NBER WORKING PAPER SERIES

MATCH QUALITY, SEARCH, AND THE INTERNET MARKET FOR USED BOOKS

Glenn Ellison
Sara Fisher Ellison

Working Paper 24197
http://www.nber.org/papers/w24197

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2018

We would like to thank Vivek Bhattacharya, Masao Fukui, Chen Lian, David Liu, and especially Hongkai Zhang for outstanding research assistance. We thank Panle Jia Barwick, Gabriel Carroll, and many seminar audiences for comments on an earlier drafts that led us to make substantial improvements to the paper. We thank the NBER Sloan Program on Digitization and the Toulouse Network for Information Technology for financial support, and the Paris School of Economics for their hospitality. The first author has received material financial support from the Toulouse Network for Information Technology. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Match Quality, Search, and the Internet Market for Used Books
Glenn Ellison and Sara Fisher Ellison
NBER Working Paper No. 24197
January 2018
JEL No. D22,L13,L81

ABSTRACT

This paper examines the effect of the Internet on markets in which match-quality is important, including an analysis of the market for used books. A model in which sellers of unusual objects wait for high-value buyers to arrive brings out match quality and competition effects through which improved search technologies may increase both price dispersion and social welfare. A reduced-form empirical analysis finds support for a number of more nuanced predictions of the model in the context of the used book market, exploiting both cross-sectional differences across books and time-series differences in the wake of Amazon's acquisition and incorporation of a large used book marketplace. The paper develops a framework for structural estimation of a model based on the theory. The estimates suggest that the shift to Internet sales substantially increased both seller profits and consumer surplus.

Glenn Ellison
Department of Economics, E52-424
MIT
77 Massachusetts Avenue
Cambridge, MA 02139
and NBER
gellison@mit.edu

Sara Fisher Ellison
Department of Economics
MIT, E52-428
77 Massachusetts Avenue
Cambridge, MA 02139
sellison@mit.edu

1 Introduction

The empirical literature on Internet pricing has found that improved search technologies did not have the dramatic price-lowering and law-of-one-price-reinforcing effects that some had forecast.[1] Ellison and Ellison (2009) argued that one explanation was that the Internet provides sellers with the means and motivation to keep markets inefficient by engaging in "obfuscation." The literature now includes a variety of models of obfuscation and empirical studies.[2] This paper can be seen as a Chicago-school critique of this literature.[3] We point out that in other markets, ones where match quality is very important, high and dispersed prices can also be a sign that search is becoming more efficient and producing welfare gains. We take advantage of features of the used book market to test a number of more nuanced predictions of our match-quality theory and find that they are borne out. We then discuss how models like ours can be structurally estimated, and estimate a model which suggests that Internet sales substantially increased both seller profits and consumer welfare.

This paper focuses on the market for used books, which is an extreme example of the Internet's failure to lead to low, uniform prices. Online prices are typically higher than offline prices had been, and are also extremely dispersed.[4] One particularly salient feature of the used book market is the importance of match quality: most people would derive zero utility from owning most used books but positive utility from owning a few specific titles. Our basic argument is that when search costs are reduced, the few people who would value owning any particular used book are more likely to come across it. This phenomenon is akin to an increase in demand. And with supply fixed (or at least highly inelastic), prices increase. We see dispersion as a separate phenomenon arising at the same time due to the Internet being an inherently asymmetric environment. Firms with many repeat customers or who are ranked highly by Google will exploit this advantage by setting a high price and

[1] Brynjolfsson and Smith (2000) found that online book and CD prices were just 9-16% lower than offline prices and price dispersion was actually greater online. Baye, Morgan and Sholten (2004) found an average range of over $100 for consumer electronics products and noted that the more refined law-of-one-price prediction that at least the two lowest prices in a market should be identical also fails dramatically. Search technologies might also be expected to reduce price dispersion at physical stores, but Kaplan and Menzio's (2015) study of a broad sample of products sold (mostly) through traditional retail stores in 2004-2009 finds substantial dispersion even after one takes out store quality effects and within-store cross-time variation.

[2] See among others Ellison (2005), Spiegler (2006), Wilson (2010), Ellison and Wolitzky (2012), Piccione and Spiegler (2012), Chioveanu and Zhou (2013), Grubb (2015), Armstrong and Zhou (2016), and Celerier and Vallee (2017).

