compared to other titles is an increasing function of the number of editions. The nesting structure in this model explicitly allows me to determine the effects of additional versions a title-format combination on the market shares of existing editions. The extent of the changes in shares depends on the σ terms.

1.3.2 Supply and Equilibrium

While I estimate several nested logit structures, I continue with the specification introduced in section 3.1 when I estimate fixed and variable costs as it allows me to express market shares as a function of the number $n = (n_H, n_P, n_E)$ of editions in each format and to obtain substitution patterns between formats.

A potential entrant's entry decision is directly affected by the work's copyright status $s, s \in \{O, I\}$ (where O indicates that the title is in the public domain (open to the

¹²See appendix A.3 for a detailed derivation.

public) and I shows that the title is under copyright (IP) protection) in two ways. First, potential entrants in the public domain decide to enter in a free-entry regime if their expected profits are positive, while they choose not to enter if their expected profits are negative. Under copyright protection an edition manager will have to *ask* the copyright holder to be allowed to enter. The copyright holder will allow edition managers to add editions (i.e. enter) until her expected profits are maximized. The copyright holder is assumed to receive a (fixed) proportion of the total profits from the editions of that title under the assumption that contracting is complete¹³. Movement of a title into the public domain means that the title moves from a profit-maximization regime into a free-entry regime.

Second, the cost structure of publishing an edition will depend on the work's IP status. An edition manager of a protected title will obviously face higher fixed costs if he has to find the copyright holder and obtain a license to publish an edition in addition to the search costs of obtaining a version of the text and the cost of setting up machines to print that are incurred in both regimes. A protected title is also likely to have a higher marginal sales cost - regardless of the edition's production quality - as the publisher usually has to pay a royalty to the copyright holder for each book sold. My supply model accounts for differences in both market structure and costs.

Formally, let c_j^s be the marginal cost of selling edition j given copyright status s , and suppose marginal costs consist of two components: the physical cost of producing a book (pr) and royalties paid to the copyright holder (cr). Then

$$c_j^s = c_j^{s,pr} + c_j^{s,cr}.$$

Royalties are paid only for books of protected titles¹⁴. Further, let F_{kw}^s be the fixed cost of publishing an edition of format k of work w (of quality ϕ_w) in copyright status s . Both marginal costs and fixed costs are allowed to vary by format *and* title quality within a copyright regime. Given these cost structures, the supply model has two stages. Firms choose whether or not to enter in stage 1 and, conditional on entry, set prices in

¹³This assumption will be relaxed in the counterfactual section.

¹⁴If editions have the same physical costs across copyright regimes, identification of royalties is obtained directly. It is, however, possible that production costs vary across copyright regimes by changing the incentive to create a "nice" version.

stage 2.

Consider first stage 2 in this entry game. Define $X_k = (X_{1,k}, X_{2,k}, \dots, X_{n_k,k})$ as the vector of observed edition characteristics for editions of format k (within title w), where $X_{j,k} = (age_j, new_j, \phi_w)$, and define $\xi_k = (\xi_{1,k}, \xi_{2,k}, \dots, \xi_{n_k,k})$ as the vector of unobserved edition characteristics from the demand estimation. Let $\mathbf{X} = (X_H, X_P, X_E)$ be the vector of the observed edition characteristics of title w , and define $\boldsymbol{\xi}$ accordingly.

Given the entry combination $n = (n_H, n_P, n_E)$ from stage 1 and the values of \mathbf{X} and $\boldsymbol{\xi}$, firms compete in Bertrand fashion. The edition manager of edition j takes as given the edition characteristics \mathbf{X} and prices \mathbf{p}_{-j} of all other editions of title w when choosing p_j to maximize his profits. The price setting process is thus assumed to be the same in each regime. That is, the copyright owner can decide to let an edition enter but is unable to coordinate the pricing of the editions, so that the edition managers compete with each other in prices¹⁵. Then the supply estimation focuses on the entry margin.

In stage 1 of the game, firms enter based on their expected profits. From the demand side model, I obtain market shares as a function of the number and characteristics of the available editions per title as

$$q_j(p_j, \mathbf{p}_{-j}, \mathbf{X}, \boldsymbol{\xi}) = M \cdot s_j(p_j, \mathbf{p}_{-j}, \mathbf{X}, \boldsymbol{\xi}),$$

where prices are determined in the second stage as functions of the number of available editions and of the editions' qualities. Variable profits from an edition are then given by

$$\pi_j^s(p_j, \mathbf{p}_{-j}, \mathbf{X}, \boldsymbol{\xi}) = \max_{p_j'} (p_j' - c_j^s) q_j(p_j', \mathbf{p}_{-j}, \mathbf{X}, \boldsymbol{\xi}).$$

A potential entrant of format k receives a draw of $\xi_j \sim F^k(\xi)$ and makes his entry decision based on his expected profits. Let

$$\begin{aligned} \Pi_k^r(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n) &= E [\pi_{jk}^s(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi})] \\ &= \int_{\xi} (p_j - c_k^s) q_j(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}) dF^k(\xi), \end{aligned} \tag{1.2}$$

¹⁵This assumption is consistent with Mankiw and Whinston (1986), who take stage 2 as non-cooperative in the social planning problem of entry as well. It finds support in the data as can be seen in section 1.2.1

be the expected profit for an entrant of format k , given that $n = (n_H, n_P, n_E)$ editions enter. Conditional on entry and prices, a work's copyright regime affects an edition's profits only through fixed and variable costs.

Public Domain, $s = O$

When a title is in the public domain, free entry conditions determine an edition manager's entry decision. I use the managers' entry decisions to obtain estimates for fixed costs of publishing for each title-format combination, given market shares, prices and costs. Similar to Ishii (2005), if I observe a vector of (n_H^*, n_P^*, n_E^*) editions, it must be that all of these editions generate positive expected profits, while an additional edition (of any format) does not. Total costs are bound by these Nash equilibrium observations for each w . Thus, for hardcover (H),

$$\begin{aligned}\Pi_{w,H}^O(n_H^*, n_P^*, n_E^*) &\geq F_H^O(\phi_w) \\ \Pi_{w,H}^O(n_H^* + 1, n_P^*, n_E^*) &\leq F_H^O(\phi_w),\end{aligned}\tag{1.3}$$

and similarly for P and E . At the same time, it cannot be profitable for edition managers to switch from one format to another, so that we get an additional set of restrictions:

$$\begin{aligned}\Pi_{w,H}^O(n_H^*, n_P^*, n_E^*) - F_H^O(\phi_w) &\geq \Pi_{w,P}^O(n_H^* - 1, n_P^* + 1, n_E^*) - F_P^O(\phi_w) \\ \Pi_{w,H}^O(n_H^*, n_P^*, n_E^*) - F_H^O(\phi_w) &\geq \Pi_{w,E}^O(n_H^* - 1, n_P^*, n_E^* + 1) - F_E^O(\phi_w).\end{aligned}\tag{1.4}$$

Similar restrictions apply to paperback and e-book editions. These inequalities and the definition of profits as given in equation (2) identify upper and lower bounds of fixed costs for publishing an edition of a title with creative quality ϕ_w in format k as all remaining terms of equation (2) are obtained in the demand estimation.

Copyright, $s = I$

When a title is protected by copyright, the copyright holder solves a profit maximization problem across the product line of the creative work. The copyright holder's variable profits are assumed to be a (fixed) proportion of each edition's profits under complete contracting. She therefore chooses the combination of formats that maximizes the title's

joint profits, given Bertrand competition after the entry decision has been made. Joint operating profits are given by

$$\Pi_w^I(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n) = n_H \Pi_H^I(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n) + n_P \Pi_P^I(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n) + n_E \Pi_E^I(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n), \quad (1.5)$$

where $\Pi_k^I(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n)$ is defined as in equation (2) above. If we observe the combination of editions $n = (n_H^{**}, n_P^{**}, n_E^{**})$, it must be that this combination is *most* profitable for the copyright holder. That is, marginal profits for less than this number of editions are positive, and marginal profits of introducing any additional edition would be negative. The joint profit maximization by title therefore gives the following set of inequalities for hardcovers:

$$\begin{aligned} \sum_{j \in n_w} \Pi_j(n_H^{**}, n_P^{**}, n_E^{**}) - \sum_{j \in n'_w} \Pi_j(n_H^{**} - 1, n_P^{**}, n_E^{**}) &\geq F_H^I(\phi_w) \\ \sum_{j \in n'_w} \Pi_j(n_H^{**} + 1, n_P^{**}, n_E^{**}) - \sum_{j \in n_w} \Pi_j(n_H^{**}, n_P^{**}, n_E^{**}) &\leq F_H^I(\phi_w) \end{aligned}$$

Again, the equivalent inequalities must hold for paperback and e-book editions. Fixed costs for protected titles are then obtained by title quality ϕ_w and format k after obtaining variable profits using the results from the demand estimation. The additional restriction that $\Pi_k(n_H, n_P, n_E) > F_k^I(\phi_w)$ for each $k \in \{H, P, E\}$ will be satisfied as long as the demand model is well-specified.

1.4 Estimation and Results

The estimation is done in three steps. I first estimate the demand parameters in a nested logit model with two nests, as described above. I then use the demand estimates to recover marginal costs by format and copyright status in the supply model. Lastly, I estimate bounds on fixed costs by title quality, format and copyright status using the equilibrium conditions of entry in the first stage.

1.4.1 Demand

Recall that the demand for an edition is estimated in a nested logit model that allows tastes to be correlated between editions within titles and title-format combinations. There is an issue of endogeneity as the share of an edition within a title is affected by the same unobserved variables as the left hand side variable. I account for this endogeneity by instrumenting the inside shares $s_{k|w,t}$ and $s_{j|kw,t}$ with the level of competition (the number of editions of each format) within the title-format combination. The number of editions per title is correlated with the level of popularity of the title, but it is not a good predictor of each edition's *market* share. It is, however, negatively correlated with the share of an edition within the *title*. To account for the possible endogeneity of prices in the demand estimation, I use each title's copyright status (a cost shifter that is unlikely to have an effect on demand for the title), and the number of British library checkouts (demand in a different market) as instruments. A first stage F-test shows that weak instruments are not a large concern in this estimation. While Shea's partial R^2 values are low (between 0.1 and 0.2), an F-statistic of 94.96 is far above the critical values in the Wald test.

Table 1.5 shows the results of the nested logit specification described above. The coefficients on the nests' inside shares are large and significantly different from zero. The regular logit model is misspecified, and the correlation of tastes for editions within titles and formats is different from the correlation of tastes across titles. I therefore use the quality estimates and price elasticities obtained from the nested logit estimation for the supply side model and policy implementations.

Table 1.5: Demand Estimation

	(1) OLS	(2) IV Logit	(3) Title	(4) Title - Format
New	0.665*** (0.0470)	3.084*** (0.136)	1.238*** (0.0869)	1.489*** (0.0989)
Price	-0.0173*** (0.00124)	-0.200*** (0.00749)	-0.0750*** (0.00470)	-0.0925*** (0.00559)
Major	0.0484 (0.0399)	-0.167* (0.0999)	0.0548 (0.0474)	0.118** (0.0522)
Age	-0.00292*** (0.000883)	-0.00882*** (0.00199)	0.00173* (0.000940)	0.00143 (0.00103)
$\sigma(\text{title})$			0.707*** (0.0228)	0.570*** (0.0295)
$\sigma(\text{format})$				0.785*** (0.0233)
Title fixed effects	✓	✓	✓	✓
Format fixed effects	✓	✓		
Mean Elasticity	-0.239	-2.772	-3.106	-4.287
Observations	4999	4999	4999	4999
R-squared	0.590	0.402	0.491	0.674

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notice that the price elasticities are rather low for the OLS and IV logit estimations. The nested logit estimation obtains a more reasonable price elasticity of -4.3, which is comparable to the pseudo-price elasticity of -9.68 obtained in section 2.1¹⁶. I focus my analysis on specification (4): tastes for editions are correlated within titles and title-format combinations. This specification allows me to obtain substitution patterns between formats of each title, and it provides reasonable coefficients and elasticities. Consumers prefer new editions over used ones. They prefer editions from major publishers over those from independent or smaller publishers. This supports the notion that

¹⁶It is lower because the pseudo-elasticity does not take preferences for certain formats and other product characteristics into account.

major publishers have an established reputation of quality or reliability. The coefficient on an edition's age is positive. It might reflect the appeal that a book reader sees in authentic original editions, but the coefficient is insignificant at the 10% level.

The σ terms show that there is a high correlation of preferences for editions within the suggested nests. The correlation of tastes for editions within a title-format combination is close to 1 at 0.785, but there is a high correlation of tastes across formats within titles as well, with a σ of 0.570. Thus, an additional edition of a title-format combination will not increase the share of this title much as any demand for this title will be mostly met by the existing editions. This is particularly true for editions of the same format.

Table 1.6 shows the ten titles with the highest and lowest estimated qualities. The estimated qualities mirror current demand for the titles and the overall quality of each work. Four of the top ten titles have won Pulitzer Prizes¹⁷ in their respective years of publication: *Gone With the Wind*, *The Good Earth*, *Years of Grace*, and *The Age of Innocence*. The authors of seven of the top ten titles have won the prize for one of their works. Five of the top ten works can still be found in a large local book store as well¹⁸. On the other hand, none of the bottom ten (in fact, none of the bottom 50) titles could be found at the book store. Nor did any earn a Pulitzer, although five of them were written by Pulitzer-winning authors¹⁹. As the ξ_j terms will be centered around zero for each title by construction, the creative quality terms ϕ_w indicate levels of production quality as well. There is a significant difference in mean values of ϕ_w between the two copyright regimes, with public domain editions having a higher quality than protected editions. This is probably due to the relatively large demand for free Gutenberg editions, which are available only for titles that are in the public domain. If the underlying assumption that all titles are a priori similar holds, this implies that publishers or the copyright holders do not extend a larger effort into producing "nice" editions of protected titles as compared to those of public domain titles²⁰. Thus, a move

¹⁷11 of the titles in my list are Pulitzer Prize winners - with three exceptions, they are all found in the top 30 titles

¹⁸I found 16 titles from my data set at a Barnes & Noble book store in Roseville, MN, in November of 2011: With two exceptions, all of these titles can be found among the top 50 titles in my estimation

¹⁹These titles may have been popular when they first appeared simply because the author was well-known and respected, although the titles themselves were not as good.

²⁰In any event, the argument by Adilov and Waldman (2012) that a longer copyright term triggers

into the public domain appears to affect consumer surplus through prices and variety only.

Table 1.6: Quality Rankings - Top 10

	Title	Pulitzer Prize
1	All Quiet on the Western Front (1929)	
2	The Good Earth (1931)	✓
3	The Age of Innocence (1921)	✓
4	Years of Grace (1930)	✓
5	Pollyanna (1913)	
6	The Lone Star Ranger (1915)	
7	Gone With the Wind (1936)	✓
8	The Forty Days of Musa Dagh (1935)	
9	Babbitt (1922)	
10	Main Street (1921)	

Bottom 10		
	Title	Pulitzer Prize
1	Lost Ecstasy (1927)	
2	Maid in Waiting (1931)	
3	Old Wine and New (1932)	
4	One Increasing Purpose (1925)	
5	Roper's Row (1929)	
6	Vein of Iron (1935)	
7	Old Pybus (1928)	
8	Within This Present (1934)	
9	Sparkenbroke (1936)	
10	The Road to Understanding (1917)	

The number of available editions is correlated with the estimated quality of a work, ϕ_w , as illustrated in figure 1.4. Works of a higher quality are available in more editions. This also depends on the work's copyright status as public domain works are much more widely available than protected works.

more ex-post investment in the work does not seem to apply to fiction novels.

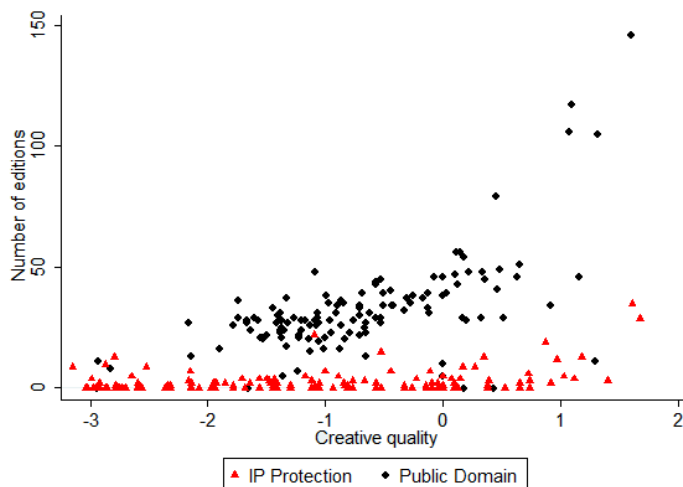


Figure 1.4: Editions per title as a function of quality

The σ terms in the demand estimation and the functional form of the market shares indicate that the degree of market expansion differs for each title and format since the number of editions per title varies as well as the quality δ_j among titles and formats. Table 1.7 shows estimated demand by edition and title as I add a hypothetical edition of each format. This affects the other editions directly by adding a competitor and indirectly by changing their optimal price. An additional hardcover edition decreases the share of the existing hardcover editions by an average of 28.5%, while existing editions of different formats lose only a small part of their shares (paperbacks lose 2.4% of their sales, while e-books lose 0.2%). This is consistent with the high value of σ_1 from the demand estimation: most substitution happens within a title-format combination. The effect of one additional paperback is generally low because there are more paperback editions to begin with.

