The Effects of Intellectual Property on the Market for Existing Creative Works

Imke Reimers[*]

University of Minnesota

January 22, 2013

Abstract

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. At the same time, many copyright-protected books, especially those of low value, have gone out of print. Under copyright, these "orphan" works are effectively unavailable as new products and putting them 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 from fiction bestsellers by decreasing variety, causing a decrease in consumer surplus that outweighs the increase in profits. Holding fixed the costs of entry a move into the public domain increases total surplus from most titles, indicating insufficient entry under copyright protection.

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 Act[1] extended the copyright term from 75 years (or "life of author + 50 years" for works that were created after 1978) to 95 years (or "life of author + 70 years"). As a result, there is a cliff at 1923 where copyright drops off to zero: works that were created after 1923 remain on copyright 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 to intellectual property - is understood to provide incentives for creation[2], these monopolies have potentially harmful impacts on consumers. The distribution of protected works is controlled by monopolists (the copyright holders). 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. 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 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 online[3].

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

---

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

[2]see, for example, Williams (2010)

[3]Googles 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.

quantification of both the cost to producers and the benefits to consumers.

In addition to the standard literature on the theory of copyright protection, a growing literature examines the impact of copyright protection via comparisons of works that are similar except in they copyright status by comparing works published before and after "cliff" date. Heald (2008) documents that coming off copyright substantially raises the probability that a title will be in print. My descriptive results match his findings. Moser (2012) does something similar with respect to prices and the Copyright Act of 1814. She finds that prices are higher if a title is on copyright. The descriptive results are informative and suggestive of welfare impacts, but a welfare assessment requires measurement of consumer and producer surplus under existing and counterfactual regimes, an exercise wich requires a structural model.

This paper thus adds several dimensions to the existing literature. First, it accounts for the differences in market structure and the possibility of excess entry. Second, it estimates 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, carries 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 is negligible whether they are protected by copyright or not. For titles whose quality lies in between these two types, their copyright status will affect their availability and consumption. Third, the paper quantifies the effect on the publishing industry by simulating a world when individual protected titles move into the public domain.

I use a dataset that utilizes the discontinuity at the cliff date by looking at fiction bestseller books that were written around 1923. It contains a title's in-print status as well as sales data 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, the creative work, 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 & Waldfogel (1999) as well as Mazzeo (2002) and Ishii (2005) by introducing product differentiation into a system that includes a structural profit function and allows me to estimate entry decisions. I estimate the exogenous

creative quality of a title and use this together with the number of existing editions as a determinant of each title-format combination's cost structure with a moment inequalities approach. I use this framework to determine the relative effects of free entry (the public domain) on consumers and publishers. I find that a move of a title into the public domain increases total surplus by increasing consumer surplus by a larger amount than it dissipates producer surplus. When decomposing the effect into a decrease in costs and a switch to a free-entry regime, I find that the Mankiw & Whinston (1986) problem of excess entry does not exist among most titles although it does hold for some titles.

More specifically, I find that a copyright decreases consumer surplus from a title by between $200 and $50,000 annually. Around one third of these changes 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 this number 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 additional entry into a title. These 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 entering an edition, potentially increasing profits. But on the other hand a move into the public domain implies a move into a free-entry system, dissipating 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, 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 will 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 will have to distinguish between the two types of titles. My results show that the copyright extension decreased total surplus from the publishing industry regardless of the composition of works.

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.

2 Data and Industry Background

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 1936[4]. 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 - 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 editions[5]. 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 title-format combinations[6]. 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

---

[4]These can be found on several websites, but Michael Korda's Book, *Making the List: A Cultural History of the American Bestseller 1900-1999*, was used as a reliable reference

[5]Throughout, I will use the terms *ISBN* and *edition* interchangeably.

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

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. While Amazon is the biggest internet retailer of physical books, Project Gutenberg is a reasonably well-known website that offers e-book versions of public domain works for downloads free of charge [7].

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, price, format and some additional characteristics 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 presents summary statistics of several measures of availability 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: Composition of Editions

Variable	Mean
Hardcover	0.3670
Paperback	0.4886
E-Book	0.1444
New	0.5403
Used	0.7978

E-books can almost exclusively be found among public domain titles. It is also easier to find each of the books in the dataset in used condition than it is to obtain them new. 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 only in one of the two conditions.

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 book[8]. I also obtain the monthly number of books that are downloaded (if available) at Project Gutenberg as listed on

---

[7]comparable to Google Books

[8]see appendix A for a more detailed description of the methodology

each title's website. The monthly download count will allow me to obtain a distribution of title qualities over the industry as well. Moreover, I obtain annual library lending figures 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 readily available in the United States. However, the number of titles that have been checked out through the library system is very small, so that I chose a bigger time frame here 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. Table 2 shows quantities demanded through each of the three channels, as well as certain 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 is at -9.68%.

Table 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 $)	21.896	11.101	5.49	98.00
Bowker Price (in $)	33.279	35.021	1.14	100.00

In the main demand estimation, I use quantity demanded in the US market as the dependent variable. That is, a sale through Amazon is equivalent to a download through Gutenberg. The BookStats wholesale prices give me a first estimate of marginal costs of a title that is protected by copyright.

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

This section shows the significance and size of the effect of a copyright on the number of editions per title that are in print (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). The setup lends itself to a regression discontinuity design (RDD) which displays the market effects of the copyright term extension. Any titles that have been published after 1923 have been "treated" by the copyright extension and hence their variety is determined by a copyright holder, while titles from before 1923 are not treated and individual editions face competition. In this section, I tighten the bound around 1923 further 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 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. 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). Thus, the equation

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

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 for the Novel.

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 shows this trend.