[3] Moraga Gonzalez, Sandor, and Wildenbeest (2017) offer another critique in noting that in models with heterogeneous search costs, reductions in search costs can sometimes lead to higher equilibrium price distributions.

[4] The first online vs. offline comparison paper we know of, Bailey (1998), reported that online prices for CDs were higher than offline prices, but we have not seen such a finding in any later papers.

waiting for a (relatively uninformed) high value consumer to arrive; whereas, less-visible sellers will need to enter the aggressive competition to serve highly informed consumers.

From a theory perspective, our story is a simple one, but one that has yet to be spelled out in detail, as far as we can tell. Accordingly, our paper is divided into three main parts. We first formalize the model of used book shopping and price-setting that we have described verbally. Then we try to establish that it is the right story for used books, through a series of reduced-form tests of more nuanced implications of our model using data we collected over several years in both online and brick and mortar used book stores. Finally, we develop a structural model along the lines of our theoretical model that is amenable to estimation, even with our limited data set, and present estimates of the parameters. In doing so, we draw out our model's implications for welfare effects.

A personal anecdote may be illustrative. Several years ago, one of us wanted a thirty year old academic book on the pharmaceutical market which the MIT library did not have. The book had long been out of print, and looking for a used copy in brick and mortar stores would be like looking for a needle in a haystack. A quick search on Alibris, however, produced four or five copies for sale. A copy was ordered, for around \$20, and it arrived shortly, with \$0.75 written in pencil on the inside front cover and subsequently erased![5] The book had evidently been languishing on the shelf of some used book store for years, and not a single customer who noticed it was willing to pay even \$0.75. A researcher needing the book happily paid \$20 and would have paid significantly more.

While this story and the empirical estimates we present are about used books, we think that similar insights should apply to the Internet's impact on a range of retail and labor markets. A wide variety of collectibles and used goods exist in limited supply and would be highly valued only by a tiny fraction of the population. For instance, eBay's founder has said that an important early experience was when he encountered a collector of broken laser pointers (Cohen, 2002). The two Boeing 747's recently sold in an online auction also fit this basic description, as might the labor of many individual craftsmen, home repair contractors, or Upwork freelancers, or the couches rented on Airbnb.

Section 2 uses simple models to bring out observations about match-quality markets. We start with a model in which sellers of unique items post prices and wait for sufficiently high-valuation buyers to arrive. The model highlights a "match-quality effect." If improved search technologies increase customer flows, then sellers will increase their prices. But this is not a bad thing: increased equilibrium match quality between buyers and items leads to

[5] The exact details of this story are lost to history, but the basic facts are correct.

increases in social welfare. We also note that price dispersion will naturally be present if some sellers are more visible than others, and we present a sufficient statistic result relevant to the estimation of welfare gains. We then use an oligopoly model in which sellers are in competition for some more sophisticated consumers to highlight a "competition effect," which tends to pull down prices at the low-end of the price distribution and can be a second channel through which improvements in search technologies lead to increased price dispersion.

Section 3 discusses the market for used books. Match quality is obviously very important in this market – most people have zero interest in owning most of the millions of books in this world – and the market has been dramatically affected by the Internet. We think that the market is of interest in its own right in part because Amazon plays a major role. An additional strong motivation for studying the market is that several features make it a natural setting for a general empirical study of match-quality markets: the fact that books have ISBN numbers makes it easy identify each offered item as one of a well-defined set of "products"; it is feasible to collect the prices at which thousands of merchants offer (or do not offer) any particular product; and there is natural variation both across titles and within-titles over time that one can examine to gain insights on models and their applicability. In particular, we highlight four sets of more nuanced predictions of the theory that one can examine in the used books context.