Table 1.7: Average Annual Demand per Edition, All Titles (no Gutenberg)

	Base results	1 more H	1 more P	1 more E
Demand	34.154	33.464	32.522	30.557
% Change, Sales, H		-28.5%	-2.8%	-1.3%
% Change, Sales, P		-2.4%	-4.0%	-0.5%
% Change, Sales, E		-0.2%	-1.3%	-19.1%

The substitution patterns become clear when looking at specific titles. For example, there currently are three hardcover versions, six paperback versions, and one electronic version of *Gone With the Wind*. Consider the average monthly sales of this title. If one hardcover edition of this title were to enter the market, the mean monthly edition sales of hardcover editions would decrease from 103 to 86.7, a relatively sizeable decrease. At the same time, mean sales of the existing paperback and electronic editions would not change as much. There would be a decrease from 24.7 to 21.5 editions among paperback editions, and a decrease from 333 to 332.6 downloads of the e-book version. Overall, the additional hardback edition would steal a relatively large market share from the existing hardcover versions while affecting the market shares of the existing editions of different formats only a little. It would increase the total quantity of books sold for *Gone With the Wind* from 790 to 808. The market shares of existing editions of different titles are essentially unchanged. The large values of σ_1 and σ_2 keep the effect of additional editions within the title²¹.

Much of the absolute change in within-title market shares depends on how many editions of a title are originally available. If there are a large number of editions for a particular title, one more edition will not affect the market shares of the existing editions as much as it does when there are only a few editions. Figure 1.5 shows the relationship between changes in market shares and the number of available editions for the hardcover format. The lines depict the fraction of original sales of existing editions of each format after one hardcover edition is added. If one hardcover edition is added to the existing editions of a title, the market shares of the incumbent hardcover editions

²¹I use this fact during the policy experiments as I assume that a change in the number of editions for one title will not affect the demand for other titles.

will drop much more than the shares of editions of other formats. The drop in the shares of existing editions is much more pronounced if there currently are only a few editions.

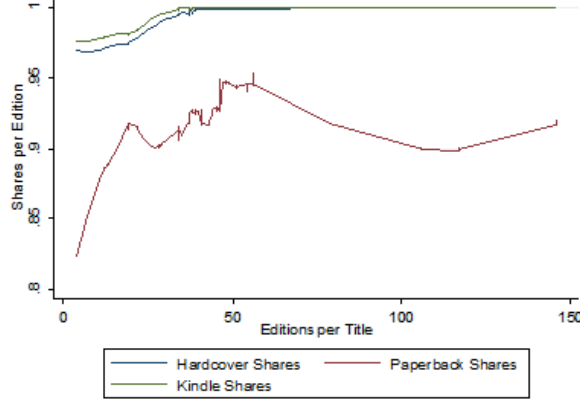


Figure 1.5: Shares as a % of Original Shares after adding a Paperback Edition

1.4.2 Supply

Through the demand model I obtain market shares as a function of the number of available editions per title and of the title’s creative quality ϕ_w . The demand side model also gives me an estimate of each edition’s price elasticity ϵ_j of demand, which is used to recover marginal costs on the individual edition level. The first-order necessary condition of profit maximization in Bertrand competition is derived as

$$\frac{p_j - c_j}{p_j} = -\frac{1}{\epsilon_j},$$

where $\epsilon_j = \frac{\alpha p_j}{1 - \sigma_1} \left[1 - \frac{\sigma_1 - \sigma_2}{1 - \sigma_2} s_{j|wk} - \sigma_2 \left(\frac{1 - \sigma_1}{1 - \sigma_2} \right) s_{j|w} - (1 - \sigma_1) s_j \right]$ as derived in appendix A.3. The marginal costs are edition specific and vary by a title’s copyright status $s \in \{O, I\}$. Note that this assumes that edition managers are single-product firms. This assumption will lead to a slight overestimation on marginal costs. However, I observe only a few publishers (less than 4% of the observations) that offer more than one edition per title, so that this assumption will not bias the results much. An average of the markups by format and copyright status, $E_k^s \left[-\frac{1}{\epsilon_j} \right]$, is used together with the entry equilibrium inequalities to identify bounds on fixed costs by title, format and

copyright status. I assume that an entry decision is made on an annual basis, so that fixed costs $F_k^s(\phi_w)$ will be incurred each year as well. Fixed costs include the costs of obtaining the license to publish an edition²² and of setting up the machinery and printing a certain level of inventory, while marginal costs include the material for each (physical) book, maintaining the distribution, and potentially royalties for protected versions.

The empirical distribution of the product-specific unobservables ξ_j follows a normal density closely. To obtain estimates of the bounds of fixed costs of publishing a title-format combination, I draw the ξ terms from format-specific normal distributions to obtain the quality δ_j of a hypothetical additional edition. The format-specific distributions are centered around zero and have standard deviations of 0.7 (hardcover), 0.9 (paperback) and 1.0 (e-book).

The lower bound of fixed costs of publishing an edition of a public domain title is given by the hypothetical revenue if one additional edition of format k (with unobserved quality drawn from ξ_k) was introduced, and the upper bound is the estimated revenue for format k given the current combination of editions of title w . That is, in the public domain, the upper bound of fixed costs is given by the expected variable profit of an edition of format k when $n = (n_H^*, n_P^*, n_E^*)$ is observed. Formally,

$$F_k^O(\phi_w) \leq \bar{F}_k^O(\phi_w) \equiv \Pi_k^O(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n),$$

where $\Pi_k^O(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n)$ is given in equation (2) and n is the observed combination of editions. Similar steps determine lower bounds of fixed costs in the public domain and fixed cost bounds of publishing protected titles²³.

One feature of my dataset deserves comment here. While I do observe quantities demanded of several editions per title and format, I only observe these quantities through one distribution channel: online retail. Any sales through more traditional channels such as brick-and-mortar bookstores and libraries are not observed although several books are obtained through these channels as well. As a consequence I observe only a fraction of variable profits for each edition. According to the Book Industry Study

²²this includes identifying the copyright holder and convincing her to let you publish

²³My current estimation assumes that a hypothetical additional edition receives $\xi_j = E[\xi_k]$. In this case, profits of each firm are deterministic, and the inequality becomes $F_{wk}^O \leq \bar{F}_{wk}^O = (p_j - c_k^r)q_j(\mathbf{p}, \mathbf{X}, \bar{\xi}_k)$

Group BookStats Report 2011 online retail accounted for around 10.7% of total books sold in the publishing industry in 2010. This number is likely to have increased since then, and it is likely to be larger for old fiction titles that are difficult to find in physical bookstores as well. A Bowker study determined that Amazon accounted for 20% of the publishing market in quarter two of 2011, and for 27% in the second quarter of 2012²⁴. For my further analysis, I assume that Amazon captures a share of $\gamma = 0.2$ of the sales of my list of editions. In the estimation of fixed cost bounds, I then use

$$F_k^O(\phi_w) \leq \frac{1}{\gamma} \cdot (p_j - c_k^O) \cdot q_j(\mathbf{p}, \mathbf{X})$$

and the equivalent inequalities corresponding to each bound and copyright regime to estimate bounds for fixed costs. As I may observe a larger fraction γ of the market than the 20% indicated by the Bowker study (due to the obscureness of some of the titles I observe), the reported results provide an upper bound on the fixed cost bounds.

Table 1.8: Marginal and Fixed Costs

Mean Marginal costs per book sold					
	Public Domain		IP Protection		Data (BookStats)
	in \$	Markup	in \$	Markup	Markup
Hardcover	18.26	0.284	18.58	0.308	0.376
Paperback	9.90	0.347	13.16	0.364	0.532
E-Book	3.29	0.588	2.53	0.737	0.227

Average Fixed cost bounds per edition published (in \$)					
	Public Domain		IP Protection		
	Lower	Upper	Lower	Upper	
Hardcover	91.80	119.31	712.28	945.08	
Paperback	58.41	88.14	836.03	1379.95	
E-Book	69.03	70.34	919.68	1653.60	

Table A.1 shows the means of estimated bounds of these costs, conditional on copyright status and format and assuming that $\gamma = 0.2$. For both hardcovers and paperback, the average marginal and fixed costs are higher among protected titles than among

²⁴see Publishers Weekly; 11/5/2012, Vol. 259 Issue 45, p6-6. Much of the increase in Amazon's market share can be attributed to Borders going out of business.

public domain titles. The sample size for e-books may be too small to give reliable estimates²⁵. Fixed costs are, on average, around ten times as high under copyright as they are in the public domain. Edition managers of all titles incur some fixed cost of obtaining a script of the title and setting up the machinery (or the platform) to produce books of that title. Apart from these setup costs, publishers of public domain titles do not incur any significant costs. Publishers of protected titles, on the other hand, incur a search and transaction cost of setting up a licensing agreement with the copyright holder. This cost is at the order of a few hundred dollars per year for most titles, but it is over \$5,000 for some of the highest-quality titles as figure A.4 shows.

Recall that this is under the assumption that the copyright holder wants to maximize the title's joint profits. If a publisher owns the copyright they maximize their own profit and are likely to restrict entry "artificially". This case explains the high estimated fixed costs for high-quality titles. The estimated fixed cost bounds are an upper bound of actual fixed costs.

While marginal costs are just slightly higher for protected titles, the profit margin is larger for protected titles as average prices are higher as well (although this difference is not significant²⁶). This is consistent with both the differences in market structures and in royalty payments. Protected titles may have a double markup: both the copyright holder and the edition manager want to make profits with each edition. At the same time, edition managers pay royalties to the copyright holder, thus incurring a higher marginal cost. Notice also from table A.1 that the cost margins fit the benchmark values from the BookStats report reasonably well. If the physical cost of producing an edition does not depend on the work's copyright status, identification of royalties is given through the marginal cost function:

$$c_j^s = c_j^{s,pr} + c_j^{s,cr}$$

with $c_j^{s,pr}$ being the production cost of an edition, and $c_j^{s,cr}$ the royalties paid to the

²⁵In the reported results I drop those observations that give me "unreasonable" fixed costs. That is, editions whose own-price elasticity is between zero and -1. I thus drop observations with unreasonably low prices and some other outliers. The set of dropped observations includes several e-books. I impose an average marginal cost by format and copyright status in an alternative specification. The estimates are similar, as shown in appendix section A.4.

²⁶see section 1.2.2

copyright holder. The difference between the marginal cost estimates (in \$) is then an estimate of the royalties paid to the copyright holder. For physical editions, my estimates suggest that royalties are between \$0.50 and \$3.00. Greco (2004) lists a standard royalty rate of 10% to 15% of the price²⁷. At a sample average price of \$20 to \$30, the estimated marginal costs are reasonable.

I allow fixed costs to vary by title quality. My estimates show that fixed costs of public domain titles remain stable across titles, whereas protected titles carry a higher fixed cost if the title is of a high quality. This is consistent with the notion that the copyright holder of a high-quality title holds more bargaining power if the edition manager is confident the edition will sell. Mechanically, it is driven by the fact that high-quality titles generate large revenues for the current editions, and they would still generate high revenues for a hypothetical additional edition²⁸. Figure A.4 shows this relationship between creative qualities and fixed costs.

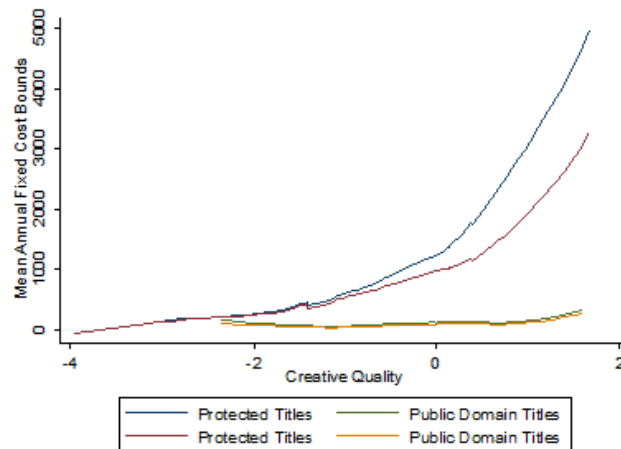


Figure 1.6: Fixed Cost Bounds by Creative Quality - Hardcover

Profits differ by copyright status as well. On average, edition managers of protected titles have larger profits than publishers of public domain titles. This follows from the differences in the market structures: potential publishers of public domain titles enter until profits are (close to) zero, while publishers of protected titles enter as long as the

²⁷page 156

²⁸Even if the fixed cost estimates in the copyright case seem high, the result of the paper will not change, as can be seen in section 5.2

copyright holder makes positive marginal profits over all editions of the title. As I allow for heterogeneity in editions within each title-format combination, profits vary across editions and several editions make positive profits even in the public domain. Table 1.9 summarizes profits for editions under their current copyright status, using the midpoints of fixed cost bounds as point estimates for costs.

Table 1.9: Profits per Edition (in \$ per year)

	Public Domain		IP Protection	
	Mean	Max	Mean	Max
Hardcover	14.72	96.60	1,222.59	28,532.95
Paperback	9.78	152.83	1,780.19	12,186.94
E-Book	15.41	69.01	2,759.75	15,098.00

The titles with the highest creative qualities ϕ_w produce the largest profits among protected titles while it is only possible for public domain titles to make positive profits through their draws of ξ_j and through integer constraints. Editions of *Gone With the Wind* (Margaret Mitchell), *The Good Earth* (Pearl Buck) and *All Quiet on the Western Front* (Erich Maria Remarque) - all protected by copyright - have the largest profits. Profits of protected low-quality title editions are lower than those of public domain titles of a similar quality due to differences in fixed costs. Protected high-quality titles, however, can generate very high profits for their editions, compared to editions of high-quality public domain titles which are operating in a free-entry system and compete with a zero-price option. Annual profits for editions of protected titles are as high as \$28,532 (*Gone With the Wind*), compared to \$97 for an average edition of the most profitable public domain titles (*The Age of Innocence*).

A change in a title's copyright status affects profits as well as the level of consumer surplus through a change in the number of competitors per title and price per edition. A combination of these effects is analyzed in the following section.

1.5 Policy Analysis

The above results provide a framework for testing the effects of a copyright on consumer and producer surplus. In each copyright regime, a title-format combination is assigned a fixed and variable cost based on the demand and supply results above. An equilibrium in the number of editions is then uniquely determined by the inequalities in the entry model. For each title, quantities sold are functions of the number of editions in each format, and bounds for marginal and fixed costs are obtained in the supply estimation. The unknown variables are the number of hardcover, paperback and e-book editions. This gives three equations in three unknowns for each title (one for each format). Since all titles' quantities demanded are interrelated, this becomes a system of $W \cdot 3$ equations in $W \cdot 3$ unknowns, where W is the set of titles in the data set.

The aim of the experiment is to examine and quantify the effect of a change in a title's copyright status on consumer surplus and on profits of the existing editions. Demand for an edition and supply of a title are determined by the title's copyright status (among other factors). A title's move from copyright protection into the public domain decreases the fixed and marginal costs of production of an edition. It also changes the identity of the entry decision maker: the copyright holder provides licenses to edition managers if the title is protected, while the edition managers themselves make the entry decisions in the public domain. A different combination of entrants follows, and differences in overall consumer surplus and producer surplus are realized. I record the subsequent changes in consumer surplus, in variable profits of existing editions, in variable profits of new entrants, and in fixed costs per title.

By definition, person i 's consumer surplus is the utility, in dollar terms, that she receives in the choice situation. Consumer i chooses the edition j that provides the greatest utility, so that consumer surplus is

$$E[CS_i] = \frac{1}{\alpha} E[\max_j (\delta_{ij} + \epsilon_{ij})].$$

Given the logit assumption that the error terms ϵ_{ij} are distributed iid extreme value, Rosen and Small (1981) show that this expectation has a tractable closed form expression. In my application, the difference in consumer surplus from title w extends

to

$$\Delta E[CS] = \frac{1}{\alpha} \left[\ln \left(1 + \sum_{w' \in \mathcal{W}^1} (D_w^1)^{1-\sigma_2} \right) - \ln \left(1 + \sum_{w' \in \mathcal{W}^0} (D_w^0)^{1-\sigma_2} \right) \right], \quad (1.6)$$

where the superscripts describe the choice sets under examination: 0 stands for the current setup, and 1 describes the counterfactual setup. While the number of *titles* in my choice set does not change, the number of *editions* of each title is allowed to vary by copyright status. The resulting change in consumer surplus describes the difference in value to the representative consumer (in dollar terms). The total change in consumer surplus is $M \cdot \Delta E[CS]$.

Variable profits for existing editions are calculated as described in section 3.2 above, and changes in producer surplus per title as a result of a change in a title's copyright status are:

$$\begin{aligned} \Delta PS_w &= \sum_{j \in J_w^1} (M \cdot (p_{wk}^1 - c_k^1) \cdot s_j(\mathbf{p}^1, \mathbf{X}; n_w^1) - F_k^1(\phi_w)) \\ &\quad - \sum_{j \in J_w^0} (M \cdot (p_{wk}^0 - c_k^0) \cdot s_j(\mathbf{p}^0, \mathbf{X}; n_w^0) - F_k^0(\phi_w)) \end{aligned}$$

Profits for new entrants are calculated as described in the model section above, and fixed costs are obtained through the supply side estimation. All values are annualized.