The effect can be broken into the different items of a title's product line. 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. The substitution patterns
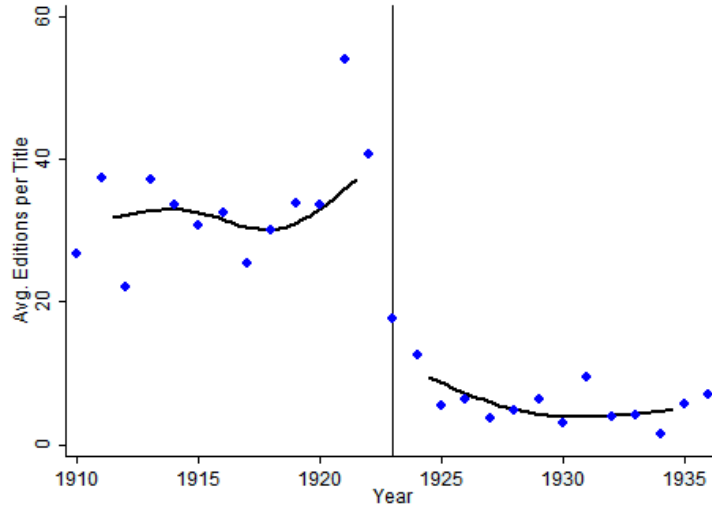
Figure 1: Number of ISBNs per Title: Copyright Effect = -28.053 versions

between formats play an important role in determining the exact effect on each of the formats. This - and the effect of a title's creative quality - will be taken into account and analyzed in the full model. Table 3 shows results of a regression of each format on quality and availability through different channels.

Treating the quantity demanded through British Libraries (PLR) as an indicator of quality, its positive effect is expected, although the size of the effect is small. A formal model of supply will capture the interaction between the formats and quantify the costs associated with obtaining the right to publish a book in a monopoly as opposed to a competitive market structure.

An RDD analysis of prices shows significant differences based on a title's copyright only among e-books. 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 when controlling for an edition's quality[9]. This indicates that the price setting decision is likely to be made in the same fashion across copyright regimes. Table 4 shows regressions of price on an edition's format and copyright status. The copyright effect is significant only for e-books.

This pattern is also illustrated in figure 2, which shows the (lack of) discontinuity in prices of hardcover editions around the cliff date of 1923. 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

---

[9]Note that this may be an artifact of my dataset.

Table 3: Regression Discontinuity - In-print Versions per Title

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

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

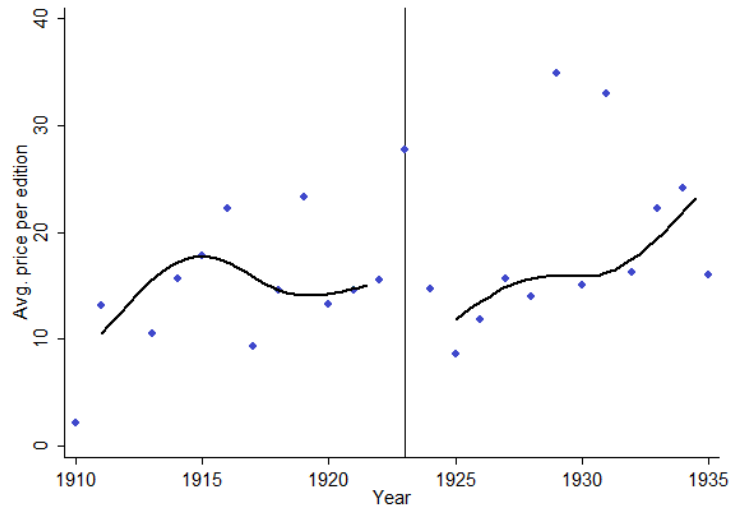focus mainly on the margin of availability.

Figure 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 difficult to determine how much of that difference is accounted for by the increased ease of downloading a book through (e.g.) Gutenberg. My demand model aims at identifying the

10

Table 4: Regression Discontinuity - Price per Edition

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

titles' creative qualities.

3 Model

The model has a consumer side and an entry component. A nested logit demand model obtains creative qualities, price elasticities and substitution patterns. The copyright holder and the edition managers[10] make their 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 regime. If the title is protected by copyright, the entry decision is made by the copyright holder who tries to maximize the profits from all editions of that title. After the entry decisions have been made, the edition managers compete in prices to maximize their own profits. The supply model allows me to recover marginal costs by format and copyright status, and

---

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

to estimate fixed costs by title, format and copyright regime.

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. 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 (any books that are not included in this data set, for instance non-fiction titles or more recent fiction titles, in addition to not reading a book at all) are likely to differ from those between editions within each title. In the demand estimation, I therefore allow for correlations of consumer tastes to differ on two levels: the correlations of tastes for editions within a title-format combination can differ from the correlations of tastes across formats within each title, and these correlations are allowed to differ from the correlations of tastes for editions across titles. As editions within each title-format combination differ in their publication date, condition and price, the editions are identified in each title-format combination. Figure 3 illustrates the proposed nesting structure:
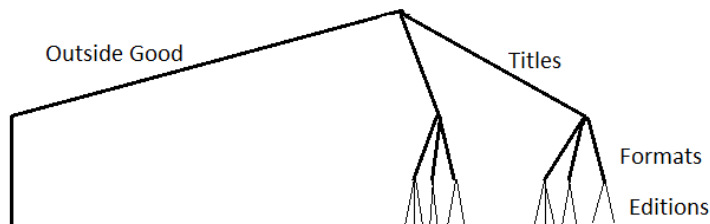

Figure 3: Nested Logit Structure

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

$$\ln s_j - \ln s_0 = \delta_j + \xi_j + \sigma_1 \ln s_{j|wk} + \sigma_2 \ln s_{k|w} + \epsilon_{ij},$$

where

$$\delta_j = \beta_0 + \beta_{age} age_j + \beta_{new} new_j - \alpha p_j + \phi_w + \xi_j.$$

In this model, p_j is the edition's price as indicated by Amazon (it is set to zero if the edition is offered through Gutenberg), and ϕ_w is 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 included in the ξ_{jt} term). The $\ln s_{j|wk}$ 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 $\ln s_{k|w}$ 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 of the old fiction titles 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. Market shares of edition j in the nested logit model are given by

$$s_j = s_{j|wk} \cdot s_{k|w} \cdot s_w.$$

This turns into a tractable equation for an edition's market share as a function of its quality and

the quality of its competitors[11]:

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

where $D_k = \sum_{j\in J_{wk}} \exp\{\delta_j/(1-\sigma_1)\}$ and $D_w = \sum_{k\in J_w} \left(D_k^{(1-\sigma_1)/(1-\sigma_2)}\right)$. An additional edition with mean quality δ_{wk} will affect each existing edition's market share. Since, generally (and in each of my specifications), the σ_2 term (the coefficient on the share of the broader nest) will be 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 extent of the changes in shares depends on the σ terms.