Section 4 discusses our dataset. We began by visiting several physical used book stores to construct a pseudo-random sample of books intended to be representative of what those stores were offering in 2009, with one group oversampled, as we will describe later. Shortly thereafter, we collected data on the prices at which a large number of online bookstores were offering the same titles by scraping data from a website, AbeBooks.com, which aggregated listing from thousands of Internet bookstores. At that time, Amazon had recently acquired AbeBooks, but the site was still operating independently. Amazon subsequently launched a program to have AbeBooks' listings appear under the "buy used" link on Amazon's site, which presumably increased the number of well-informed consumers in the market. Motivated by this change we collected listing prices for the same set of titles in November of 2012. And we collected listings again two months later so that we could get a proxy for sales by looking for listings that sellers removed over the Christmas season. Combining data from our various collection efforts, we can make several types of comparisons: within-title online versus offline prices in 2009; within-title 2009 versus 2012 online price distributions; within-time comparisons of price distributions for different types of books; and analyses of demand.

Section 5 presents our most basic observations about online used book prices which we mentioned at the very start. By 2009 online prices were substantially higher on average than contemporaneous offline prices for the same titles. And price dispersion was also much greater online than offline. (By necessity we only have a single offline price for each title in our dataset, but can still place bounds on how much offline dispersion there could have been by exploiting a variance decomposition.) The majority of the section is then devoted to examining the four sets of more nuanced cross-sectional and cross-time implications of the theory discussed at the end of section 3. In several cases we find very strong support for the predictions, bolstering the case that the model is providing a good description of the basic economics of the market.

Section 6 then develops and estimates a structural model. This model is closely based on the theoretical model of secton 2: well-informed and less well-informed potential consumers arrive at Poisson rates, firms are heterogeneous in the arrival processes they face, and products are sufficiently differentiated so that pure strategy dispersed price equilibria exist. In the model there is a one-to-one correspondence between prices and arrival rates. This fact makes it possible to back out firm-specific arrival rates from observed prices, which makes the model relatively easy to estimate via simulated maximum likelihood and lets us avoid some difficulties associated with endogeneity while using our demand data. Given the limited size of our dataset we must estimate a fairly parsimonious model, but are able to give the model some flexibility to estimate most quantities of interest. Our structural estimates indicate that arrival rates for a particular title are substantially higher at online stores than offline stores (although arrival rates are still very low for some titles), that online demand includes a very price-sensitive "shopper" segment, and that firms also receive small inflows of much less price sensitive "nonshoppers." When we use these primitives to estimate profit and welfare effects we find strong support for the view that the Internet's tranformation of the used book market has been welfare enhancing. We estimate that the shift to online sales increased sellers' discounted per-title revenues by about 80% and that welfare gains are roughly evenly shared between firms and consumers. Per-listing profit levels appear to have declined somewhat between 2009 and 2012, perhaps due to an Amazon-driven increased use of price comparison tools, but there is also at the same time an increase in the number of listed titles.

Our paper is related to a number of other literatures, both theoretical and empirical. One related empirical literature explores facts similar to those that motivate our analysis

– comparing online and offline prices for various products and documenting substantial online price dispersion.[6] Another (much smaller) related literature provides reduced-form evidence that price distributions appear consistent with models of heterogeneous search. Two noteworthy papers here are Baye, Morgan, and Scholten (2004), which discusses the implications of several theoretical models and notes that dispersion is empirically smaller when the number of firms is larger, and Tang, Smith, and Montgomery (2010), which documents that prices and dispersion are lower for more frequently searched books. A number of other papers explore other issues in the book market including Chevalier and Goolsbee (2003), Brynjolfsson, Hu, and Smith (2003), Ghose, Smith, and Telang (2006), and Chevalier and Goolsbee (2009). The focus of Brynjolfsson, Hu, and Smith (2003) is most similar in that it also estimates welfare gains from Internet book sales. In their case, the consumer surplus improvement results from Amazon making books available to consumers that they would have been unable to purchase at traditional brick and mortar stores.

Our theoretical section is intended primarily to bring out some basic intuitions as simply as possible, but we have not found an exact antecedent in the literature for our simple initial analysis of monopoly pricing with Poisson arrivals.[7] We also have not seen another derivation of our sufficient statistic result about welfare estimation, although Stiglitz (1976) does have the related fact that monopoly pricing is socially optimal in our model with constant elasticity demand. Our subsequent consideration of oligopoly pricing follows the literature on pricing and price dispersion with consumer search, including Salop and Stiglitz (1977), Reinganum (1979), Varian (1980), Burdett and Judd (1983), Stahl (1989), and Baye and Morgan (2001).[8] Relative to many of these papers, we simplify our model by focusing exclusively on the firm pricing problem without rationalizing the consumer search. Our approach of focusing on pure-strategy equilibria with heterogeneous firms harkens back to Reinganum (1979), although the structure of the population is more similar to that of Baye and Morgan (2001).