1.5.1 Moving Titles into the Public Domain

A move of a title into the public domain changes three parts of the publishing process. First, the entry decision is now made by the edition managers in a free-entry regime. Second, the costs of publishing decrease as any potential entrant does not have to incur any search costs or licensing fees. In the counterfactual experiment, the fixed costs of publishing an edition after a move into the public domain will be estimated at the public domain level for works of the same creative quality ϕ_w . Third, public domain titles are offered as a free digital version through Project Gutenberg. These three changes affect the level of availability and variety of each title. A move into the public domain will make titles available in a wider variety. In this counterfactual experiment, the number of

editions per title after a move into the public domain matches the data - the number of versions per public domain title - reasonably well. The combination of these editions is biased toward hardcover editions. This may be due to the fact that paperback versions seem more popular than hardcover editions. Thus, the fixed costs of entering paperback versions are at a larger level, while the demand model does not pick this preference up quite as well. As a consequence, the counterfactual experiment predicts between 5 and 10 fewer paperback editions and around 5 more hardcover editions than what we would expect from looking at the public domain works in the data. The model also predicts more e-books than what we observe. The predicted overall number of editions is overestimated by an average of 3 editions (10%). This will likely overstate the consumer gains described below by a small amount. This overstatement is most likely not large enough to affect the overall result.

The welfare effect of a title's move into the public domain differs by its creative quality. Table 1.10 shows that the effect of a change in a title's copyright status from copyright protection to the public domain is positive but small for all titles. Almost all titles in my dataset produce a higher total surplus after a move into the public domain. The magnitude of this effect is smaller among low-quality titles than it is for high-quality titles. When a low-quality title is in print, only a small number of consumers will read an edition of the title. Profits will be small whether a copyright holder wants to maximize profits or not. If the title is made available in both regimes, an increase in the number of editions has only a small effect on consumer surplus as there is a large degree of correlation of preferences within titles. The addition of a Project Gutenberg edition adds to this effect as it will be offered free of charge, which consumers like.

The effect of the copyright extension is similar for high-quality titles although the magnitude of the surplus changes is larger for these works. The increase in consumer surplus outweighs the decrease in profits for all high-quality titles. A move into the public domain increases total surplus by an average of \$3554.69 each year. This indicates that a copyright on a title causes insufficient entry. The magnitude of this effect varies by the title's creative quality. It also varies by the number of editions that are currently available and by other title- and edition-specific determinants of demand and supply. Appendix A.4 provides further details on the effect of a move into the public domain on a title-by-title basis.

Table 1.10: Mean Welfare Changes by Creative Quality (in \$ per title over one year)

	Bottom 25%	Mean	Top 25%
Δ CS	642.09	4,655.78	15,843.14
Δ PS (existing firms)	-47.46	-1,186.16	-5,044.80
Variable profit (new entrants)	391.23	2,576.76	9,334.52
Fixed cost (new entrants)	-333.38	-2,491.69	-9,136.90
Δ Total Surplus	652.48	3,554.69	10,995.95

There is an additional dimension to the titles that are affected by the copyright extension. Many titles that were published several decades ago have long moved out of print. Several of these titles will move back into print if they move into the public domain. These titles will change the total surplus in a different manner than titles that have never been out of print. An increase in consumer surplus stems from the title being made available while it wasn't available before (at least not in the new-books market). This change in a work's in-print status can potentially increase consumer surplus more than an additional edition of a work that is available anyway. Profits do not change much as profits are zero for titles that are not produced, and they are close to zero under perfect competition as edition managers will enter until profits are dissipated (up to an integer constraint). Table 1.11 shows the relationship between the titles' in-print status and the effect of a copyright on welfare by comparing titles that are currently out of print with titles that have the same estimated creative quality but are still in print.

Table 1.11: Mean Welfare Changes by In-Print Status (in \$ per title over one year)

	Out of print	In print - similar ϕ_w
Δ CS	2,634.09	1,535.29
Δ PS (existing firms)	0	-253.79
Variable profit (new entrants)	553.09	958.45
Fixed costs (new entrants)	-493.50	-912.59
Δ Total Surplus	2,693.68	1,327.36

Both in-print and out-of-print titles would, on average, increase total surplus if they were moved into the public domain. The magnitude of these changes differs by the title's in-print status and its creative quality. Titles that are currently out of print and move back into print as a result of a move into the public domain cause a larger

increase in consumer surplus that titles that are in print regardless of their copyright status (\$2634 per year versus \$1535 per year). The effect on total surplus (including producer surplus) is then twice as large for titles that are originally out of print than for comparable titles that are in print under each regime. The welfare effect of a decrease in the copyright term thus depends on the number of titles that are currently in print and on the number of titles that would be made available if they were in the public domain.

For instance, only 174 of over 10,000 books originally published in 1930 were still in print in 2001²⁹, whereas Project Gutenberg predicts that "virtually all" pre-1923 public domain titles could be available by the end of the decade³⁰. A calculation that takes into account the distribution of creative qualities by year can quantify the welfare effect of a copyright extension. This requires obtaining a distribution of qualities for titles that are in-print as well as titles that are out of print and would become available after a move into the public domain. Using the set of works in this paper, and obtaining download counts of a large number of works that are available at Project Gutenberg, I estimate a distribution of creative qualities of works created in a given year. I find that a copyright extension by one year would result in a welfare loss of \$10 million to \$20 million per year.

1.5.2 Decomposition of the Welfare Effect: A Switch to Free Entry

A title's move from copyright into the public domain changes both its cost structure and the entry decision. First, costs will be lowered, causing an increase in profits (keeping the number of entrants constant). Second, regardless of an edition's fixed and variable costs, a move into the public domain increases the number of in-print editions as the result of a switch from a profit-maximizing combination to a free-entry combination (assuming that the costs are not prohibitive).

In this section I assume that a title's fixed and variable costs are at the public domain level regardless of the work's copyright status. The effects on total surplus are equivalent to a world in which there are no inefficiencies due to search and transaction

²⁹see the *American Library Annal and Book Trade Almanac for 1872-1957* and Bowker's *Books in Print Online*.

³⁰see the Brief of Amici Curia, The Internet Archive, Prelinger Archives, and Project Gutenberg

costs. Instead, the costs to publishers from copyright protection are entirely transfers to the copyright holder. A focus on a switch to a free-entry system while fixing costs of entry at the public domain level shows that for most titles that are currently in print a switch to the free-entry system results in higher welfare as the decrease in profits does not offset the increase in consumer surplus. While the Mankiw & Whinston result of excess entry does not hold for most titles, there are differences between titles of different levels of creative qualities.

Table 1.12: Mean Mankiw Effect by Creative Quality (in \$ per title over one year)

	Bottom 25%	Mean	Top 25%
Δ CS	642.09	4,655.78	15,843.14
Δ PS (existing firms)	-318.18	-3,727.39	-11,674.24
Variable profit (new entrants)	358.78	3,124.93	8,944.30
Fixed cost (new entrants)	-300.93	-3,030.15	-8749.55
Δ Total Surplus	381.76	1,023.17	4,363.65

Table 1.12 shows the free-entry effect for average titles, as well as for the lowest and highest creative quality quartiles. A switch to a free-entry regime increases the number of available editions significantly, causing a loss to producers and an increase in consumer surplus. The lowest-quality titles are often only available in a few editions. For these titles a move into a free-entry system does not dissipate a lot of profits as the quantity demanded is low regardless of the level of variety. This changes significantly as the creative quality of a title increases. The highest quality titles make large profits if a copyright holder restricts entry. These will be lost due to the presence of fixed costs under free entry. Free entry thus decreases the profits to producers by a nontrivial amount. However, the gains to consumers are larger than the losses to producers. The Mankiw result of excess entry holds only for a few of the titles in my dataset. These titles generally have a low creative quality but a relatively large variety. Note that a switch into a free-entry regime does not affect titles that have run out of print. These titles will not move back into print based solely on a switch to a free-entry regime - the fixed costs of production prohibit entry into the market.

The effect of a move into a free-entry regime on individual components of total surplus depends on the work's creative quality. Overall, the effect of a switch to a

free entry regime is stronger among titles that carry a high creative quality. This is illustrated in figure 1.7. Each component of total surplus is affected more strongly if the work has a higher creative quality.

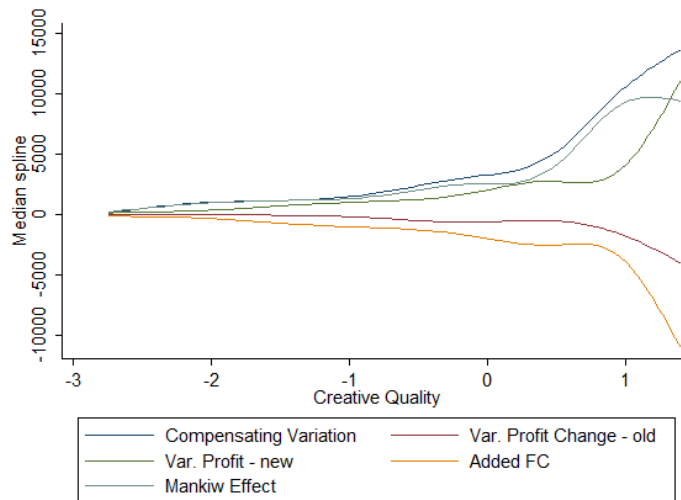


Figure 1.7: Profits by Creative Quality

The collective of new entrants produces a large additional operating profit from entering editions into high-quality titles, but their operating profit is dissipated by fixed costs. Existing editions lose profits for those titles as well. For most titles, however, differences in producer surplus are small compared to the differences in consumer surplus. The overall effect of free entry is positive for almost all titles despite the presence of fixed costs. The Mankiw & Whinston result of excess entry does not hold in this market.

Note that the differences in total surplus also reflect a world in which the differences in fixed and variable costs across copyright systems consist only of transfers from the edition managers to the copyright holder. The producer surplus changes here are the changes in combined profits to publishers *and* the copyright holder. While the previous section and the supply estimation assume that there is no licensing fee and all differences in entry are purely due to search and transaction fees, this section assumes that there are no search costs. Rather, the copyright holder (who might be an edition manager in this case) limits entry artificially.

In reality, the fixed costs of publishing a title will be a combination of search costs and a transfer to the copyright holder in the form of licensing fees and royalties. Sections 5.1 and 5.2 depict an upper bound and a lower bound for costs. The copyright extension decreases total surplus under each of the two extreme cases. A linear combination of the costs under copyright protection will still lead to a decrease in total surplus.

1.5.3 Decomposition of the Welfare Effect: Project Gutenberg

In addition to moving into a free-entry regime and lowering the costs of entry, a title's move into the public domain is accompanied by the introduction of a free digital option through Project Gutenberg, Google Books and several other institutions. The above analysis includes this option, thus simulating the world that we would be in today, with a shorter or longer copyright term. However, Google Books was not introduced until 2004. And although Project Gutenberg was established in 1971, only four of the 125 public domain titles had been posted before the 1998 Copyright Term Extension Act. I therefore simulate a world that the policy makers may have envisioned - without the option of downloading a free digital version of public domain works. A switch into the public domain then has only two effects: lower costs and a switch to a free-entry system.

Table 1.13: Mean Welfare Changes Without Zero-Price Options (in \$ per title over one year)

	Bottom 25%	Mean	Top 25%
ΔCS	436.47	2,655.44	9,352.58
ΔPS (existing firms)	-54.78	-847.17	-4,300.48
Variable profit (new entrants)	1464.92	11,660.52	49,840.63
Fixed cost (new entrants)	-1380.68	-11,279.06	-48,363.14
Δ Total Surplus	465.93	2,189.73	6,529.58
Effect of Gutenberg on ΔTS^*	28.5%	38.4%	40.6%

* This is the fraction of total surplus changes that are not covered by Gutenberg:

$$\frac{\Delta(TS)_{table10} - \Delta(TS)_{table13}}{\Delta(TS)_{table10}}$$

These two effects will result in a new combination of entrants and subsequently different prices and quantities. Table 1.13 shows the effect of a move into the public domain when the public domain does not include a free digital option. Comparing these changes in surplus to the changes we see when a move into the public domain

includes the introduction of a free digital version, several patterns become apparent. If there is no zero-price competitor, even more edition managers will enter. There will be more variety per title, which is good for consumers. But on the other hand consumers cannot obtain the title for free. On average, the effect from the zero-price option is larger than the effect from the additional increase in variety, so that consumers benefit even more from the existence of Project Gutenberg than if there is no zero-price option. On the producer side, profits will be dissipated whether there is a free option or not. The magnitude of the effect of the copyright extension thus is primarily affected by the effect on consumers. The average contribution of free digital options on the total surplus changes is around 38%. Project Gutenberg plays a larger role among high-quality titles. Consumers would greatly appreciate being able to read a work like *Gone With the Wind* for free, while they do not care as much about being able to read an obscure title for free. The effect of a move into the public domain is overwhelmingly positive regardless of the presence of Project Gutenberg. But now that a consumer can easily download public domain works for free, the effect is even larger. In light of the Google Books litigation, this result indicates that making books available for free (especially if they have been out of print) can increase total surplus significantly.

1.6 Conclusion

This paper provides a framework for calculating the effects of a change in an intellectual property regime and a subsequent change in the availability of varieties of books on consumer surplus and on profits. Entry and variety decisions in the publishing industry depend on a work's creative quality and its copyright status. I develop a model of the publishing industry consisting of a discrete choice demand model that allows me to determine a work's quality and substitution patterns between formats of each title as well as between editions of different titles. I find that an additional edition increases a title's sales only slightly. I identify variable costs and bounds for fixed costs of producing an edition of each title given its copyright status, in an entry model that draws upon the moment inequalities literature. I find that fixed costs of publishing an edition are much higher if the title is protected by copyright, while marginal costs of each book sold are similar across IP regimes.

The publishers' collective entry decisions are based on the work's estimated quality and copyright status. Titles that are in the public domain are available in a wider variety. This trend is especially prevalent among timeless classics - titles with a high creative quality. Editions of titles from the public domain will enter until their expected profits are close to zero, while editions of protected titles will be entered until the copyright holder makes zero marginal profits. Thus, editions of high-quality protected titles produce a large profit, while editions of high-quality public domain titles make small profits.

I find that a copyright increases producer surplus by limiting entry and competition, while an additional entrant does not increase the work's market share much due to a large degree of substitutability between editions of a title. Consumer surplus increases more than producer surplus decreases when a title moves into the public domain, indicating that there is insufficient entry under copyright protection. A significant part of this is due to recent developments in technology, making public domain titles available for free in digital formats.

While I show that the copyright extensions in the 20th century were welfare decreasing, the increased popularity of Project Gutenberg (and the introduction of Google Books in 2004) increased the negative effects of the copyright extension. This is true especially for the large stock of titles that have become orphans in the sense that the copyright holder cannot be found. Whatever we thought were the disadvantages of a copyright extension on existing works when the policy was made, these effects are amplified in today's digital world.

Chapter 2

Storming the Gatekeepers: Digital Disintermediation in the Market for Books

2.1 Introduction

Technological change has transformed content industries such as recorded music, newspapers, movies, television, and books. The recorded music industry was the first to face challenges from digitization with the arrival of the Napster file-sharing service in 1999. Endowed with the opportunity to obtain music files without payment, consumers withdrew from purchasing recorded music. A large body of research documents harmful effects of file-sharing on recorded music revenue, plausibly explaining all of the reduction in recorded music revenue¹. Newspapers too have faced substantial challenges from digitization, and their revenues have fallen as sharply as the revenues to recorded music, by half since the late 1990s².

While the past decade has been challenging for participants in content industries, observers and participants have begun to appreciate that new technology brings benefits,

¹See, for example, Oberholzer-Gee and Strumpf (2007), Rob and Waldfogel (2006), ? for empirical evidence on piracy in music. Liebowitz (2006) provides evidence that file sharing explains the bulk of revenue reduction in music.

²See Greenfield (2013).

in the forms of lower costs of production and distribution, along with threats arising from a handicapped ability to harvest revenue. An emerging body of work documents counterbalancing beneficial effects of new technology on the flow of new musical works. Waldfogel (forthcoming) documents that the service flow from new music has increased since the late 1990s, along with some evidence explaining how this could be so.

The beneficial effect of new technology on the availability of new products has been obscured in both the view of researchers and in public policy debates by the negative impacts of new technology on revenue. This is understandable given the timing of the various effects of new technology on music. It took four years from the dawn of widespread online stealing with Napster until the availability of a viable and attractive outlet for purchasing digital music in the form of the iTunes Music Store. Thus, for four years, digitization was visibly harming music producers without apparent offsetting effects.

Digitization has unfolded differently in the market for books. Until the launch of the Amazon Kindle in 2007, there was no widely adopted platform for legal or illegal consumption of digital books. While one could view, say, a pdf file on a computer, the legal and illegal markets for digital books remained small prior to Kindle. Since Kindle's launch, e-readers have diffused rather rapidly. By the beginning of 2012, the share of households with a digital reader had grown to 30 percent. The US market for digital books has grown correspondingly, to 5 percent of the market for trade books in 2010, to about 15 percent in 2011, and to a quarter of the market during the first half of 2012³.