3.2 Supply and Equilibrium

While I do estimate several nested logit structures for robustness checks, I continue with the specification introduced in section 3.1. in the estimation of 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 affected by the work's copyright status $r, r \in \{O, I\}$ (where O indicates that the title is in the public domain and I shows that the title is under IP protection) on two dimensions. First, a potential entrant in the public domain decides 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, on the other hand, an edition manager will have to *ask* the copyright holder to be allowed to enter. The copyright holder will let edition managers enter an edition until her expected profits are maximized. That is, the copyright holder is assumed to receive a (fixed) proportion of the total profits from the editions of that title. A move of a title into the public domain thus implies that the title moves from a profit-maximization setting 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 is expected to have higher fixed costs as he has to obtain a

---

[11]See appendix A.2 for a detailed derivation.

license to publish an edition of that title in addition to the costs of obtaining the text and setting up machines to print that are incurred in both regimes. A protected title is also likely to have a higher marginal cost of selling a book - 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 these differences in market structure and costs.

Let c_k^r be the marginal cost of selling a book of format k of a work in regime r, 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

$$c_k^r = c_k^{r,pr} + c_k^{r,cr}.$$

Royalties are paid only for books of protected titles[12]. Further, let F_{wk}^r be the fixed cost of publishing an edition of format k of work w in regime r. Marginal costs are common across titles within a format and regime, while fixed costs are allowed to vary by format *and* title 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, stage 2 describes the pricing decisions by the edition managers.

Consider first stage 2 in this entry game. Define $X_k = (X_{1,k}, X_{2,k}, \ldots, X_{n_k,k})$ 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 $\xi_k = (\xi_{1,k}, \xi_{2,k}, \ldots, \xi_{n_k,k})$ as the vector of unobserved edition characteristics. 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 $\boldsymbol{\xi}$ from the demand estimation, 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 prices[13]. Then,

---

[12]If editions have the same physical costs across copyright regimes, identification of royalties is obtained directly.

[13]This assumption is consistent with Mankiw & 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 2.3.

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 qualities 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

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

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

$$
\begin{aligned}
\Pi_k^r(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n) &= E\left[\pi_{jk}^r(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi})\right] \\
&= \int_\xi (p_j - c_k^r)q_j(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi})dF^k(\xi),
\end{aligned}
\tag{2}
$$

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 title only through fixed and variable costs.

3.2.1 Public Domain, $r = O$

When a title is in the public domain, free entry conditions will 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. 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 can be bound by these Nash equilibrium observations for each w:

$$
\begin{aligned}
\Pi_{w,H}^O(n_H^*, n_P^*, n_E^*) &\geq F_{w,H}^O \\
\Pi_{w,H}^O(n_H^* + 1, n_P^*, n_E^*) &\leq F_{w,H}^O,
\end{aligned}
\tag{3}
$$

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:

$$\Pi_{w,H}^O(n_H^*, n_P^*, n_E^*) - F_{w,H}^O \geq \Pi_{w,P}^O(n_H^* - 1, n_P^* + 1, n_E^*) - F_{w,P}^O \tag{4}$$

$$\Pi_{w,H}^O(n_H^*, n_P^*, n_E^*) - F_{w,H}^O \geq \Pi_{w,E}^O(n_H^* - 1, n_P^*, n_E^* + 1) - F_{w,E}^O.$$

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

3.2.2 Copyright, $r = I$

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. She therefore chooses the combination of formats that maximizes the title's joint profits, given Bertrand competition after the entry decision has been made. These are given by:

$$\Pi_w^I(\mathbf{p}, \boldsymbol{\xi}, n) = n_H \Pi_H^I(\mathbf{p}, \boldsymbol{\xi}, n) + n_P \Pi_P^I(\mathbf{p}, \boldsymbol{\xi}, n) + n_E \Pi_E^I(\mathbf{p}, \boldsymbol{\xi}, n) \tag{5}$$

If we observe the combination of editions $(n_H^{**}, n_P^{**}, n_E^{**})$, 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 us a different set of inequalities:

$$\sum_{j \in n_w} \pi_j(X_j; n_H^{**}, n_P^{**}, n_E^{**}) - \sum_{j \in n_w'} \pi_j(X_j; n_H^{**} - 1, n_P^{**}, n_E^{**}) \geq F_{wH}^I$$

$$\sum_{j \in n_w'} \pi_j(X_j; n_H^{**} + 1, n_P^{**}, n_E^{**}) - \sum_{j \in n_w} \pi_j(X_j; n_H^{**}, n_P^{**}, n_E^{**}) \leq F_{wH}^I$$

The equivalent inequalities must hold for paperback and e-book editions. Fixed costs for protected titles are then obtained by title-format combination after obtaining variable profits using the results

17

from the demand estimation. The additional restriction that $\pi_k(n_H, n_P, n_E) > 0$ for each $k \in \{H, P, E\}$ will be satisfied as long as the demand model is well-specified.

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. I then use the demand estimates in the supply model to recover marginal costs by format and copyright status. Lastly, I estimate bounds on fixed costs by title, format and copyright status using the equilibrium conditions of entry in the first stage[14].

4.1   Demand

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 $s_{k|w}$ and $s_{j|wk}$ with the level of competition (the number of editions of each format) within the title, and with the work's copyright status. While the number of editions of a specific title can reveal something about the overall market share of one single edition within this title, it provides more information on the edition's share *within* the title. The work's copyright status is a similar indicator of an edition's share within a title as titles that are protected by IP generally have fewer editions and thus each edition is likely to have a larger within-title share. It may add additional information on an edition's share within a title that may not be captured by the number of editions.

Table 5 shows the results of the nested logit specification described above. The coefficients on the nests' inside shares are significantly different from zero. The regular logit model is mis-specified, 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.

Price elasticities are rather low for the OLS and IV logit estimations. The nested logit

---

[14]I take prices as given in the data. This allows me to focus on the costs of production.