Another active recent literature demonstrates how one can back out estimates of search costs from data on price distributions under rational search models. An early paper was

[6]See, for example, Bailey (1998) Brynjolfsson and Smith (2000), Clay, Krishnan, and Wolff (2001), Baye, Morgan, and Scholten (2004), and Ellison and Ellison (2009).

[7]The model could be thought of as a simplified special case of the dynamic inventory model of Arrow, Harris, and Marschak (1951). Also, similar stopping time problems with price offers made by buyers rather than sellers are in a number of papers, including Karlin (1962) and McCall (1970). In addition, there are substantial literatures covering more complex dynamic monopoly problems with inventory costs, finite time horizons, learning about demand, etc. See Talluri and Van Ryzen (2004).

[8]See Baye, Morgan, and Scholten (2006) for a survey that brings together many of these models.

Sorensen (2001), which performed such an estimation in the context of prescription drug prices. Hortacsu and Syverson (2004) examine index mutual funds. Hong and Shum (2006) discuss both a nonparametric methodology and an application involving used book prices. Subsequent papers extending the methodology and examining other applications include Moraga Gonzalez and Wildenbeest (2008), Kim, Albuquerque, and Bronnenberg (2010), Brynjolfsson, Dick, and Smith (2010), De los Santos, Hortacsu, and Wildenbeest (2012) (which also studies consumers shopping for books), Moraga Gonzalez, Sandor, and Wildenbeest (2013), and Koulayev (2014). Relative to this literature, we will not try to estimate search costs to rationalize demand and we will not posit identical firms – instead we focus on estimating a consumer arrival process from price distributions (and some quantity data) in a model that allows for substantial firm-level heterogeneity. Our motivation is also quite different: these papers focus on estimating the distribution of search costs which rationalizes price distributions, whereas we are most interested in what we can learn about consumer demand and welfare from those price distributions.

2 A Model

In this section we discuss monopoly and duopoly models of markets with fixed supply in which match quality is important. The models are quite simple, but illustrate several ways in which search technology improvements may affect price levels and price dispersion, and also bring out some important observations related to the estimation of welfare effects.

2.1 A monopoly model

We begin with a simple dynamic monopoly model. One can think of it as a model of a brick-and-mortar store or of an Internet store serving customers unaware of any other sites that offer the same product. It will also serve as a starting point for our subsequent analysis of an oligopoly model in which some consumers do also search across stores.

Suppose that a monopolist has a single unit of a good to sell. It may store the good in inventory at zero cost, but earnings from sales at any future date t will be discounted by a factor of e^{-rt}. Consumers randomly arrive at the monopolist's store according to a Poisson process with rate γ. The valuation v of each arriving consumer is an independent draw from a distribution with CDF $F(v)$. Consumers buy if and only if their value exceeds the firm's price so the probability that a consumer who arrives will buy is $D(p) = 1 - F(p)$. Assume that $\lim_{p\to\infty} pD(p) = 0$, which implies that optimal prices are finite.

One can think about the dynamic optimal monopoly price in two different ways. One is simply to compute the discounted expected profit $\pi(p)$ obtained from any fixed price p. Intuitively, expected profit is simply $E(pe^{-r\tilde{t}})$ where $\tilde{t}$ is the random variable giving the time at which the good is sold and r is a discount rate. Consumers willing to pay at least p arrive at Poisson rate $\gamma D(p)$. The density of the time of sale is then $f(t|p) = \gamma D(p)e^{-\gamma D(p)t}$ and the expected profit is

$$\begin{aligned}
\pi(p) &= \int_0^\infty pe^{-rt} f(t|p)dt \\
&= \int_0^\infty pe^{-rt}\gamma D(p)e^{-\gamma D(p)t}dt \\
&= \frac{\gamma p D(p)}{r + \gamma D(p)}
\end{aligned}$$