The evolution of the e-book market is interesting in itself as a case study of a new product; it is perhaps even more interesting as a context where digitization's possible impacts operating through cost reductions are not obscured by widespread digitally-enabled theft. Rather than piracy, digitization has had two different major impacts on the market for books. First, digital distribution has reduced the marginal cost of books to essentially zero, which has substantially reduced prices.

The low costs of e-book distribution also allow for a second effect. Making a new book meaningfully available to consumers has traditionally required the assistance of one of the six major world publishing houses, which have acted as gatekeepers of literary commerce. Now, by contrast, online platforms such as Amazon's Kindle Direct

³See Milliot (2012) for data on 2010 and 2011, and the 2013 BookStats report for data on 2012.

Publishing arm, Smashwords, Lulu, and others make it possible for authors to circumvent the traditional publishing gatekeepers to make their products directly available to consumers. Self-publishing has grown substantially since 2011: the number of new self-published works now exceeds the number released traditionally. Because these products would not previously have made their way to consumers, self-published works have augmented the available choice set. This is potentially a consequential phenomenon: some of the best-selling titles of 2012 (the *Fifty Shades* series) all began their commercial lives as self-published works.

New technologies may therefore have brought about welfare benefit through both a lower-priced format as well as a growth in product offerings. This paper seeks to document the welfare effects of the growth in electronic books, and we proceed in seven sections.

Section 1 presents background on the changes in the book market brought about by e-books (including the diffusion of e-readers, the appearance of effective channels for self-publishing, and the growth of product discovery institutions). Section 2 presents a simple theoretical characterization of the ways in which digitization may have changed the welfare of market participants. Section 3 describes the three basic data sets, from nine underlying sources, that we assemble for the study. Section 4 presents descriptive results on the evolution of book sales quantities, prices and the number of products brought to market over time, along with results on the evolution of sales concentration and the market penetration of self-published works.

Section 5 turns to estimates of the welfare impacts of digitization. First, we present a crude estimate based on falling prices and a rising quantity of books purchased. Second, we use data on physical and electronic sales of bestsellers in 2012 to estimate a nested logit demand model, which we can use to separately estimate the welfare benefit from low prices vs. new kinds of titles coming to market. Section 7 then turns to an aesthetic perspective on the new products, asking whether the large sales of self-published work has crowded out the consumption of critically acclaimed books. We conclude that self-published books offer substantial welfare gains, both through reduced prices and by making available new varieties that would not earlier have been available to consumers. We find no evidence that the availability of such products has coarsened US book consumption. The evidence in this paper adds to emerging evidence elsewhere that

digitization has important benefits for both consumers and creators, even as it creates challenges for many existing intermediaries.

2.2 Background

2.2.1 The Traditional Publishing Industry

In order to make a book available to consumers, an author has traditionally needed the support of two kinds of entities. First, the author needed a major publisher to publish and promote the work. Second, the author also needed bookstores to choose to stock the work. Both of these kinds of entities have been dominated by a small number of gatekeepers.

The publishing industry has traditionally been dominated by a handful of major publishing houses. In 2012 there were six: Hachette, HarperCollins, MacMillan, Penguin, Random House, and Simon & Shuster. The merger of Penguin and Random House in 2013 reduced this to five⁴. Until the launch of Amazon, book retailing was dominated by bookstore chains such as Barnes & Noble and Borders and other retail chains (supermarkets, warehouse and discount clubs), which collectively had 81 percent of the retail market for trade books (Greco (2004)). Independent bookstores had the remainder. While the big box retailers maintained larger selection than most independent stores, even their large selection (as many as 140,000 titles in a Borders store) was small compared with the number of extant books⁵. More books were published each year than physical stores could maintain in inventory. Hence, physical retailing added a second filter between creators and consumers.

2.2.2 The Growth of E-Readers

While e-books can be read on computers, and have therefore been in principle available for over a decade, e-books are most useful to consumers when consumed on small handheld devices, such as e-book readers or tablet computers. The e-book market has grown quickly since 2007, driven largely by the success of e-readers from Amazon, Apple, and

⁴See Bosman (2013).

⁵See Austen (2011) for information on titles at Borders. Estimates of Amazon's title count top 1 million.

Barnes & Noble. While Sony had released some electronic book readers as early as 2004, the e-book market began in earnest with Amazon’s release of the Kindle, priced at \$399, in November 2007. The Kindle was well received by technology critics. Built in wi-fi for quick book downloading was widely applauded (Pogue, 2007). The Kindle 2 was released in February 2009, priced at \$259. Later that year (November) Barnes & Noble released the Nook, priced at \$259. The Apple iPad was introduced in April 2010, priced between \$499 and \$829, depending on options. In November 2010, BN released the Nook color at \$249. In November of 2011, Amazon released the Kindle Fire. As of late 2011, Kindles were available at a range of prices, depending on options. For \$79, one could buy an ad-supported Kindle with wi-fi. The same model was available at \$109 without ads⁶.

These readers have spread quickly. As figure 2.1 shows, the share of US adults owning an e-book reader grew from 2 percent in April 2009 to 19 percent in November 2012. The share owning a tablet grew from 3 percent in May 2010 to 25 percent in November 2012. The share some viable method for consuming electronic books either a tablet or an e-reader reached 33 percent in November 2012 (up from 18 percent before the December 2011 holiday (Rainie (2012))).

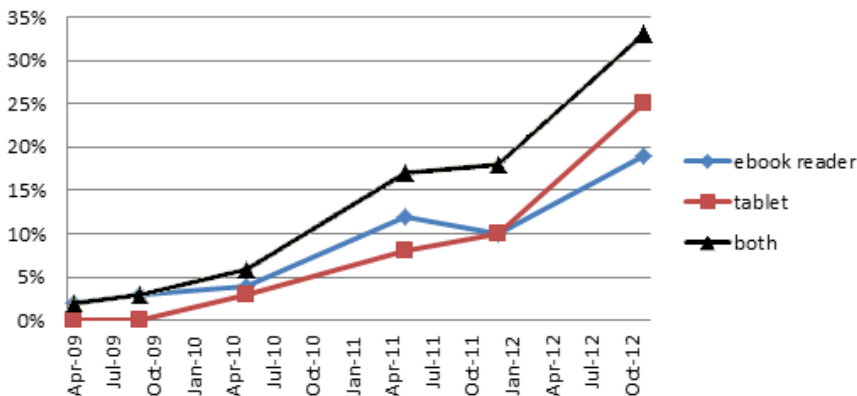


Figure 2.1: Percent of US Adults with Tablets and E-readers

⁶See Rainie, Zickuhr, Purcell, Madden, and Brenner (2012), page 15.

2.2.3 Electronic Books and Lower Marginal Costs

E-books have far lower costs than physical books. In 2010 the cost structure for a hardback book priced at \$26 included the following components: the publisher received \$13 in revenue and paid \$3.90 to the author, \$3.25 for printing, storage, and shipping, \$0.80 for design, typesetting, and editing, and \$1.00 for marketing⁷. The average per-copy cost was therefore roughly \$9. By contrast, for an e-book priced at \$9.99, the publisher received \$6.99 in revenue. The per-copy costs the publisher faces were about \$2.25 for the author, \$0.38 for digitization, typesetting, and editing, and \$0.60 for marketing. The average per-copy cost was roughly \$3. Of course, design, typesetting, editing, and marketing are not really marginal costs, while royalties and printing are more clearly marginal costs. Even including only these plausibly marginal costs gives a marginal cost of about \$7 for a physical book compared with \$2 for an electronic book.

2.2.4 The Appearance of Self-Publishing Channels

In addition to digitization's effect on book prices, digitization has also affected the supply side of the market. It is now possible for authors to make their works available direct to the public, without using a traditional publisher. Major providers of self-publishing services include Smashwords, Author Solutions, and Lulu, as well as Amazon. Amazon's Kindle Direct Publishing allows authors to sell their works through Amazon, receiving 70 percent of the sales price as a royalty (authors also pay Amazon some delivery fees). Smashwords offers a similar service, although Smashwords does not (as of March 27, 2013) distribute more than a few successful books through Amazon⁸.

2.2.5 Product Discovery Institutions

Consumers have traditionally learned about new books through book reviews authored by professional critics and published in established media outlets. The number of such book reviews has traditionally been small in comparison with the number of books released each year. In 2010, when roughly 200,000 books were released, traditional media outlets published 50,000 book reviews. The largest source of reviews are publications

⁷See Rich (2010).

⁸See the Smashwords distribution information page at <http://www.smashwords.com/distribution>.

aimed at bookstores and libraries (Publishers Weekly, Library Journal, and Kirkus). Table 2.1 shows the main institutions. Because many works are reviewed by multiple outlets, the number of reviews exceeds the number of works reviewed. Media outlets aimed directly at consumers issued far fewer book reviews. For example, the New York Times reviewed about 1,250 books per year, and the Washington Post reviewed roughly 1,000. While it is difficult to say how many works were reviewed, it is clear that only a small share of works released were reviewed.

Table 2.1: Major Book Review Outlets

Source	2010	2011	target
Booklist	8,457	7,978	librarians
Bookmarks	712	727	general readers, book groups
BookPage	828	682	Bookstores, public libraries
Bulletin of the Center for Children's Books	726		school and public librarians
Chicago Tribune Sunday Book Section	500	500	general readers
Choice	6,851	6,833	academic librarians
Horn Book Guide	3,967	4,266	teachers, librarians
Horn Book Magazine	428	482	industry professionals
Kirkus Reviews	4,524	6,197	industry professionals
Library Journal	6,099	6,590	industry professionals
Multicultural Review	378		educators
New York Journal of Books	1,225	1,372	general readers
New York Review of Books	394		general readers
New York Times Sunday Book Review	1,250	1,250	general readers
Publishers Weekly	7,884	7,835	industry professionals
School Library Journal	5,774	6,219	librarians
Washington Post Book World	884	1,136	general readers
TOTAL	50,881	52,067	

Source: Library and Book Trade Almanac 2012, p. 545.

The past few years have seen the development of other sources of information about books. These include crowd-sourced information, notably reviews and star ratings posted at retailers themselves (e.g. Amazon) as well as reviews and ratings posted at third party sites such as Goodreads, AllReaders, or BookPage.

Of the crowd-sourced sites, Goodreads is the largest. According to Alexa.com,

Goodreads is the 167th-ranked site in the US, with over 10 million users⁹. Their book coverage is broad: "Goodreads has 10 million reviews across 700,000 titles - one of the largest and deepest collections of quality book reviews on the internet."¹⁰ Goodreads was launched in 2007 and has therefore generated reviews for over 100,000 works per year since its launch. This represents substantially more book reviews than the traditional media sector produced over time.

The new information environment also includes bloggers and small-scale organizations that review books online. One can find lists of these independent reviewers in online guides to self-publishing. According to the Step-by-Step Guide to Self-Publishing¹¹,

"Book reviews are the best way to promote your book for free! A regular print ad in a newspaper can cost thousands of dollars. A book review only costs you the postage and the cost of the book. Got it?"

Although submitting books for review can be costly (with postage costs, etc), the rewards are enormous. And it's still a LOT cheaper than print ads. A positive book review on Amazon is like a shower of gold for writers."

The development of new information sources, along with the growth in new products, raises the possibility that consumers will be able to find a better set of books than would have been possible when the informed choice set was smaller. Whether this is so is an empirical question.

2.3 Theoretical Framework

Digitization, as will be detailed empirically below, has two effects on the markets for books. First, the ability to distribute books electronically reduces costs and has reduced prices as well. This gives rise to a movement along the demand curve for books and possible increases in both consumer and producer surplus.

Second, digitization has also allowed creators to circumvent traditional gatekeepers publishing houses and make their works directly available to consumers. The number of

⁹See G.F. (September 4, 2012).

¹⁰See the Goodreads website at <http://www.goodreads.com/api>.

¹¹See <http://www.stepbystepselfpublishing.net/reviewer-list.html>

new titles brought to market has increased substantially. If some of these are appealing to and discoverable by consumers, then the availability to this larger set of products can deliver welfare benefit to consumers.

Even before digitization, the book industry brought thousands of new products to market each year. Given the large number of available products, additional new products need not have added much of value to the choice set. Yet, the nature of media products makes it possible for growth in available products to have substantial benefits to consumers.

Books, like music and movies, are products whose success is difficult to predict at the time that investments are made. Caves (2000) describes the book business, along with music and movies, as industries in which "nobody knows" which products will succeed with consumers at the time that investments are made¹². While it is far less costly to bring a new book to market than a new movie or music album, the costs of traditional publishing still limit the number of new products. Traditionally, a publisher needed to acquire rights, usually with a five-figure advance, edit and print books, promote these works to critics and, in some cases, directly to consumers. Publishers also needed to ship books to stores and, usually, to incur the costs of shipping the unsold books back from stores¹³.

New digital technology effectively reduces costs and allows creators to bring new works to market without enlisting the use of the publishing industry's costly apparatus. As we detail below, this has given rise to large growth in the number of new titles available. Whether new varieties benefit consumers depends largely on the predictability of books' appeal at the time of investment. If publishers and authors had perfect foresight about books' appeal to consumers, they would then release all works with (ex ante and therefore ex post) revenue in excess of costs. A decrease in the cost of releasing works would raise the number of titles released, but the additional works would, by construction, have limited appeal (with revenues below the old threshold and above the new, lower one). The benefit of title expansion would be modest. But in the more realistic case of unpredictable appeal, however, an increase in the number of titles available can substantially increase the number of titles that turn out, ex post, to have

¹²Screenwriter William Goldman once famously declared that in Hollywood, "nobody knows anything" about which movies will succeed. See Turan (January 17, 2007).

¹³See Greco (2004) for detailed information about the book publishing industry.

significant appeal to consumers (see Terviö (2009) for a model that embodies this logic).

In our context the question is whether the newly available works, which would previously not have been available to consumers, end up as a significant share of the commercially successful books. For empirical purposes we take the self-published works to be examples of works which would previously not have been available to consumers¹⁴.

The recent history of the book industry, along with this theoretical background suggests a number of questions for this paper to explore. First, what has happened to the evolution of prices, units sold, and the number of book titles available over the past few years as digital technologies have diffused in the book market? Second, have the newly available titles brought much benefit to consumers, i.e. have they had much sales success? We turn to these questions below, beginning with a discussion of available data.

2.4 Data

Data availability is a major obstacle to the authoritative study of the market for books in general and electronic books in particular. Ideally, we would observe the full list of available new titles from the last few years, along with title-level sales and prices for both physical and electronic books by week. The available data fall far short of this ideal but still allow some meaningful analysis. Some data are available at the aggregate level, while others are available for individual titles over time. While Nielsen's Book Scan product has data on weekly sales of physical titles, there is not yet an analogous product for e-books. In this section we describe the sources of data along with some of the recent patterns as prelude to more detailed analysis of the following sections.

We create four different data sets using information drawn from ten distinct sources. The four datasets are 1) an aggregate dataset on the prices, quantities, and numbers of new books released by year, 2008-2011; 2) a title-level dataset on the weekly top 150 USA Today bestsellers, including their weekly ranks, genres, as well as information

¹⁴We realize that some self-published works are produced by authors with a history of publishing via traditional channels and who may have been able to release their self-published works through traditional channels. These appear to be a minority of cases, however. When authors can get released through traditional publishers, they typically choose that route. Many of the works in the data below originally appeared as self-published works but were picked up by publishers as the work became popular. In those cases the self-published works leaves the market, replaced by a traditionally published version.

on whether each title was originally self-published, 3) a 2012 annual title-level dataset on the prices and quantities sold for over 1000 top-selling titles according to e-book sales, and 4) a list of 100 critically acclaimed books (*New York Times* "notable books") published each year, 1997-2012, that can be matched with the USA Today bestsellers.

2.4.1 Aggregate Data Sources

We have aggregate data on US quantities sold, for physical and electronic books, for 2008-2012. Data on top-level quantities sold are from BookStats, which calculates total sales based on surveys of a large number of publishers.

We have aggregate data on retail prices from a few different sources. One systematic source is the Library and Book Trade Almanac 2012 (Bogart 2012), which reports data on the number of new titles by format and subject along with their retail prices. The Almanac obtains these data from book wholesaler Baker & Taylor. The prices they report are the simple averages across titles; they are not weighted by sales of books.

The Bureau of Labor Statistics CPI provides another source of information on book prices. The BLS collects data on the prices of recreational books. These data appear to include electronic books. The instructions for interviewers include two book categories, "books purchased through book clubs" and "other books, audio books, or e-books."¹⁵ The latter category includes "all new books, audio books, and downloadable e-books purchased individually."

Fragmentary data on e-book prices are available from other sources. A firm called iobyte solutions collects data on the prices of bestselling books, and they report the distribution of prices over the past 6 months by cells (\$0-2.99, \$3-7.99, \$8-9.99, \$10 and over).

We have two sources of data on the number of new titles. Bogart (2012) reports the number of new titles available in each year, overall and for fiction in particular. Bowker, the publisher of the Books in Print database, also produces an annual time series on new titles. Their overall time series does not distinguish physical from electronic books. Bowker also produces counts of the number of self-published works 2006-2011, distinguishing physical from electronic¹⁶.

¹⁵See the BLS website.

¹⁶See Bowker (2012).