Table 5: Demand Estimation

	(1) OLS	(2) IV Logit	(3) Title - Format
New	1.069***	1.379***	0.276***
	(0.0339)	(0.0306)	(0.0248)
Price	-0.0503***	-0.124***	-0.0246***
	(0.00107)	(0.00158)	(0.00152)
Major	-0.171***	-0.131***	0.0576***
	(0.0313)	(0.0276)	(0.0208)
Age	-0.00254***	-0.00485***	-0.00352***
	(0.000625)	(0.000554)	(0.000432)
σ(title)			0.680***
			(0.00958)
σ(format)			0.952***
			(0.0122)
Constant	-18.75***	-18.23***	-14.48***
	(0.0676)	(0.0606)	(0.0609)
Mean Elasticity	-0.714	-1.761	-5.646
Observations	4169	4169	4169
R-squared	0.451	0.568	0.768

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

estimation obtains a more reasonable price elasticity of -5.7, which is comparable to the pseudo-price elasticity of -9.68 obtained in section 2.1. I focus my analysis on specification (3): 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 obtains reasonable coefficients and elasticities. The coefficients are significant, and their signs are sensible. Consumers prefer new editions over used ones, and they prefer editions from major publishers over editions from independent or smaller publishers, as major publishers have an established reputation of quality or reliability. Moreover, a consumer prefers a more recent edition over an older edition, holding all other factors equal.

The σ terms suggest 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

particularly close to 1 at 0.952, but there is a high correlation of tastes within titles as well, with a σ of 0.680. Thus, an additional edition of a title-format combination will not increase the share of this title much as any demand for this title can be mostly met by the existing editions.

Table 6 shows the ten titles with the highest and lowest estimated qualities. The estimated qualities mirror today's demand for the titles and the overall quality of each work reasonably well. Four out of the top ten titles have won Pulitzer Prizes for the Novel (11 of the titles in my list are Pulitzer Prize winners - with one exception, they are all found in the top 60 titles) 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 well[15]. On the other hand, none of the bottom ten (in fact, none of the bottom 50) titles could be found at the book store. Neither did any earn the author a Pulitzer Prize (although five of them were written by prize-winning authors). As the ξ_j terms will be centered around zero by construction, the creative quality terms ϕ_w indicate levels of production quality as well. The difference in mean values of ϕ_w is not significant between the two copyright regimes. If the underlying assumption that all titles are a priori similar holds, this implies that publishers do not extend a larger effort into producing "nice" editions of protected titles as compared to those of public domain titles. Thus, a move into the public domain is assumed to affect consumer surplus through prices and variety only.

The number of available editions is correlated with the estimated quality of a work, ϕ_w, as illustrated in figure 4.

The demand estimation shows 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 7 shows the estimated sales by edition and title as I add a hypothetical edition of each format. An additional hardcover edition thus decreases the share of the existing hardcover editions by an average of 15.6%, while the existing editions of different formats lose only a very small part of their shares. This is consistent with the high value of σ_1 from the demand estimation: most substitution happens within a title-format combination.

---

[15]I 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

Table 6: Quality Rankings - Top 10

	Title	Pulitzer Prize
1	The Age of Innocence (1921)	✓
2	Babbitt (1922)	
3	All Quiet on the Western Front (1929)	
4	The Good Earth (1931)	✓
5	Gone With the Wind (1936)	✓
6	Main Street (1921)	
7	Pollyanna (1913)	
8	Years of Grace (1930)	✓
9	Penrod (1914)	
10	The Enchanted April (1923)	

Bottom 10

	Title	Pulitzer Prize
1	Roper's Row (1929)	
2	One Increasing Purpose (1925)	
3	Old Pybus (1928)	
4	Lost Ecstasy (1927)	
5	Sparkenbroke (1936)	
6	A Good Woman (1927)	
7	Tomorrow Morning (1927)	
8	The Dim Lantern (1923)	
9	Within This Present (1934)	
10	Old Wine And New (1932)	

Table 7: Average Sales per Edition (in $), All Titles

	Base results	1 more H	1 more P	1 more E
Sales	109.79	106.79	102.79	108.01
% Change, Sales, H		-15.6%	-0.5%	-0.1%
% Change, Sales, P		-0.2%	-9.1%	-0.05%
% Change, Sales, E		-0.05%	-0.15%	-13.3%

The substitution patterns become clear when looking at particular titles. For example, there currently are 3 hardcover versions, 6 paperback versions, and 1 electronic version of *Gone With the Wind*. Consider the sales of this title on October of 2011. If one hardcover edition of this title enters the market, the mean edition sales of hardcover editions will decrease from 161
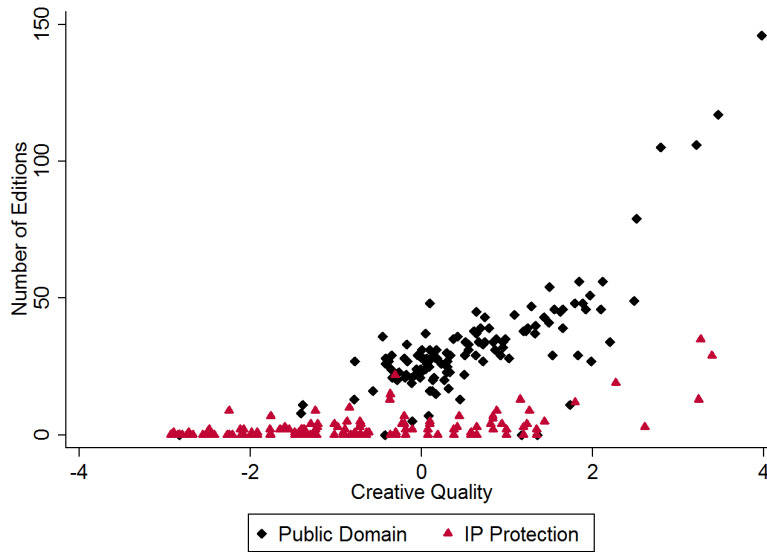
Figure 4: Editions per title as a function of quality

to 124 in that month, a relatively sizeable decrease. At the same time, mean sales of the existing paperback and electronic editions does not change as much. There is just a small decrease from 29.3 editions among paperback editions, and a decrease from 325 to 320 downloads of the e-book version. Overall, the additional hardback edition steals a relatively large market share from the existing hardcover versions while changing the market shares of the existing editions of different formats only a little and increasing the total quantity of books sold for *Gone With the Wind*. The market shares of versions of different titles remain rather unchanged.