Hence, one way to think of the dynamic optimal monopoly price p^m is as the maximizer of this expression:

$$p^m = \underset{p}{\operatorname{argmax}}\, \pi(p) = \underset{p}{\operatorname{argmax}}\, \frac{\gamma p D(p)}{r + \gamma D(p)}.$$

Note that expected profits are zero in both the $p \to 0$ and $p \to \infty$ limits, so an interior optimum exists if $D(p)$ is continuous. The monopoly price will satisfy the first-order condition obtained from differentiating the above expression if $D(p)$ is differentiable. Note also that $\pi(p)$ only depends on γ and r through the ratio γ/r. This is natural because the scaling of time is only meaningful relative to these two parameters, arrival rate and discount rate.

The second way to think about the dynamic profit maximization problem is as a dynamic programming problem. Let π^* (which depends on $\gamma, r,$ and $D()$) be the maximized value of $\pi(p)$. This is the opportunity cost that a monopolist incurs if it sells the good to a consumer who has arrived at its shop. Hence, the dynamic optimal monopoly price is also the solution to

$$p^m = \underset{p}{\operatorname{argmax}}(p - \pi^*)D(p).$$

Looking at the problem from these two perspectives gives two expressions relating the dynamic monopoly price to the elasticity of demand:

Proposition 1 *Suppose $D(p)$ is differentiable. The dynamic monopoly price p^m and the elasticity of demand ϵ at this price are related by*

$$\frac{p^m - \pi^*}{p^m} = -\frac{1}{\epsilon},$$

and

$$\epsilon = -\left(1 + \frac{\gamma}{r} D(p^m)\right).$$

We leave the proof of this and other propositions to the appendix to streamline the presentation.

Remarks:

1. In contrast to the static monopoly pricing problem with zero costs where a monopolist chooses p so that $\epsilon = -1$, the monopolist in this problem prices on the elastic portion of the demand curve to reflect the opportunity cost of selling the good.

2. The expressions in Proposition 1 are first-order conditions that one can solve to obtain expressions for the monopoly price given a particular $D(p)$. For example, if values are uniform on $[0,1]$ so $D(p) = 1-p$, they can be solved to find $p^m = \frac{\sqrt{1+(\gamma/r)}}{1+\sqrt{1+(\gamma/r)}}$.
Another tractable example is a truncated constant elasticity demand curve: $D(p) = \min\{1, hp^{-\eta}\}$. Here, the monopoly price is

$$p^m = \begin{cases} \left(\frac{h}{\eta-1}\right)^{1/\eta} \left(\frac{\gamma}{r}\right)^{1/\eta} & \text{if } \frac{\gamma}{r} > \eta - 1 \\ h^{1/\eta} & \text{otherwise} \end{cases}$$

If more consumers visit online stores than offline stores, then the comparative statics with respect to the arrival rate are relevant to online-offline price differences.

Proposition 2 *The monopoly price p^m is weakly increasing in $\frac{\gamma}{r}$.*

Remarks:

1. The $\gamma/r \to \infty$ limit of the monopoly price depends on the support of the consumer value distribution. When the value distribution has an upper bound, the monopoly price will approach the upper bound. When there is no upper bound on consumer valuations, the monopoly price will go to infinity as $\gamma/r \to \infty$.[9]

2. The rate at which p^m increases in γ/r depends on the thickness of the upper tail of the distribution of consumer valuations. In the uniform example, the monopoly price increases rapidly when γ/r is small, but the effect also diminishes rapidly. In the truncated constant elasticity example, the monopoly price is proportional to $(\gamma/r)^{1/\eta}$. In the extremely thick-tailed version of this distribution with η slightly larger than

[9]To see this, note that for any fixed p and ϵ, $\pi(p+\epsilon, \gamma/r) = \frac{p+\epsilon}{1+r/(\gamma D(p+\epsilon))} \to p+\epsilon$ as $\gamma/r \to \infty$. Hence, the monopoly price must be larger than p for γ/r sufficiently large.

1, the monopoly price is almost proportional to the arrival rate. When the tail is thinner, i.e., when η larger, it increases more slowly.