2.4.2 USA Today Bestseller List Data

We have two title level datasets with information about the sales of both electronic and physical books. First, we have the USA Today weekly bestseller ranking. USA Today produces a single weekly list based on overall physical and electronic sales. In addition to producing a ranking of the top 150-selling books each week, the USA Today list indicates the format (hardback, paperback, or electronic) selling the most copies of the title this week. The USA Today list, available back to 1993, includes over 1,000 separate titles per year. The list includes the name of the publisher, which is helpful for determining whether a book is self-published, as well as a genre designation.

Determining which books are self-published requires some detective work. Some are easy. We deem a work self-published if the listed publisher contains the word "self" (e.g. "self-published via Amazon"). We also include works published by the major self-publishing services listed in Bowker (2012). These include Smashwords, Lulu Enterprises, and various divisions of Author Solutions (Xlibris, Authorhouse, IUniverse, and Trafford). These services collectively account for about three quarters of self-published electronic books. We also found works that had originally been self published using on-line sources. Table A1 lists all of the works that we identify as originally self-published.

2.4.3 Publishers Weekly Top 1000 E-books for 2012

Our second title level dataset has quantities and prices, for both electronic and physical formats, for the 1000 top-selling electronic books of 2012. Data are derived from three separate sources. The list of books comes from Publisher's Weekly's list of the 1000 top-selling electronic books for 2012. We obtain physical sales for these titles from Nielsen BookScan, which contains title-level data on sales of trade books.

We obtain price data for these 1000 titles, in electronic and physical formats, from Amazon.com. Title-level price data for electronic books are not readily available. It is of course easy to get the prices of current bestsellers at the Amazon Kindle Store; and there are some collections of average e-book prices for the last few years. But there is no obvious way to get the prices of particular e-book titles from the past. While Amazon provides an annual bestseller list archive that preserves the sales ranking order, the prices posted on the list are current prices. If a book had been self-published in the

year of the list but is now available through a major publisher, the list includes no price information. We obtained the price data in early 2013, so these prices are those prevailing at the time of data collection rather than during 2012.

2.4.4 Title-level Information on Physical Books

The aggregate data are useful for making a rudimentary calculation of the welfare benefit from lower prices, but without information on sales volume of new self-published titles we cannot say much about the benefit that they produce for consumers. To remedy this, we attempt to construct a title-level dataset including sales of each title in physical and electronic forms, as well as an indicator of whether the work was originally self-published.

Nielsen BookScan collects title-level data on sales of trade books, and they have these data back to 2003. BookScan has two shortcomings. First, Nielsen allows only-title-by-title access to their data. Given the time it takes to obtain a single title, it is difficult to obtain the universe of literally hundreds of thousands of titles published per year. Second, Nielsen does not include e-books in their data.

While these shortcomings limit what one can do with Nielsen data, they have some offsetting benefits. Mainly, Nielsen does compile weekly bestseller lists, which include the 100 bestselling trade titles that week, along with the weekly sales of each title. These are available back to 2003. Each entry includes both the quantity sold and the retail list price. The Nielsen data also include the publisher and imprint, along with the ISBN number. The data do include self-published books but only if they have physical sales. Aggregating the Nielsen weekly top 100 gives 94.9 million units sold in 2009, a year in which the total sales of physical trade books was 2.2 billion. By 2011, when total physical sales were 2.1 billion, the Nielsen weekly top 100 books accounted for 72.2 million.

2.4.5 Title-level Information on Electronic Books

We are aware of no systematic direct source of weekly information on the title-level sales of electronic books. That said, Publisher's Weekly (PW) has released lists of the top-selling electronic book titles for 2010, 2011 and 2012. For 2010 they included titles selling more than 10,000 copies; for 2011 and 2012, they included titles selling over

25,000. Their list does in principle include self-published books, but it excludes titles priced below \$5.00. De facto, this excludes many self-published titles. PW includes 30 adult, and 82 children's title electronic sales for 2010. PW's 2011 list included 329 adult titles. The 2012 list includes 1006 adult titles and 153 children's titles. Electronic sales of these titles total 6.6 million, 26.6 million and 48.8 million for 2010, 2011 and 2012, respectively. In about a quarter of cases, PW reports only a sales rank rather than a quantity. We can impute a quantity by regressing $\log(q)$ on the book's log-rank, where the rank is within year and type (adult vs children). Using this imputation we observe title-level electronic book sales of 7.2 million in 2010, 33.8 million in 2011 and 95.1 million in 2012.

Bestseller lists do convey some information about the sales of both electronic and physical books. USA Today produces a single weekly list based on overall physical and electronic sales, and self-published books are included. In addition to producing a ranking of the top 150-selling books each week, the USA Today list indicates the format (hardback, paperback, or electronic) selling the most copies of the title this week. The USA Today list, available back to 1993, includes over 1,000 separate titles per year. The list includes the name of the publisher, which is helpful for determining whether a book is self-published.

2.5 Descriptive Results

2.5.1 Evolution of Physical and E-book Quantities

Table 2.2 shows data derived from BookStats on sales of physical and electronic books. They report 2008 sales of 2.16 billion trade books, a category that includes adult fiction, adult nonfiction, juvenile books, and religious books. (The major categories not included in trade books are professional and educational books). Of these sales, the majority (68 percent) were various forms of softcover books (trade and mass market paperbacks), while just over a quarter (26 percent) were hardcover books. E-books made up a very small share (0.4 percent). Since 2008, trade paperbacks and hardcover books have held relatively steady. By contrast, sales of mass market paperbacks (the inexpensive paperback format printed on low-quality paper and designed to be sold on racks in airports, etc.) have fallen, and sales of e-books have risen sharply. In 2011 total trade

units sold stood at 2.50 billion, an increase of 10.7 percent in units sold over 2010. Electronic books accounted for 15.5 percent of the trade units sold in 2011.

Table 2.2: Trade Book Sales by Book Format and Type, 2008-2011 (millions of units)

	Format				
	Total	hardcover	mass market	softcover	e-book
2008	2,164.1	623	383.4	1,078.0	30.1
2009	2,211.1	678	355.5	1,077.8	58.3
2010	2,261.3	648	408	1,200	128.2
2011	2,503.2	596	279	1,200	389.8
2012*	2,625.2				524.3

	Type				
	Total	Adult Fiction	Adult Nonfiction	Juvenile Kids & Young Adults	Religion
2008	2,164.1	593.0	524.8	844.7	201.6
2009	2,211.1	616.2	508.7	909.3	176.9
2010	2,261.3	704	474	936	272
2011	2,503.2	658	480	972	374
2012*	2,625.2				

Source: various media reports on BookStats 2011, 2012.

2012 data are estimated based on August, 2012 year-to-date revenue.

The symbol * denotes estimated.

Full-year data for 2012 are not yet available, but as of November 2012 the Association of American Publishers has released year-to-date data on the change in category revenue relative to 2011. Relative to 2011 year-to-date revenue in adult hardbound books was down by 5.3 percent, up by 7.7 percent in softcover trade, down in mass market paperbacks by 12.5 percent, and up in e-books by 34.5 percent. Applying these revenue percentage changes to the unit totals for 2011 gives rough estimates of 2012

quantities. E-book units are thus estimated at 524.3 million, while combined physical units are estimated to be 2.1 billion, putting total units at 2.6 billion¹⁷. This is an increase of 4.8 percent over 2011 unit sales.

We get another glimpse at the growth of electronic books from the share of USA Today listings for which the electronic format was the best selling. Figure 2.2 shows the share of listings for which the electronic version was the bestselling edition, by week. The share was essentially zero prior to 2010 and rose to about 3 percent during 2010. The share jumped markedly at the start of 2011, most likely because of the heavy volume of e-readers given as gifts at Christmas during 2010. The share was roughly steady at 25-30 percent during 2011, then fell toward the end of the year, presumably because e-books are not popular as holiday gifts. The e-book share then jumped to 70 percent at the start of 2012, again because recipients of e-book readers made their first e-book purchases immediately after Christmas. During 2012, the share of listings selling best as e-books has fluctuated between 30 and 40 percent. After another dip in December 2012, the e-book share is now around 50 percent. This figure suggests that e-book sales are continuing to grow, although the growth between 2010 and 2011 appears larger than the growth between 2011 and 2012.

¹⁷We weight the percentage change for distinct physical categories (hardbound, softcover trade, mass market) using their respective 2010 shares of physical sales. These imply 2.7 percent increase in physical units 2011 to 2012.

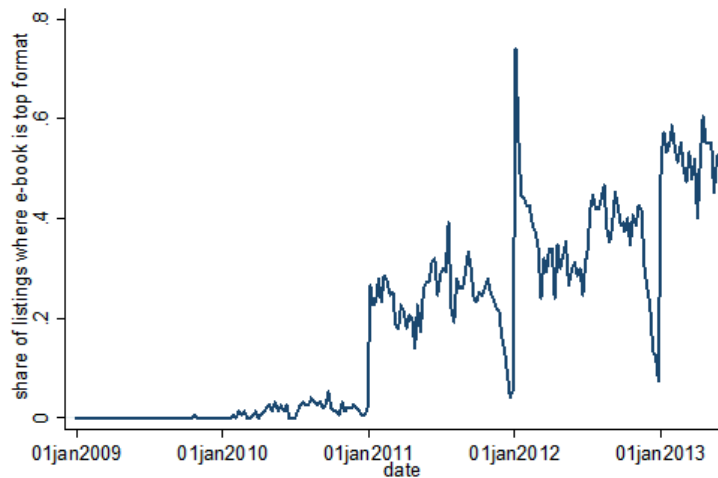


Figure 2.2: Growth of 'Mainly E-book' Entries in USA Today List

2.5.2 Evolution of Prices

Table 2.3 reports average prices (in 2011 dollars, adjusted using the CPI) for overall titles as well as for fiction alone, by books formats, from the Almanac. A few patterns are clear. First, especially for fiction, mass market paperbacks and e-books are substantially less expensive than other formats. Second, physical book prices appear to be relatively constant, while e-book prices are falling. For example, the average fiction e-book had a retail price of \$9.10 in 2008 and \$5.31 in 2011, a reduction of 42 percent.

Table 2.3: New Book Prices

total				
year	hardcover	e-book	trade paperback	mass market
2008	35.35	59.95	41.12	6.79
2009	35.78	46.45	41.10	7.02
2010	35.16	42.92	43.39	7.05
2011	34.42	27.34	36.98	6.97

fiction				
year	hardcover	e-book	trade paperback	mass market
2008	30.33	9.10	17.03	6.77
2009	30.18	8.61	18.16	7.00
2010	33.22	7.28	18.56	7.01
2011	29.60	5.31	18.25	6.95

Source: Library and Book Trade Almanac 2012.

Prices inflated to 2011 levels using the CPI.

The CPI price series for recreational books for all urban consumers is displayed in Figure 2.3, 2002-2012. This series is not seasonally adjusted, and December 1997=100. The price level for books is steady at 104 between 2002 and 2007. It then rises from 2007 until the end of 2009, reaching a peak at 107. Since the beginning of 2010, the series has been falling fairly steadily, reaching 100 in late 2012. In the past two years, the price of books has fallen by 6.9 percent according to the CPI. Given that the overall CPI rose by nearly 9 percent between late 2009 and late 2012, the price of books has fallen 14.6 percent in real terms over this period.

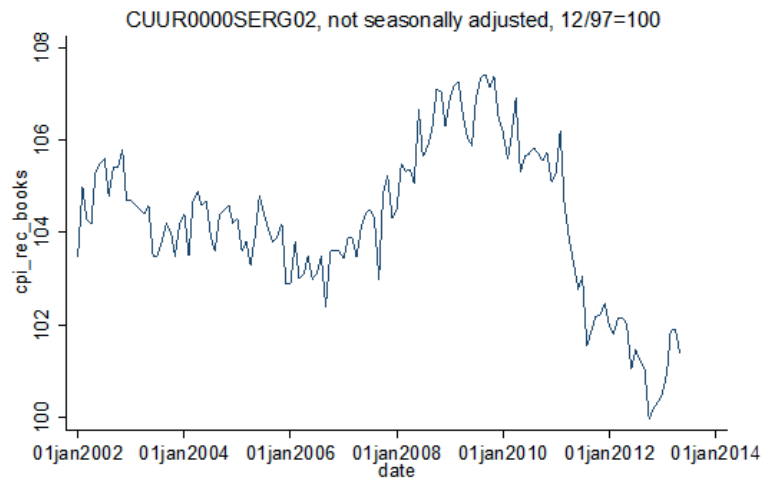


Figure 2.3: US City Average CPI for Recreational Books

Figure 2.4 report the iobyte evidence on the share of Kindle top 100 Bestsellers priced below \$8 and \$10, respectively, between April 2012 and March 2013. The share priced below \$10 has risen from 60 percent in late April 2012 to almost 90 percent in March 2013. The share priced below \$8 has risen even more sharply: from 40 percent in late April to nearly 80 percent in a year later. The iobyte evidence corroborates that declining list prices for electronic books resulted in lower prices paid for books with substantial sales during 2012.

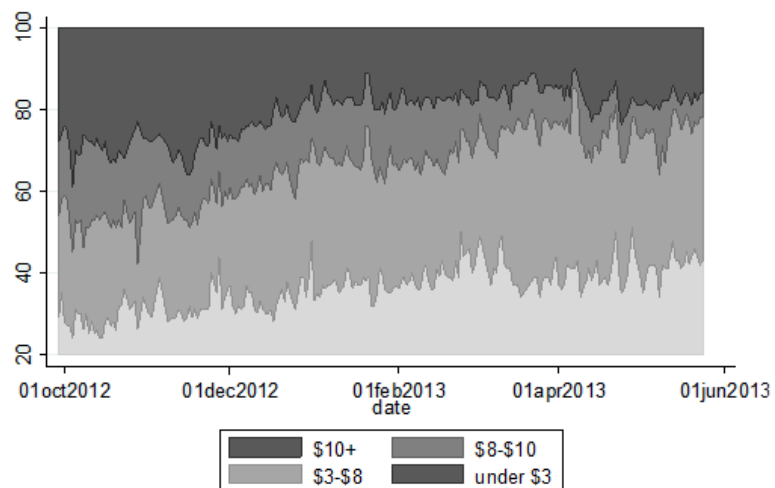


Figure 2.4: Price Distribution for Top 100 Amazon Kindle Bestsellers

It seems clear that book prices are falling, but list prices (such as those in Table 2.3) likely overstate actual prices paid for books. For an estimate of current price levels for physical and electronic books we monitored Amazon's top 100 bestselling physical and electronic books during November and December 2012. Weighting prices by a rough estimate of sales (the reciprocal of the sales rank), the average e-book price was \$7.95, and the average physical book price was \$13.83. During 2012 e-books have made up 19.6 percent of units sold (using our estimates in Table 2.2), so the average price paid for a book during 2012 is $0.196(7.95) + (1-0.196)(13.83) = \12.68 . Using the annual average of the real CPI for recreational books (books CPI/overall CPI), we can construct a series of book price levels. The average 2012-dollar price of a book by this approach was \$14.13 in 2008, \$14.31 in 2009, \$13.93 in 2010, \$13.19 in 2011, and \$12.68 in 2012.

2.5.3 The Number of New Books

Table 2.4 reports the number of new titles available in each year, overall and for fiction in particular. These data are drawn from Bogart (2012). The number of new physical titles overall was 180,032 in 2008, with 18,638 in fiction. Since 2008, the number of new physical titles has been essentially constant: total titles numbered 177,126 in 2011, while physical fiction titles numbered 19,760. The number of electronic titles, by contrast, has increased substantially. Total new electronic titles grew from 35,495 in 2008 to 111,150 in 2011. For fiction alone, the number of new electronic titles grew from 7,414 in 2008 to 39,886 in 2011.

Table 2.4: New Titles (Baker & Taylor)

year	total titles		fiction titles	
	physical	electronic	physical	electronic
2008	180,032	35,496	18,638	7,414
2009	178,841	53,731	18,272	13,364
2010	186,344	67,145	17,971	18,043
2011	177,126	111,150	19,760	39,886

Source: Library and Book Almanac 2012.

Bowker, the publisher of the Books in Print database, also produces an annual time

series on new titles, which we report in Table 2.5. Their overall time series does not distinguish physical from electronic books. According to Bowker, the number of new titles grew from 215,138 in 2002 to 347,278 in 2011. Fiction alone increased from 25,102 to 60,075 over the same period. Bowker also produces counts of the number of self-published works 2006-2011, distinguishing physical from electronic. In 2006 there were 60,875 new self-published works released, roughly 90 percent of them physical. The number of new self-published physical works has tripled, reaching 148,424 in 2011. The number of self-published electronic works has grown by more than 10 times, from 7,758 in 2006 to 87,201 in 2011.

Table 2.5: Total and Self-Published Titles

	Total works		Self-published		
	total titles	fiction	total	print	electronic
2002	215,138	25,102			
2003	240,098	24,666			
2004	275,793	38,832			
2005	6251,903	34,927			
2006	274,416	42,777	60,875	53,117	7,758
2007	284,370	53,590	74,400	66,459	7,941
2008	289,729	53,058	83,751	75,800	7,951
2009	302,410	48,738	109,019	94,826	14,193
2010	328,259	53,139	149,594	111,551	38,043
2011	347,278	60,075	235,625	148,424	87,201

Sources: Bowker's "New Book Titles and Editions, 2002-2011"

The first column ("total") is what Bowker terms "subtotal".