Much of the change in 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 would if there are only a few editions. Figure 5 shows the relationship between changes in market shares and the number of available editions for the hardcover format. If one hardcover edition is added to the existing editions of a title, the market shares of the incumbent hardcover editions 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 are only a few editions.
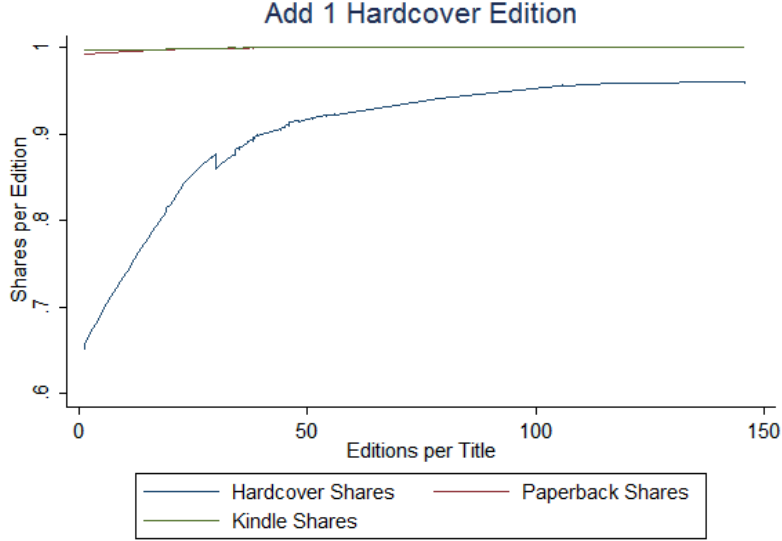
Figure 5: Shares as a % of Original Shares after adding an Edition

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[16], which is used to recover marginal costs. The first-order necessary condition of profit maximization in Bertrand competition is derived as:

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

The marginal costs are format specific and vary by a title's copyright status $r \in \{O, I\}$. An average of the markups by format and copyright status, $E_k^r \left[-\frac{1}{\epsilon_j} \right]$, is used together with the entry equilibrium inequalities to identify bounds of 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_{wk}^r will be incurred each year as well. Fixed costs include the costs of obtaining the license to publish an edition (this includes convincing the copyright holder to let you publish) and of setting up the machinery and printing a certain level of inventory, while marginal costs include the material for each (physical) book and royalties for protected versions.

---

[16]as derived in appendix A.2

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. That is,

$$F_{w,k}^O \le \overline{F}_{w,k}^O \equiv \Pi_k^O(\mathbf{p}, \mathbf{X}, \boldsymbol{\xi}, n),$$

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 obtain lower bounds of fixed costs in the public domain and fixed cost bounds of publishing protected titles[17].

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 are not observed although several books are sold through these channels as well. As a consequence, I observe only a fraction of variable profits for each edition as well. According to the Book Industry Study 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 2 of 2011, and for 27% in the second quarter of 2012[18]. 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

---

[17]My 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 $F_{wk}^O \le \overline{F}_{wk}^O = (p_j - c_k^r)q_j(\mathbf{p}, \mathbf{X}, \overline{\xi}_k)$

[18]see 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.

fixed cost bounds, I then use

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

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), I estimate fixed costs and evaluate the policy experiments for higher values of γ for robustness.

Table 8 shows the means of estimated bounds of these costs, conditional on copyright status and format and assuming that $\gamma = 0.2$.

Table 8: Marginal and Fixed Costs

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 Bound	Upper Bound	Lower Bound	Upper Bound	
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	

For each format, the average marginal and fixed costs are higher among protected titles than among public domain titles. Fixed costs are, on average, around ten times as high under copyright as they are in the public domain. There is a significant cost to obtaining the license to publish a protected title. Edition managers of all titles incur some fixed cost of obtaining a script of the title and setting up the machinery (or the plattform) 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 thousand dollars for most titles, but it is over $10,000 for some of the highest-quality titles as figure 6 shows.

While marginal costs are just slightly higher for protected titles, the profit margin is larger

for protected titles as average prices are slightly higher as well (although this difference is not significant[19]). 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. Note that the cost margins fit the benchmark values from the BookStats report reasonably well. Assuming that 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:

$$c_k^r = c_k^{r,pr} + c_k^{r,cr}$$

with $c_k^{r,pr}$ being the production cost of an edition, and $c_k^{r,cr}$ are the royalties paid to the 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 \$1.50 and \$2.50. Greco (2011) lists a standard royalty rate of 10% to 15% of the price[20]. At a sample average price of \$20 to \$30, the estimated marginal costs seem reasonable.

Unlike marginal costs, I allow fixed costs to vary by title. I obtain 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 and can charge a higher licensing fee to potential publishers if the edition manager is confident the edition will sell[21]. Figure 6 shows this relationship between creative qualities and fixed costs/licensing fees.

Profits differ by copyright status as well. On average, edition makers 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 copyright holder (the social planner) 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

---

[19]see section 2.3

[20]page 156

[21]Even if the fixed cost estimates in the copyright case seem high, the result of the paper will not change
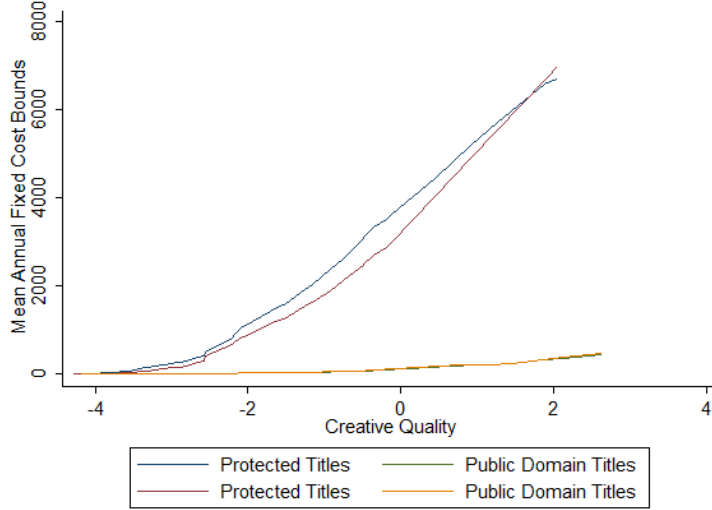
Figure 6: Fixed Cost Bounds by Creative Quality

even in the public domain. Table 9 summarizes profits for editions under their current coypright status, using the midpoints of fixed cost bounds as a point estimate for costs.