Online and offline used book dealers may also differ in the distribution of consumer values. For example, the probability that a consumer searches for a particular book online may be increasing in the consumer's valuation, whereas consumers who are browsing in a physical bookstore will come across titles for which they have low and high valuations. One way to capture such an effect would be to assume that offline searchers' valuations are random draws from $f(v)$, while consumers with value v only search online with probability $q(v)$, so the density of valuations in the online searcher population will be $g(v) = af(v)q(v)$ for some constant a. If $q(v)$ in increasing, then g is higher than f both in the sense of first-order stochastic dominance and in having a thicker upper tail: $\frac{1-G(x)}{1-F(x)}$ is increasing in x. The following proposition shows that shifts in the distribution satisfying the latter condition increase the monopoly price holding the arrival rate constant.

Proposition 3 *Let $p^m(\gamma/r, F)$ be the monopoly price when the distribution of valuations is $F(x)$. Let $G(x)$ be a distribution with $\frac{1-G(x)}{1-F(x)}$ increasing in x. Then $p^m(\gamma/r, G) \geq p^m(\gamma/r, F)$.*

The monopoly model with constant elasticity demand has some important welfare properties relevant to our and other empirical implementations. The first is an alignment result related to Stiglitz's (1976) observation about monopolistic exploitation of exhausible resources. The second is a sufficient statistic result about computing social welfare.

Proposition 4 *Suppose that the distribution of consumer valuations is such that demand has the truncated constant elasticity form and that the monopolist's price is not at the kink in the demand curve. Then,*
(i) The monopoly price maximizes social welfare.
(ii) Expected social welfare is $E(W) = p^m$.

Result (i) is part of the motivation for our earlier description of this paper as a Chicago-school critique. It suggests that market power need not lead to inefficiency.

On the positive side, result (ii) can be seen as a powerful observation for estimating social welfare. Any given price can be rationalized by a variety of (γ, h, η) combinations, but the proposition implies we do not need to estimate these underlying parameters to estimate social welfare. We also, however, view the result as an important cautionary note. Assuming that firms are pricing optimally and that demand has the constant elasticity form are strong assumptions. Welfare would not be parameter-independent if we depart from these assumptions. But estimated welfare in a rational-seller model will be approximately equal to the price of the good unless one estimates a demand equation that is sufficiently flexible so that $E(v - p^m | v > p^m)$ need not be approximately equal to what it would be with a constant elasticity demand curve that matched the estimated demand elasticity at the observed price. Estimating demand flexibly for prices outside the range of most of the data will typically be extremely difficult.

2.2 An oligopoly model

We now discuss related oligopoly models. We begin with a simple symmetric full-information model which serves as a building block. And we then discuss an asymmetric oligopoly model in which firms serve both comparison shoppers and a local market.

Suppose that there are N firms in the market. Suppose there is a flow arrival rate γ_0 of shoppers who visit all N firms. Assume that shoppers buy from firm i with probability $D(p_i, p_{-i})$ and that this demand function is symmetric, twice-differentiable, weakly decreasing in p_i, and weakly increasing in p_{-i}. Assume also that the set of feasible prices is a compact interval so $\text{argmax}_p \, pD(p, p_{-i})$ always exists. As in the monopoly model, we are interested in modeling a firm endowed with a single unit of the good to sell that faces a dynamic waiting-time problem. In the oligopoly case it is natural that the dynamic problem would have a time-varying component: a firm should anticipate that competition will become more and less intense as additional sellers enter and current sellers sell their goods. Optimal pricing in such a setting could be an interesting topic to explore, but in this paper we consider a simpler stationary model: we assume that whenever one of a firm's rivals makes a sale, the rival is instantaneously replaced by an identical entrant.[10] Profits in the dynamic model then relate to those of the static model as in the monopoly case:

$$\pi(p_i, p_{-i}) = \frac{\gamma_0 p_i D(p_i, p_{-i})}{r + \gamma_0 D(p_i, p_{-i})}.$$

In the static version of this model with a nonzero marginal cost c, it is common to assume that demand is such that $\pi^s(p) = (p_i - c)D(p_i, p_{-i})$ has increasing differences in p_i and p_{-i}. The game is then one with strategic complements: best response correspondences $BR_i(p_{-i})$ are increasing, and results on supermodular games imply that a symmetric pure

[10] Our model can also be thought of as a model of boundedly rational sellers who are able to estimate the expected rate of arrival of consumers willing to buy from them at p given the observed set of rival prices (which is probably already unrealistically demanding), but who are not sophisticated enough to be able to assign probabilities to future paths of rival prices and integrate expected demand over all such paths.

strategy Nash equilibrium always exists (Milgrom and Roberts, 1990). These results would carry over to our dynamic model.