Alone and together, these data sources indicate a substantial growth in the number of new titles available each year since 2006. The increase in the number of new works has been driven entirely by self-published books; and many of these are electronic.

2.5.4 Sales Concentration

We have seen above that there has been a substantial growth in the number of new products available to consumers, driven largely by self-published books. Has the availability

of these new products attracted consumption? We address this question in two parts. A first question is simply whether the sales concentration of books has declined. If the new products are sufficiently appealing to attract consumption away from traditional products, then we should see a decline in sales concentration.

We can construct a rudimentary measure of sales concentration directly from the USA Today ranking. Each week the ranking includes 150 books; the list therefore includes 7,800 (52 x 150) entries over the course of the year. Figure 2.5 shows the number of distinct works on the list each year, which provides a measure of sales concentration. Between 1994 and 2000 the number hovered between 1,000 and 1,100. Between 2001 and 2005 the number fluctuated between 1,100 and 1,200. Since 2005 the number has increased more quickly and reached 1,600 in 2012 (the last full year of data). The decrease in sales concentration is consistent with the availability of new products drawing consumption away from existing products.

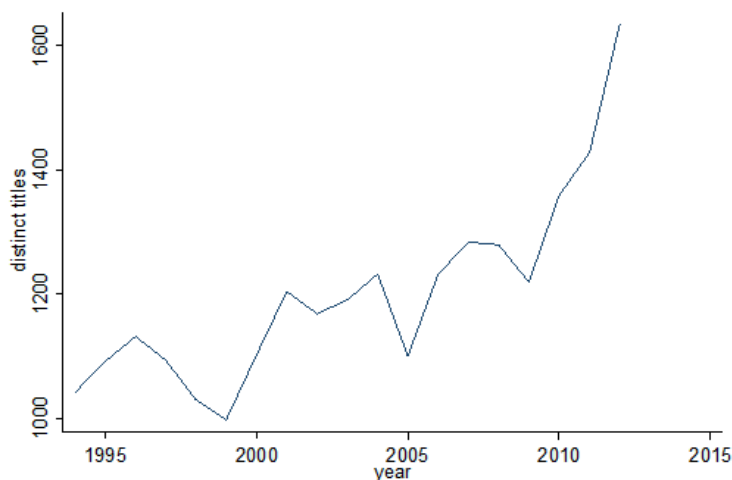


Figure 2.5: Number of Distinct Titles on USA Today List per Year

2.5.5 Are Self-Published Works Consequential

A second and more direct measure of whether the new products are appealing to consumers is simply the share of bestseller list entries accounted for by works that were originally self-published, displayed in Figure 2.6. Zero prior to 2011, the share rose to 4 percent by mid-2011, fell to 2 percent, then rose above 6 percent by the end of the

year. In 2012 the share has continued to fluctuate but has reached a higher peak of 10 percent mid-year. In the first quarter of 2013, the share has consistently been above 10 percent. Self-published works have had their largest impact in the romance category.

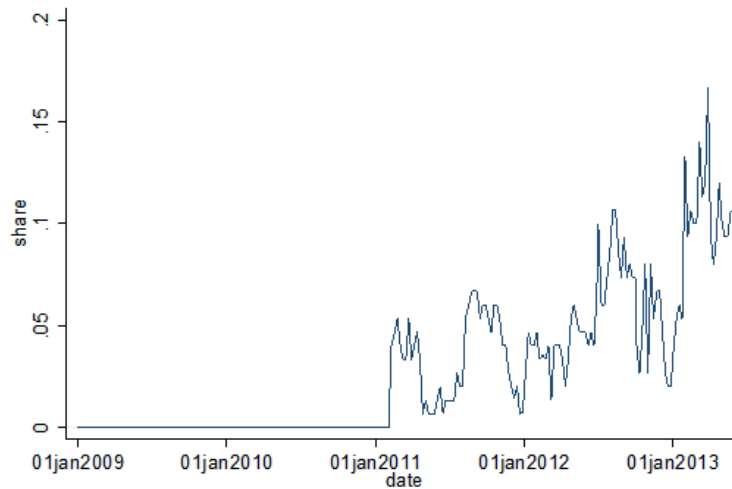


Figure 2.6: Share of Bestseller Listings Originally Self-Published: All Genres

Figure 2.7 shows the self-published share of listings among romance books. The share reaches 20 percent during 2011 and 30 percent during 2012. In the first months of 2013 it has been as high as 50 percent. It is clear that self-published works, which previously would not have meaningfully made their way to consumers, have rapidly become a significant share of total sales. That self-published books make up a large share of titles is suggestive that self-published works also make up a large share of consumption, but drawing that inference requires data on sales quantities of both physical and electronic books by title.

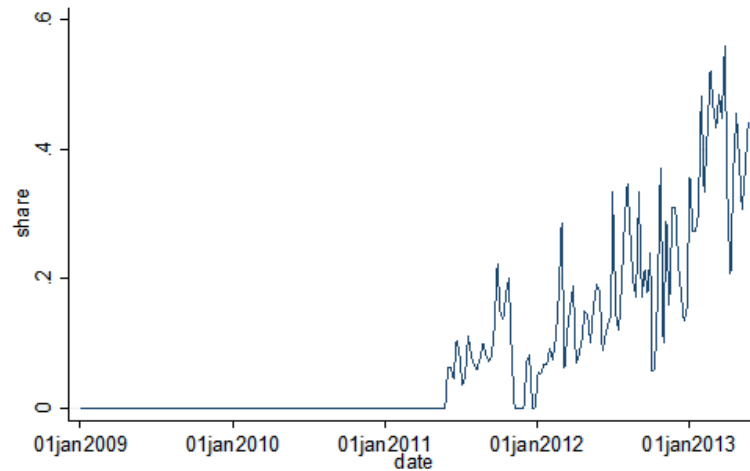


Figure 2.7: Share of Bestseller Listings Originally Self-Published: Romance Genre

2.6 Digitization and Welfare

2.6.1 Surplus for Market Participants: Basic Approach

Cost reductions and reduced entry barriers can give rise to three sorts of changes. First, price reductions can raise consumer surplus. Cost reductions in conjunction with price changes may change producer surplus. Finally, entry of new products will consume some additional fixed costs. Here we attempt to tally these welfare changes.

How much have consumers benefitted from e-books? To a first approximation we can view the past few years of price reductions driven by new cost-reducing technology as a movement along the demand curve. In 2008, the average price of a book was \$14.13, and consumers purchased \$2.164 billion books. In 2012, the average price was \$12.68, and consumers purchased 2.707 billion books. With a linear approximation to demand, consumer surplus in 2012 is \$3.5 billion higher than in 2008. This is 10.2 percent of 2012 revenue. Since 2008, the cumulative increase in consumer surplus, relative to the 2008 level, is \$5.7 billion. The results are summarized in table 2.6.

Table 2.6: Increased Consumer Surplus from Price Reduction

year	price	quantity	Revenue	Δ CS year-to-year	Δ CS rel to 2008	Cumulative Δ CS
2008	14.13	2,164.1	30,569.5			
2009	14.31	2,211.1	31,635.7	-398.1	-398.1	-398.1
2010	13.93	2,261.3	31,501.7	842.8	431.4	33.3
2011	13.19	2,503.2	33,014.6	1,767.2	2,186.1	2,219.4
2012	12.68	2,707.6	34,332.4	1,326.0	3,521.6	5,741.0

Note: all quantities and dollar figures (except price) in millions of 2012 dollars.

What are the effects on producers? The marginal cost of a physical book is essentially \$7, while the marginal cost of an e-book is essentially \$2 (this calculation treats royalties paid to authors as a component of marginal cost). To a first approximation, then, the gain in producer surplus from is the $(Q_{e-bookssold}) * (5(P_{physical} - P_{e-book}))$. Based on the Amazon prices cited above, this is $Q \cdot (5(13.83 - 7.95)) = Q \cdot (55.88) = -0.88 \cdot Q$. In 2012, this is a reduction of \$0.47 billion.

Finally, what has happened to the fixed costs of creation? Between 2007 and 2011 the number of new titles released annually has grown by about 75,000. The additional cost is the cost of producing 75,000 new books, which consists largely of the time spent researching and writing the books. For a conservative estimate of the time required to produce a book we can look at a prolific writer. Stephen King has written many novels and a large number of short stories over his career. His first book, *Carrie*, was published in 1974, and he has written steadily since then. Figure 2.8 shows King's cumulative page output over time. King produces about 800 pages of published fiction, or about 2 novels, per year. Given King's legendary prolific nature, using his productivity gives a conservative estimate of time cost, of 6 months of full time work per novel. It is not clear what wage the aspiring writers might otherwise have earned. If we use the average hourly pay of \$23.50, it costs society an average of \$27,000 per work. If we instead use the federal minimum wage of \$7.25, a work costs \$7,250 to create. The additional 75,000 works thus cost society between \$544 million and \$2.025 billion per year.

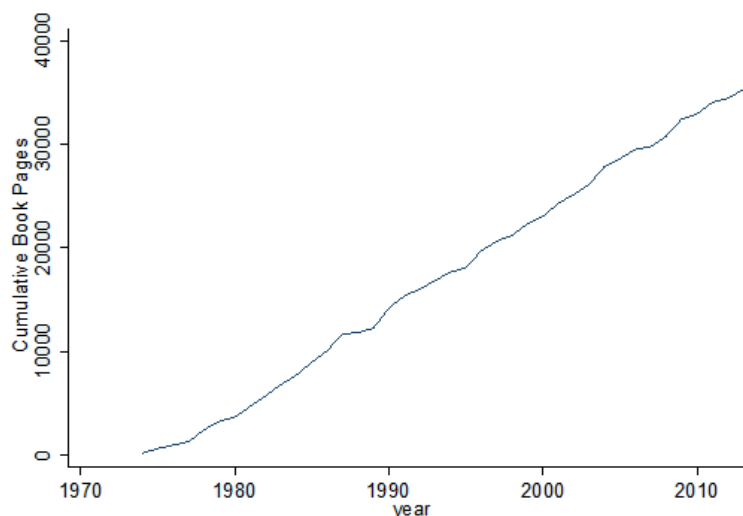


Figure 2.8: Stephen King's Cumulative Page Output Since 1977

The gains to consumers appear to exceed the costs to other parties.

2.6.2 Differentiated Products Approach to Welfare

The previous section provides a crude estimate of the welfare changes under the assumption that e-books are perfect substitutes for physical books and that self-published titles are perfect substitutes for more traditionally published works. In reality e-books may be associated with a different utility than their physical counterparts. Likewise, self-published books (the "new" set of titles that would not have made it to the consumer before) may not have the same appeal to consumers as traditionally published books do. Consumers might expect higher quality titles from traditional publishers as they would only publish those works that show promise to be successful. On the other hand some of the most successful titles of 2012 were self-published (most notably the *Fifty Shades* books). Whether electronic books and self-published titles provide more or less utility then becomes an empirical question.

Tables 2.7 and 2.8 suggest that consumers prefer physical copies over their electronic counterparts. Table 2.7 shows mean prices for physical and electronic versions of the titles in the dataset as obtained through Amazon. On average self-published books have a lower price than titles that were published "traditionally" (by around \$1), and e-books

are cheaper than physical books (by \$2.50).

Table 2.7: Amazon Prices in 2013 - by Format and Publisher

		Mean	Std. Deviation	Min	Max
Traditional	Physical	11.15	6.714	0.99	150.20
	E-book	8.53	3.319	0	36.99
Self	Physical	9.86	4.225	2.99	21.6
	E-book	7.62	2.153	2.99	9.99

While physical books are on average \$2.50 more expensive than e-books, the sales quantities of the two formats are comparable, with physical books being sold slightly more than e-books. This holds true for traditionally published titles while self-published titles tend to sell more e-books. This is expected as self-published works emerged as a consequence of the increased ease of bringing a work to the consumer through e-books. Several self-published books have indeed been available in electronic formats before they became available as physical copies. The mean quantities are higher for self-published titles than for those that are not self-published. This is mainly due to the success of the *Fifty Shades* books but it holds true even if we disregard the series.

Table 2.8: Quantities Sold in 2012 - by Format and Publisher

		Mean	Std. Deviation	Min	Max
Traditional	Physical	77,250	110,571	121	919,991
	E-book	73,826	93,482	25,000	1,000,000
Self	Physical	1,206,024	2,149,537	307	6,417,207
	E-book	1,358,383	2,207,777	25,898	5,000,000
Total	Physical	91,584	286,185	121	6,417,207
	E-book	90,138	293,487	25,000	1,000,000

To account for the possibility that the new products provide a different utility to consumers than their more traditional counterparts do, we estimate a differentiated products demand system by using the micro level data from Publishers Weekly and

Nielsen to separately estimate the welfare benefits from low prices and the new products (e-books and self-published titles). The dataset contains 968 different titles, each in two formats (electronic and physical). Of the 1936 observations, 26 are identified as self-published.

We assume that a consumer's mean utility for a book is linear in its format and price, conditional on the title's quality. We let

$$\delta_j = \beta_0 + \beta_e ebook_j + \beta_p p_j + \xi_i + \epsilon_j$$

be the utility that a consumer derives from obtaining title-format combination j , where β_e describes the consumers' mean utility for e-books relative to physical books, β_p denotes the utility for price and ξ_i is the utility for title i (a title dummy variable). Define the relevant market as all Americans deciding whether to buy a book each month. Given the US population, the total market size is 3.6 billion. Assuming that ϵ_j follows a type I extreme value distribution, we can estimate the following nested logit equation:

$$\ln(s_j) - \ln(s_0) = \beta_0 + \beta_e ebook_j + \beta_p p_j + \xi_i + \sigma \ln\left(\frac{s_j}{1 - s_0}\right) + \epsilon_j,$$

where s_j and s_0 are the market shares for product j and the outside good, respectively, and σ shows the degree of substitutability of the bestselling books compared to the outside good (reading a book that is not among the top 1000 e-books of 2012 or not reading at all).

It is difficult to obtain reasonable values for β_p and σ in a full demand estimation given the limitations of the dataset, which is constricted to title-format level observations and a single time period. We instead estimate the equation assuming different values of β_p and σ . We choose values of σ between 0 and 1, where a value close to 1 indicates that the bestsellers are close substitutes and a value close to 0 means that there is no particular correlation of tastes for bestselling books. The price coefficient is chosen based on what is known about marginal costs and markups in the book publishing industry and under the assumption of profit maximization: an edition's markup is the inverse of its price elasticity of demand. We expect a 20 percent markup for physical books (which translates into a price elasticity of -5) and close to a 100 percent markup for e-books due to the low marginal cost (which leads to a price elasticity close to -1).

The true average price elasticity will be between these values.

In alternative specifications, we set the price elasticities to -1, -2 and -5, respectively. Using the nested logit formula for a product's own-price elasticity (see appendix A.3), the coefficient β_p is then calculated as:

$$\beta_p = E \left[\frac{-(\epsilon_j \cdot (1 - \sigma))}{p_j \left(1 - \sigma \frac{s_j}{1 - s_0} - (1 - \sigma)s_j \right)} \right]$$

Rewrite the estimation equation, given our values of β_p and σ , to estimate β_e and the title coefficients ξ_i :

$$\ln(s_j) - \ln(s_0) - \beta_p p_j - \sigma \ln \left(\frac{s_j}{1 - s_0} \right) = \beta_0 + \beta_e ebook_j + \xi_i + \epsilon_j$$

Estimation is done after merging the data for annual demand for physical books (Nielsen) and for electronic books (Publishers Weekly) with current prices from Amazon.

Table 2.9 shows the calculated price coefficient as well as the estimated utility from e-books as compared to physical books and the consumers' mean preference for e-books versus physical versions given several taste correlations σ . Controlling for prices, for any value of σ consumers prefer physical editions over electronic books. Depending on the chosen price elasticity and taste correlation, consumers prefer physical editions over e-books by between 50 cent (unit elasticity) and \$2 (mean elasticity of -5). Considering the relative prices and quantities across formats, we report the demand estimation results for a mean price elasticity of -5¹⁸.

Table 2.9: Demand Estimates: Mean Price Elasticity is Set to -5

σ	0	0.25	0.5	0.75	0.9
E-book	-1.249*** (0.129)	-0.9370*** (0.0969)	-0.6247*** (0.0646)	-0.3124*** (0.0323)	-0.1249*** (0.0129)
Price	-0.6064	-0.4548	-0.3032	-0.1516	-0.0606
Title FE	Yes	Yes	Yes	Yes	Yes

Std. errors in parentheses; Significance levels: *** < 0.01, ** < 0.05, * < 0.1

We determine differences in consumer surplus between several different scenarios to

¹⁸The results for other price elasticities can be found in appendix B.2.

separately identify the price effect and the variety effect (self-publishing) from e-books. First, we simulate a world without e-books and without self-published titles. This is an approximation of the world in 2008 - before e-books gained popularity. Second, we imagine that e-books are available (at various price levels) but there are no self-published titles. Third, e-books and self-published titles exist, but their prices are not the same as traditional physical book prices. Last, we use today's world, with e-books and self-published titles and their current prices¹⁹

The differences in consumer surplus are given by:

$$\Delta E[CS] = \frac{1}{\alpha} \left[\ln \left(1 + \left(\sum_{j \in J_1} \exp \left(\frac{\delta_j}{1 - \sigma} \right) \right)^{1 - \sigma} \right) - \ln \left(1 + \left(\sum_{j \in J_0} \exp \left(\frac{\delta_j}{1 - \sigma} \right) \right)^{1 - \sigma} \right) \right], \quad (2.1)$$

The δ_j terms are taken from the previous estimation, in which we used price elasticities of demand to determine a preference for e-books versus physical editions.