Table 9: Profits per Edition (in $ per year)

	Public Domain		IP Protection	
	Mean	Max	Mean	Max
Hardcover	14.72	96.60	1222.59	28532.95
Paperback	9.78	152.83	1780.19	12186.94
E-Book	15.41	69.01	2759.75	15098.00

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) have the largest profits. Profits of protected low-quality title editions are lower than those of public domain titles of a similar quality. But protected high-quality titles 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.

A change in a title's copyright status affects profits as well as the level of consumer surplus

27

through a change in the number of competitors per title. A combination of these effects is analyzed in the following section.

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. I thus obtain three equations in three unknowns for each title (one for each format). Accounting for the fact that all titles' quantities demanded are interrelated, this becomes a system of $W \cdot 3$ equations in $W \cdot 3$ unknowns.

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 others). 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 mangers themselves make the entry decisions in the public domain. A different combination of entrants follows, and differences in overall consumer surplus and producer surplus ensue. 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

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

In my application, the difference in consumer surplus from title w is then

$$\triangle E[CS] = \frac{1}{\alpha}\left[\ln\left(1 + \sum_{w' \in \mathcal{W}^1} (D_w^1)^{1-\sigma_2}\right) - \ln\left(1 + \sum_{w' \in \mathcal{W}^0} (D_w^0)^{1-\sigma_2}\right)\right] \tag{6}$$

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 $M \cdot \triangle E[CS]$.

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

$$\triangle E[\pi] = M\left(\cdot \sum_{j \in J_w} (p_j - c_k^1)s_j(n_H^1, n_P^1, n_E^1) - \sum_{j \in J_w} (p_j - c_k^0)s_j(n_H^0, n_P^0, n_E^0)\right)$$

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.

5.1 Moving Titles into the Public Domain

A move of a title into the public domain changes three factors in 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 differences 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 the data. The model also predicts a few more e-books than expected. 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. 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 titles. The increase in consumer surplus outweighs the decrease in profits for all high-quality titles. This indicates that a copyright on a title causes insufficient entry. Table 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.

Table 10: Mean Welfare Changes by Creative Quality (in $ per title over one year)

	Bottom 25%	Mean	Top 25%
$\triangle$CS	642.09	4655.78	15843.14
$\triangle$PS (existing firms)	-47.46	-1186.16	-5044.80
Variable profit (new entrants)	391.23	2576.76	9334.52
Fixed cost (new entrants)	-333.38	-2491.69	-9136.90
$\triangle$ Total Surplus	652.48	3554.69	10995.95

A move into the public domain increases total surplus by an average of $3554.69 each year. Table 10 shows that the magnitude of this effect varies by the title's creative quality. It also varies

30

by the number of editions that are currently available and by other determinants of demand and supply. Appendix A.3 provides further details on the effect of a move into the public domain on a title-by-title basis.

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 differently 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 zero (up to an integer constraint). Table 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 11: Mean Welfare Changes by In-Print Status (in $ per title over one year)

	Currently out of print	Currently in print - similar ϕ_w
$\triangle$CS	2634.09	1535.29
$\triangle$PS (existing firms)	0	-253.79
Variable profit (new entrants)	553.09	958.45
Fixed costs (new entrants)	-493.50	-912.59
$\triangle$ Total Surplus	2693.68	1327.36

Both in-print and out-of-print titles would, on average, increase total surplus if they are moved into the public domain. The magnitude of these changes differ both by the title's in-print status and its creative quality. 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, of over 10,000 books originally published in 1930, only 174 were still in print in 2001[22], whereas Project Gutenberg predicts that "virtually all" pre-1923 public domain titles

---

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

could be available by the end of the decade[23]. A calculation that takes into account the distribution of creative qualities by year can quantify the welfare effect of a copyright extension[24]. Using the set of works in this paper, and a first estimation of the distribution of creative qualities as obtained through download counts of a large number of works that are available at Project Gutenberg, a copyright extension by one year would result in a welfare loss of $10 million to $20 million.

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).

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. But 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 - similarly to the overall effect.

Table 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.

Table 12: Mean Mankiw Effect by Creative Quality (in $ per title over one year)

	Bottom 25%	Mean	Top 25%
$\triangle$CS	578.67	5193.59	15181.68
$\triangle$PS (existing firms)	-318.18	-3727.39	-11674.24
Variable profit (new entrants)	358.78	3124.93	8944.30
Fixed cost (new entrants)	-300.93	-3030.15	-87.54
$\triangle$ Total Surplus	318.35	1560.98	3702.19

---

[23]see the Brief of Amici Curia, The Internet Archive, Prelinger Archives, and Project Gutenberg, here

[24]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 and is left for future work.

The lowest-quality titles are often difficult to come by and are only available in a few editions. For these titles a move into a free-entry regime 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. Note that a switch into a free-entry regime does not affect titles that have run out of print. These titles will not 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.

Table 12 shows that 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 7.
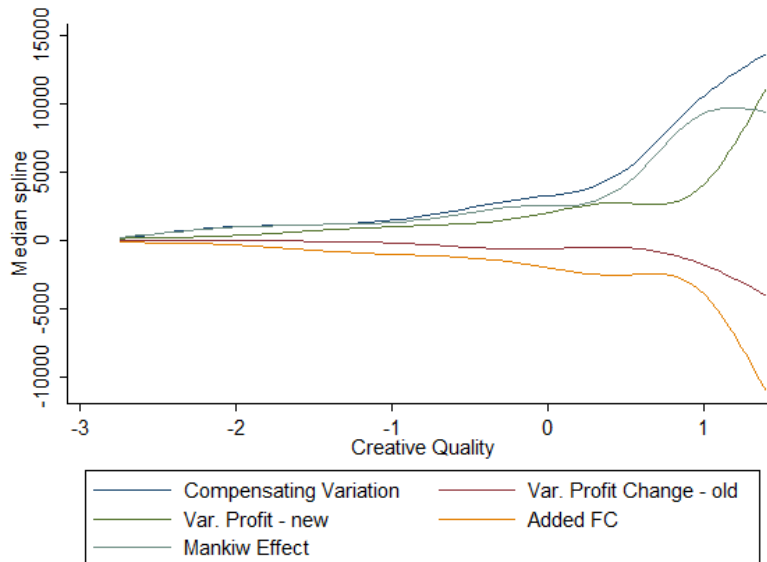

Figure 7: Profits by creative quality

The collective of new entrants produces a large additional operating profit from entering editions into high-quality titles, but much of their operating profit is dissipated by fixed costs. Existing editions lose more profits for those titles as well as they had larger profits before the switch

to free entry. Differences in consumer surplus are small compared to the differences in publishers' profits. The overall effect of free entry is negative for almost all titles due to the presence of fixed costs. The Mankiw & Whinston result of excess entry holds in this market.

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 also 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.

These changes will result in a new combination of entrants and subsequently different prices and quantities. Table 13 shows the effect of a move into the public domain when the public domain does not include a free digital option.

Table 13: Mean welfare changes by creative quality (in $ per title over one year)

	Bottom 25%	Mean	Top 25%
$\triangle$CS	436.47	2655.44	9352.58
$\triangle$PS (existing firms)	-54.78	-847.17	-4300.48
Variable profit (new entrants)	1464.92	11660.52	49840.63
Fixed cost (new entrants)	-1380.68	-11279.06	-48363.14
$\triangle$ Total Surplus	465.93	2189.73	6529.58

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 effects 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 more variety, so

34

that consumers benefit from the existence of Project Gutenberg. 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 30%. 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.

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 IP 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 for 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.

References

[1] Anderson, S., de Palma, A. and J.F. Thisse (1992), "Discrete Choice Theory of Product Differentiation", *Cambridge: MIT Press.*

[2] Berry, S.T. (1994), "Estimating Discrete-Choice Models of Oligopoly Differentiation", *The RAND Journal of Economics*, 25 (2), 242-262.

[3] Berry, S., Levinsohn, J. and a. Pakes (1995), "Automobile Prices in Market Equilibrium", *Econometrica*, 63 (4), 841-890.

[4] Berry, S. and J. Waldfogel (1999), "Free Entry and Social Inefficiency in Radio Broadcasting", *The RAND Journal of Economics*, 30 (3), 397-420.

[5] Bresnahan, T. and P. Reiss (1991), "Entry and Competition in Concentrated Markets", *The Journal of Political Economy*, 99 (5), 977-1009.

[6] Cardell, N.S. (1997), "Variance Components Structures for the Extreme-Value and Logistic Distributions with Application to Models of Heterogeneity", *Econometric Theory*, 13 (2), 185-213.

[7] Chevalier, J. and A. Goolsbee (2003), "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com", *Quantitative Marketing and Economics*, 1, 203-222.

[8] Ferreira, F., Petrin, A., & J. Waldfogel (2012), "Trade, Endogenous Quality, and Welfare in Motion Pictures".

[9] Greco, A. N. (2011), "The Book Publishing Industry", *Taylor & Francis.*

[10] Heald, P. (2008), "Property Rights and the Efficient Exploitation of Copyrighted Works", *Minn. L. Rev. 92*, 1031.

[11] Ishii, J. (2005), "Compatibility, competition, and investment in network industries: ATM networks in the banking industry", *Unpublished working paper.*

[12] Korda, M. (2002), "Making the List: A Cultural History of the American Bestseller 1900-1999", *New York: Barnes & Noble Books.*

[13] Landes, W. M., & Posner, R. A. (1989), "An Economic Analysis of Copyright Law", *The Journal of Legal Studies*, 18(2), 325-363.

[14] Landes, W. M.& Posner, R. A. (2003), "The economic structure of intellectual property law", *Harvard University Press.*

[15] Lee, D. (2008), "Randomized Experiments from Non-Random Selection in U.S. House Elections", *Journal of Econometrics*, 142 (2), 675-697.

[16] Lemley, M. A. (2004), "Ex ante versus ex post justifications for intellectual property", *The University of Chicago law review*, 129-149.

[17] Li, Xing, MacGarvie, Megan & Petra Moser (2012) , "Dead Poets' Property - The Copyright Act of 1814 and the Price of Books in the Romantic Period".

[18] Mankiw, N. G., & M. D. Whinston (1986), "Free entry and Social Inefficiency", *The RAND Journal of Economics*, 48-58.

[19] Mazzeo, M. (2002), "Product Choice and Oligopoly Market Structure", *RAND Journal of Economics*, 33 (2), 221-242.

[20] Morey, E., Rowe, R. and M. Watson (1993), "A Repeated Nested-Logit Model of Atlantic Salmon Fishing", *American Journal of Agricultural Economics*, 75 (3), 578-592.

[21] Mussa, M. and S. Rosen (1978), "Monopoly and Product Quality", *Journal of Economic Theory*, 18 (2), 301-317.

[22] Seim, K. (2006), "An Empirical Model of Firm Entry with Endogenous Product-Type Choices", *RAND Journal of Economics*, 37 (3), 619-640.

[23] Varian, H. R. (2006), "Copyright Term Extension and Orphan Works", *Industrial and Corporate change*, 15(6), 965-980.

[24] Verboven, F. (1996), "International Price Discrimination in the European Car Market", *RAND Journal of Economics*, 27 (2), 240-268.

[25] Vogel, H. L. (2010). "Entertainment Industry Economics: A Guide for Financial Analysis", *Cambridge University Press*.

[26] Williams, H. (2010), "Intellectual Property Rights and Innovation: Evidence from the Human Genome", *NBER Working Paper No. 16213*.