This approach and the estimates from above give the mean consumer surplus differences in table 2.10 (as compared to a world without e-books or self-published titles). Much of the increase in consumer surplus comes from a decrease in the price level while the addition of e-books per se did not increase consumer surplus as much. Consider the case when $\sigma = 0.75$ - the bestselling titles are relatively good substitutes for one another. Adding the new format without decreasing the price level or introducing self-published books increases consumer surplus by about \$815 million, while introducing the lower prices of electronic versions increases consumer surplus by \$2.17 billion. The subsequent introduction of new titles (the self-published works) raises consumer surplus by an additional \$450 million (a total consumer surplus increase of \$2.62 billion), and introducing the self-published books at their new, lower price increases consumer surplus by a total of \$2.84 billion. This change in consumer surplus is smaller than the \$5.7 billion calculated as a first approximation because e-books provide less utility per edition sold (an equivalent of about \$2.50 per book), and because self-published works may be giving less utility as well, although the title coefficients ξ_i are not significantly different between the traditionally published and self-published titles.

¹⁹While the question we ask is similar to the question Hausman and Leonard (2002) address, our approach is more straightforward as we do not observe and hence do not account for price changes of the existing products.

Table 2.10: Compensating Variation, Million \$, Total in 2012 (Compared to no E-books and no Self-Publishing), Price Elasticity = -5

σ	0.0	0.25	0.5	0.75	0.9
p_e high, no self	119.320	256.106	495.065	815.139	1,006.582
p_e high, p_{self} high	187.373	394.219	748.074	1,214.783	1,492.445
p_e low, no self	378.289	758.122	1,377.205	2,166.377	2,630.664
p_e low, p_{self} high	486.372	951.778	1,693.524	2,624.941	3,170.363
p_e low, p_{self} low	539.805	1,044.786	1,841.556	2,835.475	3,416.340

Note that as the substitutability σ between bestsellers increases, the differences in consumer surplus increase as well. This is partly due to the functional form of the estimation and the consumer surplus calculation, which takes a function of utility to the power of $(1 - \sigma)$. It can also be explained by the fact that a large σ corresponds with a strong preference for books. If options are taken from a highly liked choice set, a consumer's utility will decrease more than if we don't like the options as much to begin with.

2.7 The Coarsening of Product Offerings and Consumption

The Fifty Shades trilogy includes by far the best-selling self-published books. While they have sold extremely well, the work is not respected by critics. For example Jen Doll, writing in the Atlantic Wire wrote,

"Look, I'm not afraid to say it: 50 Shades of Grey is a terrible book. I know this because I have started reading it. It didn't take long to figure out. The writing is stilted and relies on tropes that anyone who's ever sat through 15 minutes of a high school writing workshop would know to avoid. The

characters are two-dimensional and stereotypical.²⁰”

Others have criticized the quality of self-published works. Bestselling author Sue Grafton raised a firestorm of controversy during 2012, saying in an interview that,

To me, it seems disrespectful that a 'wannabe' assumes it's all so easy s/he can put out a 'published novel' without bothering to read, study, or do the research. Self-publishing is a short cut and I don't believe in short cuts when it comes to the arts. I compare self-publishing to a student managing to conquer Five Easy Pieces on the piano and then wondering if s/he's ready to be booked into Carnegie Hall.²¹

These concerns raise the question of whether the availability of work that critics view as low-quality has coarsened consumption patterns. One way to address this is to quantify the overlap between the bestseller list and the books deemed by critics to be of high quality. The New York Times produces a list of 100 "notable books" each year. We can ask how many of these overlap the USA Today bestsellers as well as how the overlap varies over time.

Figure 2.9 shows the number of NYT Notable works from each year that also appear among the USA Today bestsellers at some point during the year. About 40 of the 100 NYT notable works typically appear among the USA Today bestsellers.

²⁰See Doll (May 22, 2012).

²¹See Vinjamuri (August 15, 2012).

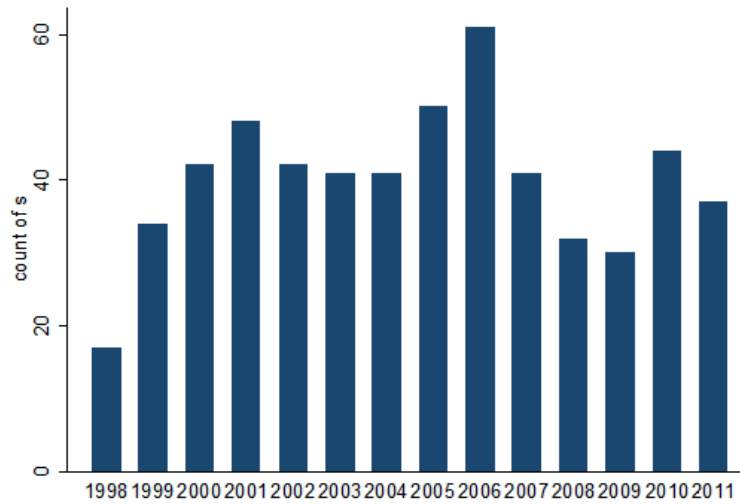


Figure 2.9: Number of NYT Notable Titles on USA Today List

Figure 2.10 approximates the USA Today works sales as $1/(\text{weekly rank})$. Using this approach, we can see that NYT notable works account for about 5 percent of the estimated sales of USA Today bestsellers. There appears to be a downward trend from a high of 13 percent in 2000 to 3 percent in 2011, but it is difficult to determine whether the trend is statistically meaningful.

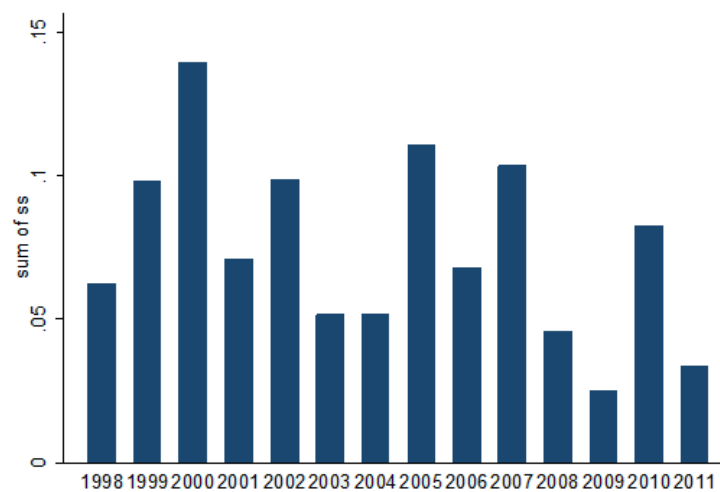


Figure 2.10: NYT Notable Titles' Share of USA Today Estimated Sales

2.8 Conclusion

New technology has reduced the costs of creating and distributing books and has given creators the opportunity to circumvent the traditional gatekeepers of publishing. Price reductions have spurred book consumption, with benefits to consumers on the order of \$3.5 billion per year. Self-published works, along with new institutions for product discovery, have expanded the choice set available to consumers. As of mid-2012 self published works account for about 10 percent of bestselling titles and sales.

Disintermediation presents a challenge to traditional publishers and retailers. Amazon has responded by becoming a major facilitator and retailer for self-published titles. Traditional publishers are also responding. Penguin (a division of Pearson publishing) purchased one of the largest self-publishing companies Author Solutions in July of 2012 for \$116 million²². Coming years promise to bring challenges to the traditional industry along with benefits to consumers.

²²See Sonne and Trachtenberg (July 19, 2012).

Chapter 3

References

- ADILOV, N., AND M. WALDMAN (2012): “Optimal copyright length and ex post investment: a Mickey Mouse approach,” Economic Inquiry.
- AUSTEN, B. (2011): “The End of Borders and the Future of Books,” Retrieved from Bloomberg Businessweek Magazine: <http://www.businessweek.com/magazine/the-end-of-borders-and-the-future-of-books-11102011.html#p2>.
- BERRY, S., AND J. WALDFOGEL (1996): “Free entry and social inefficiency in radio broadcasting,” Discussion paper, National Bureau of Economic Research.
- BERRY, S. T. (1994): “Estimating discrete-choice models of product differentiation,” The RAND Journal of Economics, pp. 242–262.
- BOSMAN, J. (2013): “Penguin and Random House Merge, Saying Change Will Come Slowly,” Retrieved from New York Times: http://www.nytimes.com/2013/07/02/business/media/merger-of-penguin-and-random-house-is-completed.html?_r=2&.
- BOWKER (2012): Self-Publishing in the United States, 2006-2011: Print vs. Ebook.
- CAVES, R. E. (2000): Creative industries: Contracts between art and commerce, no. 20. Harvard University Press.
- DIXIT, A. K., AND J. E. STIGLITZ (1977): “Monopolistic competition and optimum product diversity,” The American Economic Review, 67(3), 297–308.

- DOLL, J. (May 22, 2012): “The Alleged Sexiness of ’50 Shades of Grey’,” Retrieved from The Atlantic Wire: <http://www.theatlanticwire.com/entertainment/2012/05/alleged-sexiness-50-shades-grey/52667/>.
- G.F. (September 4, 2012): “The world’s biggest book club,” Retrieved from The Economist: <http://www.economist.com/blogs/babbage/2012/09/books-and-internet>.
- GRECO, A. N. (2004): The book publishing industry. Taylor & Francis.
- GREENFIELD, J. (2013): “E-Retailers Now Accounting for Nearly Half of Book Purchases by Volume,” Retrieved from Digital Book World: Digital Publishing News for the 21st Century: <http://www.digitalbookworld.com/2013/e-retailers-now-accounting-for-nearly-half-of-book-purchases-by-volume>.
- HAUSMAN, J. A., AND G. K. LEONARD (2002): “The competitive effects of a new product introduction: A case study,” The Journal of Industrial Economics, 50(3), 237–263.
- HEALD, P. (2007): “Property rights and the efficient exploitation of copyrighted works: an empirical analysis of public domain and copyrighted fiction best sellers,” UGA Legal Studies Research Paper, (07-003).
- IMBENS, G. W., AND T. LEMIEUX (2008): “Regression discontinuity designs: A guide to practice,” Journal of Econometrics, 142(2), 615–635.
- ISHII, J. (2005): “Compatibility, competition, and investment in network industries: ATM networks in the banking industry,” Unpublished working paper.
- KORDA, M. (2001): Making the List: A Cultural History of the American Bestseller, 1900-1999: as Seen Through the Annual Bestseller Lists of Publishers Weekly. Barnes & Noble Publishing.
- LI, X., M. MACGARVIE, AND P. MOSER (2012): “Dead Poets’ Property-The Copyright Act of 1814 and the Price of Books in the Romantic Period,” .
- LIEBOWITZ, S. J. (2006): “File Sharing: Creative Destruction or Just Plain Destruction?,” Journal of Law and Economics, 49(1), 1–28.

- MANKIW, N. G., AND M. D. WHINSTON (1986): “Free entry and social inefficiency,” The RAND Journal of Economics, pp. 48–58.
- MAZZEO, M. J. (2002): “Product choice and oligopoly market structure,” RAND Journal of Economics, pp. 221–242.
- MILLIOT, J. (2012): “Industry Sales Pegged At \$27.2 Billion,” Retrieved from Publishers Weekly: <http://www.publishersweekly.com/pw/by-topic/industry-news/financial-reporting/article/53112-industry-sales-pegged-at-27-2-billion.html>.
- OBERHOLZER-GEE, F., AND K. STRUMPF (2007): “The effect of file sharing on record sales: An empirical analysis,” Journal of political economy, 115(1), 1–42.
- PAKES, A., J. PORTER, K. HO, AND J. ISHII (2011): “Moment Inequalities and Their Application,” Unpublished Manuscript.
- RAINIE, H., K. ZICKUHR, K. PURCELL, M. MADDEN, AND J. BRENNER (2012): The rise of e-reading. Pew Research Center’s Internet & American Life Project.
- RAINIE, L. (2012): “Tablet and e-book reader ownership nearly double over the holiday gift-giving period,” Tablet.
- RICH, M. (2010): “Math of publishing meets the e-book,” New York Times.
- ROB, R., AND J. WALDFOGEL (2006): “Piracy on the high C’s: music downloading, sales displacement, and social welfare in a sample of college students,” Discussion paper, National Bureau of Economic Research.
- ROSEN, H. S., AND K. A. SMALL (1981): “Applied welfare economics with discrete choice models,” .
- SONNE, P., AND J. TRACHTENBERG (July 19, 2012): “Penguin Group Dives into Self-Publishing,” Retrieved from Wall Street Journal: <http://online.wsj.com/article/SB10000872396390444464304577537092288601370.html>.
- TERVIÖ, M. (2009): “Superstars and mediocrities: Market failure in the discovery of talent,” The Review of Economic Studies, 76(2), 829–850.

- TURAN, K. (January 17, 2007): “What Dark Horse with next ‘Sunshine’?,” Retrieved from Chicago Tribune: <http://www.chicagotribune.com/topic/zap-et-sundance17jan17,0,5602793.story>.
- VINJAMURI, D. (August 15, 2012): “Publishing is Broken, We’re Drowning in Indie Books - And That’s a Good Thing,” Retrieved from Forbes Magazine: <http://www.forbes.com/sites/davidvinjamuri/2012/08/15/publishing-is-broken-were-drowning-in-indie-books-and-thats-a-good-thing/>.
- WALDFOGEL, J. (forthcoming): “And the bands played on,” *The Economics of Digitization: An Agenda*. Shane Greenstein, Avi Goldfarb and Catherine Tucker, eds., National Bureau of Economic Research.
- WILLIAMS, H. L. (2010): “Intellectual property rights and innovation: Evidence from the human genome,” Discussion paper, National Bureau of Economic Research.

Appendix A

Appendix to chapter 1

A.1 Obtaining Demand Data

There is little literature on the book publishing industry, mostly because it is difficult to obtain reliable data on demand. Several papers aim at solving this problem. Most notably, Chevalier & Goolsbee (2003) uses publically available data on sales ranks at amazon.com and barnesandnoble.com to derive quantity proxies at these online book stores. They use these results to determine relative price sensitivities. The set of books that I consider allows me to avoid the estimation of an elasticity parameter because I consider the long tail of demand. Almost all of the titles under consideration consistently have sales ranks below 100,000. These books are not sold very often - most likely not more than once an hour. I use a website called www.novelrank.com to track hourly sales ranks of a book. Each improvement in ranking (over the previous hour) can then be treated as one book sold. Counting the instances of sales rank improvements over a month gives me a reasonably accurate proxy for the quantity of books sold. Consider, for example, the sales ranks of a mass market paperback version of *Arrowsmith* by Sinclair Lewis (1925). From April 2011 to March 2012, this book has been sold 13 times, an average of once per month. The individual sales are easily distinguishable from each other, as figure A.1 can attest.

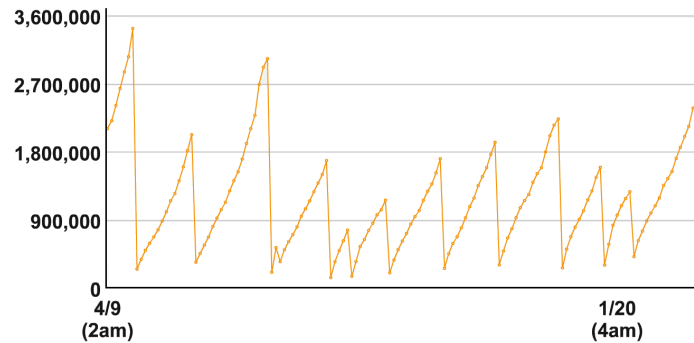


Figure A.1: Sales Ranks, Arrowsmith (Mass Market Paperback)

Some books are slightly more difficult to analyse as they face a higher demand than this particular example. Since I have hourly ranking data, the sales can still be counted by virtue of counting individual spikes in the rankings if the ISBN hasn't been sold more than once an hour.

The demand for e-books is partly taken from the same source, but these data do not seem as reliable. In addition to Amazon's sales rankings and its documentation through novelrank.com, I collect monthly download counts of titles that are available through the Project Gutenberg website. Although all titles that are available through this website are in the public domain, it is the closest approximation on quantities I could obtain given the restricted nature of data availability on e-book demand.

A.2 Discontinuity in Availability and Prices

In section 2.2 I illustrate the effect of the copyright extension in a regression discontinuity design (RDD) setup. While a copyright does not have a significant effect on the prices of physical editions, it does have an effect on the number of editions that a title is offered in. This effect is universal across formats and controlling for a title's "creative quality". Considering only the most popular works, there is a drop of around 40 editions per title at 1923, while for the least popular work there is a drop of almost 20 editions per title. This is illustrated in figures A.2 and A.2. Similar patterns hold true across formats.