A   Appendix

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 can avoid the estimation of an elasticity parameter because I consider mostly 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 a sold book. 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 the figure below can attest.
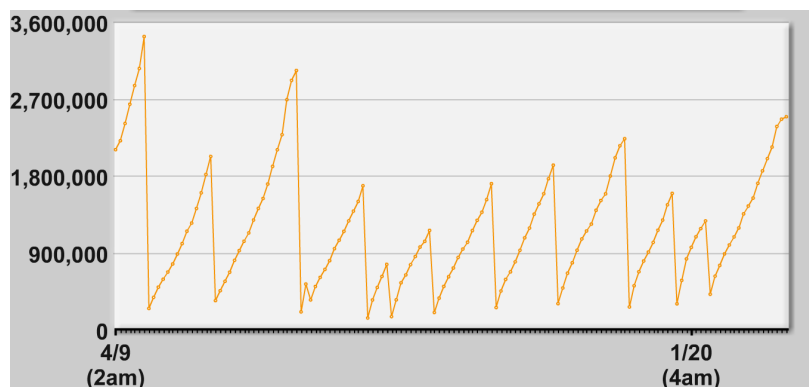

Figure 8: Sales Ranks, Arrowsmith (Mass Market Paperback)

Some books are slightly more difficult to analyse as theye face a higher demand than this particular example, but since I have hourly ranking data, the sales can still mostly be counted by virtue of counting spikes in the rankings.

The demand of 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

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 "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 almost 20 editions per title. This is illustrated in figures 9 and 10.
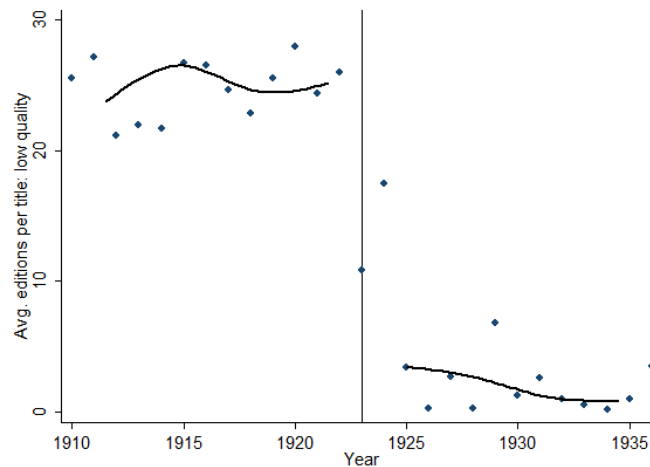

Figure 9: RDD: high-popularity works

Figure 10: 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,

$$
\begin{aligned}
s_{j|wk} &= \frac{\exp\left\{\frac{\delta_j}{1-\sigma_1}\right\}}{D_k} \\
s_{k|w} &= \frac{D_k^{\frac{1-\sigma_1}{1-\sigma_2}}}{D_w} \\
s_w &= \frac{D_w^{1-\sigma_2}}{1+\sum_{w'\in W}\left(D_w^{1-\sigma^2}\right)}
\end{aligned}
$$

This obtains the market shares indicated above:

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

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:

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

A.4 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 13 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 and the title's quality.

Table 14: Low-Quality Titles - All figures in $

Title	$\triangle$(CS)	$\triangle$(PS) - existing	Variable π - new	Fixed cost - new	$\triangle$(welfare)
One Increasing Purpose	167.90	0	151.87	100.13	219.64
Old Pybus	320.34	25.09	154.00	50.12	449.31
Sparkenbroke	180.00	0	154.64	100.02	234.62
A Good Woman	177.03	0	153.96	99.70	230.30
Maid in Waiting	311.09	-124.16	260.77	249.89	197.80
Come and Get It	593.79	-10.90	247.86	220.01	610.73
The Woman of Andros	518.98	-28.53	282.37	260.50	512.31
Doomsday	298.19	13.59	182.68	109.40	385.06
The Blue Window	1018.11	5.17	362.66	284.34	1110.27
Mary Peters	1007.46	-79.07	514.01	490.02	952.37

Medium-Quality Titles - All figures in $

Title	$\triangle$(CS)	$\triangle$(PS) - existing	Variable π - new	Fixed cost - new	$\triangle$(welfare)
The Constant Nymph	4034.92	-275.09	2198.62	2170.00	3788.46
The Plutocrat	1914.91	-60.64	715.42	649.60	1920.08
Swan Song	4368.02	-1.55	1632.86	1591.38	4407.96
Dodsworth	5241.51	-627.06	2521.64	2512.28	4623.81
The Greene Murder Case	3813.48	-96.32	1789.41	1758.48	3748.10
The Green Hat	4556.79	-203.43	3745.61	3659.23	4439.75
Show Boat	1809.38	-2404.38	1452.28	1406.63	-549.34
Shadows on the Rock	2198.58	-960.34	1666.33	1588.33	1316.24
White Banners	3388.37	320.17	1650.43	1601.65	3116.98
A White Bird Flying	2850.38	-1018.13	2268.87	2153.18	1947.93

High-Quality Titles - All figures in $

Title	$\triangle$(CS)	$\triangle$(PS) - existing	Variable π - new	Fixed cost - new	$\triangle$(welfare)
Arrowsmith	14847.23	-6894.52	2323.98	2263.69	8013.00
Of Time and the River	14555.96	-1536.17	3340.02	3169.31	13190.51
The Keeper of the Bees	20910.97	-2427.63	15933.55	15825.80	18591.09
It Can't Happen Here	12741.91	-4024.65	6257.76	6105.74	8869.28
Little Man, What Now?	14235.43	-3228.18	13173.60	12777.52	11403.33
Sons	13680.63	-4410.67	11139.64	10998.32	9411.28
Lost Horizon	20884.93	-4313.72	7961.42	7826.67	16705.96
Years of Grace	23458.94	-2593.45	5765.74	5436.32	21194.91
Gone With the Wind	45428.53	-26470.24	184149.57	180984.69	22123.17
The Good Earth	51433.81	-3103.11	28636.06	28369.51	48597.25