Figure A.2: RDD: High-Popularity Works

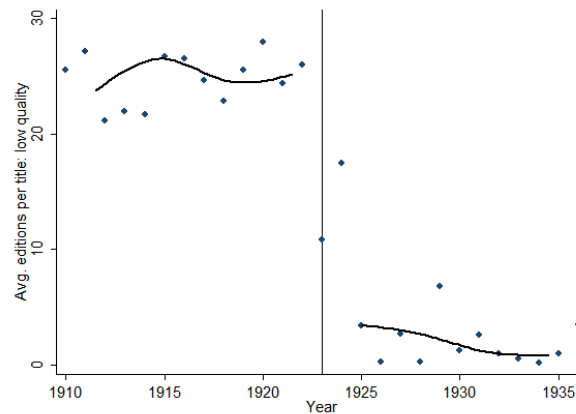


Figure A.3: RDD: Low-Popularity Works

A.3 Market Shares and Elasticities

While formulas for market shares and elasticities of logit and one-level nested logit models are widely known in demand estimation (see, for example, Berry (1994)), the extension to two levels of nests is tedious. My demand model includes one level of nests for titles, and another level of nests for each format within a title. Verboven (1996)

shows that market shares can be written as:

$$\begin{aligned} s_j &= s_{j|wk} \cdot s_{k|w} \cdot s_w \\ &= \frac{\exp\{\delta_j/(1-\sigma_1)\}}{\exp\{I_{wk}/(1-\sigma_1)\}} \cdot \frac{\exp\{I_{wk}/(1-\sigma_2)\}}{\exp\{I_w/(1-\sigma_2)\}} \cdot \frac{\exp\{I_w\}}{1 + \exp\{I_w\}} \end{aligned}$$

where $I_{wk} = (1-\sigma_1) \ln \left(\sum_{l \in J_{wk}} \exp \left\{ \frac{\delta_l}{1-\sigma_1} \right\} \right)$ and $I_w = (1-\sigma_2) \ln \left(\sum_{k \in J_w} \exp \left\{ \frac{I_{wk}}{1-\sigma_2} \right\} \right)$.

Define $D_k = \sum_{l \in J_{wk}} \exp \left\{ \frac{\delta_l}{1-\sigma_1} \right\}$. Then $I_{wk} = (1-\sigma_1) \ln(D_k)$, and I_w can be rewritten as:

$$I_w = (1-\sigma_2) \ln \left(\sum_{k \in J_w} D_k^{\frac{1-\sigma_1}{1-\sigma_2}} \right)$$

Letting $D_w = \sum_{k \in J_w} D_k^{\frac{1-\sigma_1}{1-\sigma_2}}$, we obtain $I_w = (1-\sigma_2) \ln(D_w)$. Then,

$$\begin{aligned} s_{j|wk} &= \frac{\exp \left\{ \frac{\delta_j}{1-\sigma_1} \right\}}{D_k} \\ s_{k|w} &= \frac{D_k^{\frac{1-\sigma_1}{1-\sigma_2}}}{D_w} \\ s_w &= \frac{D_w^{1-\sigma_2}}{1 + \sum_{w' \in W} (D_w^{1-\sigma_2})} \end{aligned}$$

This obtains the market shares indicated above:

$$s_j = \frac{\exp\{\delta_j/(1-\sigma_1)\}}{D_k^{(\sigma_1-\sigma_2)/(1-\sigma_2)} D_w^{\sigma_2} (1 + \sum_{w' \in W} D_w^{1-\sigma_2})}$$

Elasticities are then obtained in the usual fashion as $\epsilon_j = \frac{\partial s_j}{\partial p_j} \frac{p_j}{s_j}$. While the derivation is tedious, the formula for calculating elasticities is neat:

$$\epsilon_j = \frac{\alpha p_j}{1-\sigma_1} \left[1 - \frac{\sigma_1 - \sigma_2}{1-\sigma_2} s_{j|wk} - \sigma_2 \left(\frac{1-\sigma_1}{1-\sigma_2} \right) s_{j|w} - (1-\sigma_1) s_j \right]$$

Note that in the logit case where $\sigma_1 = \sigma_2 = 0$ we get the expected logit price elasticity formula $\epsilon_j = \alpha p_j (1 - s_j)$.

A.4 Imposing Marginal Costs by Format and Copyright Status

The dataset includes several editions that are sold at a steep discount (often less than \$1). The nature of the dataset and of a logit demand estimation cause price elasticities of several books to be estimated between 0 and -1. This gives a negative cost estimate under profit maximization.

There are several ways of dealing with these observations. In the paper I drop observations that give unreasonable marginal cost estimates in order to keep the marginal costs flexible at the edition level. In this section I impose that marginal costs are fixed at their mean level by format and copyright status, so that we have a marginal cost c_k^s as opposed to c_j^s . While the marginal costs are thus kept lower than the mean marginal costs in the paper (where I drop negative marginal costs), the fixed cost estimates become higher. In some cases the estimated lower bounds for fixed costs are higher than the upper bounds, which is an artifact of the moment inequalities approach. Upper bounds are biased downward and lower bounds are biased upward in small samples. Both table A.1 and figure A.4 illustrate the trends.

Table A.1: Marginal and Fixed Costs

Mean Marginal costs per book sold					
	Public Domain		IP Protection		Data (BookStats)
	in \$	Markup	in \$	Markup	Markup
Hardcover	15.34	0.232	17.66	0.314	0.376
Paperback	9.89	0.253	10.64	0.416	0.532
E-Book	1.21	0.756	7.14	0.285	0.227

Average Fixed cost bounds per edition published (in \$)					
	Public Domain		IP Protection		
	Lower	Upper	Lower	Upper	
Hardcover	104.98	116.22	1304.17	1137.23	
Paperback	199.63	217.83	850.66	1673.65	
E-Book	148.05	163.64	3322.70	3420.54	

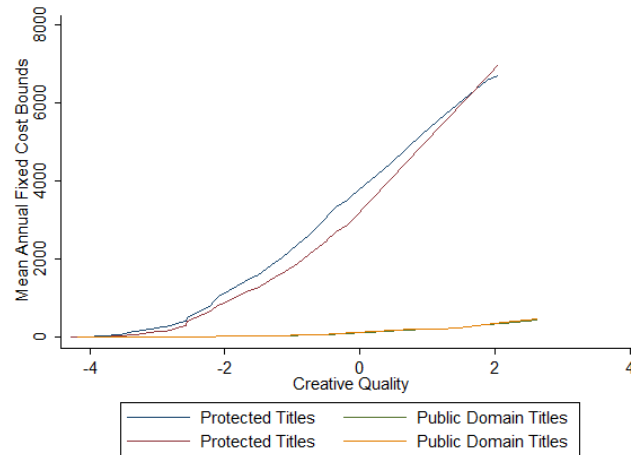


Figure A.4: Fixed Cost Bounds by Creative Quality

This approach will lead to larger losses by publishers if a title moves into the public domain. The overall effect of moving a title into the public domain is still overwhelmingly positive.

A.5 A Move into the Public Domain - Selected Titles

In this section I present how a move of a few representative titles of different quality levels affects total surplus. Table A.2 shows welfare effects for selected low-quality, medium quality and high-quality titles. The effect of a move into the public domain differs on a title-by-title basis. Welfare effects depend on the number of editions available, the title's quality and the editions' characteristics. Other things equal, if a large number of editions is already available, a move into the public domain is more likely to have a negative effect. Similarly, a less popular work is more likely to cause excess entry if it moves into the public domain. That is, a title with a more exhausted market will be better off under copyright protection than in the public domain.

Table A.2: Low-Quality Titles - All figures in \$

Title	$\Delta(\text{CS})$	$\Delta(\text{PS})$ - existing	Total π - new	$\Delta(\text{welfare})$
One Increasing Purpose	167.90	0	51.74	219.64
Old Pybus	320.34	25.09	3.88	449.31
Sparkenbroke	180.00	0	54.62	234.62
A Good Woman	177.03	0	54.26	230.30
Maid in Waiting	311.09	-124.16	10.88	197.80
Come and Get It	593.79	-10.90	27.85	610.73
The Woman of Andros	518.98	-28.53	21.83	512.31
Doomsday	298.19	13.59	73.28	385.06
The Blue Window	1018.11	5.17	78.32	1110.27
Mary Peters	1007.46	-79.07	23.99	952.37

Medium-Quality Titles - All figures in \$

Title	$\Delta(\text{CS})$	$\Delta(\text{PS})$ - existing	Total π - new	$\Delta(\text{welfare})$
The Constant Nymph	4034.92	-275.09	28.62	3788.46
The Plutocrat	1914.91	-60.64	65.82	1920.08
Swan Song	4368.02	-1.55	41.48	4407.96
Dodsworth	5241.51	-627.06	9.36	4623.81
The Greene Murder Case	3813.48	-96.32	30.93	3748.10
The Green Hat	4556.79	-203.43	86.38	4439.75
Show Boat	1809.38	-2404.38	45.65	-549.34
Shadows on the Rock	2198.58	-960.34	78.00	1316.24
White Banners	3388.37	320.17	48.78	3116.98
A White Bird Flying	2850.38	-1018.13	115.69	1947.93

High-Quality Titles - All figures in \$

Title	$\Delta(\text{CS})$	$\Delta(\text{PS})$ - existing	Total π - new	$\Delta(\text{welfare})$
Arrowsmith	14847.23	-6894.52	60.29	8013.00
Of Time and the River	14555.96	-1536.17	70.71	13190.51
The Keeper of the Bees	20910.97	-2427.63	107.75	18591.09
It Can't Happen Here	12741.91	-4024.65	152.02	8869.28
Little Man, What Now?	14235.43	-3228.18	396.08	11403.33
Sons	13680.63	-4410.67	141.32	9411.28
Lost Horizon	20884.93	-4313.72	134.75	16705.96
Years of Grace	23458.94	-2593.45	329.42	21194.91
Gone With the Wind	45428.53	-26470.24	3164.88	22123.17
The Good Earth	51433.81	-3103.11	266.55	48597.25

Appendix B

Appendix to chapter 2

B.1 Titles that were originally self-published

Determining which books are self-published requires some detective work. The USA Today bestseller lists include each title's publisher in addition to their rank, longevity and most popular format. Some of these titles are easy to categorize. We deem a work self-published if the listed publisher contains the word "self" (e.g. "self-published via Amazon"). We also include works published by the major self-publishing services listed in Bowker (2012). These include Smashwords, Lulu Enterprises, and various divisions of Author Solutions (Xlibris, Authorhouse, IUniverse, and Trafford). These services collectively account for about three quarters of self-published electronic books. We also found works that had originally been self published using online sources. Table B.1 lists all of the works that were in the USA Today top 150 weekly lists in 2012 and that we identify as originally self-published.

Table B.1: Originally Self-Published Works and their Sales, 2010-2012 ytd

Title	Author
Accidentally Married to ... a Vampire?	Mimi Jean Pamfiloff
All In	Raine Miller
All She Ever Wanted	Barbara Freethy
Angelfall (Penryn & the End of Days, Book 1)	Susan Ee
Anything He Wants	Sara Fawkes
Anything He Wants 2: All's Fair	Sara Fawkes
Anything He Wants 4: Collateral Damage	Sara Fawkes
Bared to You: A Crossfire Novel	Sylvia Day
Beautiful Disaster	Jamie McGuire
Because of Low	Abbi Glines
Bedding the Billionaire	Ruth Cardello
Ceaseless	Abbi Glines
Chasing Rainbows	Kathleen Long
Checkmate	R.L. Mathewson
Clash	Nicole Williams
Crash	Nicole Williams
Devoured	Emily Snow
Down to You	M. Leighton
Easy	Tammara Webber
Fade Into Me	Kate Dawes
Fade Into You	Kate Dawes
Fifty Shades Darker	E L James
Fifty Shades Freed	E L James
Fifty Shades of Grey	E L James
Flat-Out Love	Jessica Park
For His Forever	Kelly Favor
For His Honor	Kelly Favor
Gamble on Engagement	Rachel Astor
Golden Lies	Barbara Freethy
Hard Mated	Jennifer Ashley
High Heels Mysteries Boxed Set	Gemma Halliday
His Every Choice	Kelly Favor
His Every Defense	Kelly Favor
His Every Desire	Kelly Favor
His Every Move	Kelly Favor
His Every Touch	Kelly Favor

In Leah's Wake	Terri Giuliano Long
Just the Way You Are	Barbara Freethy
Last Breath	Michael Prescott
Lord of Vengeance	Lara Adrian
Losing It	Cora Carmack
Love Left Behind	S.H. Kolee
Mortal Faults	Michael Prescott
My Favorite Mistake	Chelsea M. Cameron
My Mother Was Nuts	Penny Marshall
Naked	Raine Miller
Nobody's Perfect	Kallypso Masters
On Dublin Street	Samantha Young
On the Island	Tracey Garvis-Graves
Perfection: A Neighbor From Hell Novel	R.L. Mathewson
Playing for Keeps	R.L. Mathewson
Point of Retreat	Colleen Hoover
Reason to Breathe	Rebecca Donovan
Rising Storm	Kathleen Brooks
Samantha Moon	J.R. Rain
Sarah's Surrender	Lynda Chance
Slammed	Colleen Hoover
Softly at Sunrise	Maya Banks
Taken	Kelli Maine
Taking Chances	Molly McAdams
The Bride Wore Chocolate	Shirley Jump
The Edge of Never	J.A. Redmerski
The Hangman's Daughter	Oliver Potzsch
The Look of Love: The Sullivans, Book 1	Bella Andre
The Mill River Recluse	Darcie Chan
The Perfect Game	J. Sterling
The Secret of Ella and Micha	Jessica Sorensen
The Sweetest Thing	Barbara Freethy
The Wild Ones	M. Leighton
Training Tessa	Lyla Sinclair
Vampire Dawn	J.R. Rain
Weekends Required	Sydney Landon
While It Lasts	Abbi Glines
White Trash Beautiful	Teresa Mummert
Why Me?	Sarah Burleton
Wife by Wednesday	Catherine Bybee
Wild Montana Sky	Debra Holland
Wish List	Sylvia Day

B.2 Alternative specifications for demand estimation

The demand estimation and consumer surplus results presented in chapter 2 assume that the mean price elasticity of a bestselling edition is -5. It is however quite possible that the true price elasticity is different. Under profit maximization e-books (with almost no marginal costs) are likely to have an inverse price elasticity of $-\frac{p-mc}{p} = -1$, thus giving us a price elasticity of -1. At the same time physical books are said to have a markup of around 20%. This implies a mean price elasticity of -5 as assumed in the chapter. The true price elasticity will most likely be in between these two values. We present demand estimates and consumer surplus results when the mean price elasticity is -1 and -2 in tables B.2 and B.4, and B.3 and B.5, respectively.

Table B.2: Demand estimates: Mean price elasticity is set to -1

σ	0	0.25	0.5	0.75	0.9
E-book	0.0235 (0.0433)	0.0176 (0.0324)	0.0117 (0.0216)	0.00587 (0.0108)	0.00235 (0.0433)
Price	-0.121	-0.0910	-0.0606	-0.0303	-0.0121
Title FE	Yes	Yes	Yes	Yes	Yes

Std. errors in parentheses;
Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table B.3: Demand estimates: Mean price elasticity is set to -2

σ	0	0.25	0.5	0.75	0.9
E-book	-0.295 0.0612	-0.221*** (0.0459)	-0.147*** (0.0306)	-0.0737*** (0.0153)	-0.0295*** (0.00612)
Price	-0.243	-0.182	-0.121	-0.0606	-0.0243
Title FE	Yes	Yes	Yes	Yes	Yes

Std. errors in parentheses;
Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

A larger price elasticity corresponds with a larger monetary disutility for e-books. This is driven by the observed relative popularity of physical and electronic books. If a consumer with a high price elasticity buys e-books as often as physical books although

e-books have a lower price, then they must prefer physical books a lot. If on the other hand a consumer with a low price elasticity buys as many e-books as physical books, the difference in utility from the two formats need not be as large.

Consequently, a lower price elasticity and a lower disutility for e-books will lead to larger effects on consumer surplus: if people like e-books as much as they like physical books, then the introduction of e-books (at lower prices) increases consumer surplus more than would be the case if people don't like e-books as much. Notice also that with a low price elasticity consumers benefit more from the additional variety through self-published books than they do from the lower prices of e-books.

Table B.4: Compensating variation, million \$, total in 2012 (compared to no ebooks or self-publishing), price elasticity = -1

σ	0.0	0.25	0.5	0.75	0.9
p_e high, no self	1,351.947	2,781.593	5,172.463	8,275.055	10,110.906
p_e high, p_{self} high	2,030.146	4,042.885	7,303.654	11,443.289	13,875.594
p_e low, no self	1,891.446	3,790.611	6,886.023	10,831.883	13,153.313
p_e low, p_{self} high	2,614.455	5,077.629	8,976.213	13,849.359	16,699.094
p_e low, p_{self} low	2,699.027	5,223.931	9,207.779	14,177.375	17,081.719

Table B.5: Compensating variation, million \$, total in 2012 (compared to no e-books or self-publishing)

σ	0.0	0.25	0.5	0.75	0.9
p_e high, no self	522,428	1,092.566	2,061.902	3,334.684	4,090.766
p_e high, p_{self} high	806.757	1,638.285	3,011.005	4,776.316	5,817.797
p_e low, no self	945.723	1,895.306	3,443.011	5,415.941	6,576.656
p_e low, p_{self} high	1,274.882	2,482.580	4,398.699	6,797.629	8,201.141
p_e low, p_{self} low	1,349.514	2,611.966	4,603.890	7,088.688	8,540.859