Proposition 5 *Suppose $\pi^s(p) = (p_i - c)D(p_i, p_{-i})$ has increasing differences in p_i and p_{-i} when $p_i > c$. Then, best response correspondences in the the dynamic oligopoly model are weakly increasing, and a symmetric pure strategy Nash equilibrium exists.*

Remarks:

1. As in the monopoly model, equilibrium prices in the full information oligopoly model are increasing in the arrival rate γ_0. Each individual best response is increasing in γ_0 by the same argument as in the monopoly case. And then the comparison of equilibria follows as in Milgrom and Roberts (1990). The static oligopoly model corresponds to $\gamma_0 = 0$, so this implies that prices in the dynamic oligopoly model are higher than those in the static model.

2. A more precise statement of the previous remark is that the set of Nash equilibrium prices increases in γ_0 in the strong set order. The dynamic oligopoly model may have multiple equilibria even when the static model has an unique Nash equilibrium. For example, in a duopoly model with $D_i(p_1, p_2) = \frac{1}{9}(1 - p_i + \frac{3}{2}p_{-i})$, the static ($\gamma_0 = 0$) model has $p^* = 2$ as its unique symmetric PSNE, whereas the dynamic model with $\gamma_0/r = 1$ has both $p^* = 4$ and $p^* = 10$ as symmetric PSNE. Intuitively, the dynamic effect creates an additional complementarity between the firms' prices: when firm 2's price increases, firm 1's opportunity cost of selling the good increases, which provides an additional motivation for increasing p_1.

3. Although it is common to assume that demand is such that $(p - c)D(p_i, p_{-i})$ has increasing differences, it is implausible that the assumption would hold at all price levels. For example, the assumption is globally satisfied in the linear demand case $D_i(p_i, p_{-i}) = 1 - p_i + ap_{-i}$, but assuming that this formula holds everywhere involves assuming that demand is negative for some prices.[11] In such static models it is common to modify the demand function in some cases, for example assuming demand is zero whenever the formula gives a negative answer. The modifications only affect cases that are unimportant so the equilibrium set is unchanged and best

[11]Prices for which demand is greater than one are also inconvenient for our interpretation of demand as a probability of purchase, but this can often be dealt with by scaling demand down by a constant and increasing all arrival rates by the same constant.

responses remain upward sloping. Similar modifications should also typically produce a more plausible dynamic oligopoly model without affecting the equilibrium or the best response functions. But the modified models will not globally have the increasing differences property.

In practice, there is a great deal of price dispersion in markets for used books (and other items). The most common approach to explain such dispersion in the IO theory literature is to assume that some consumers are not fully informed about prices.[12] A simple way to incorporate a similar mechanism in the above framework is to consider a hybrid of the monopoly and full-information oligopoly models above and the gatekeeper model of Baye and Morgan (2001). In particular, let us assume that there are $N+1$ populations of consumers. There is a flow with arrival rate γ_0 of shoppers who visit all N online firms. And for each $i \in \{1, 2, \ldots, N\}$, assume there is a flow with arrival rate γ_i of nonshoppers who visit only firm i. Assume that nonshoppers again buy from firm i with probability $D^m(p_i) \equiv 1 - F(p_i)$ as in the monopoly model. Assume that shoppers buy from firm i with probability $D(p_i, p_{-i})$ as in the full information oligopoly model. So, in other words, online stores have a flow of consumers for whom they are effectively monopolists, the nonshoppers, and a flow of consumers for whom they are effectively oligopolists competing with other stores carrying the same title, the shoppers. We treat offline stores as only having a flow of nonshoppers.

Again, we assume each firm that makes a sale is immediately replaced by an identical entrant. Expected firm profits can then be calculated just as in the monopoly model:

$$\pi_i(p_i, p_{-i}) = \frac{p_i(\gamma_i D^m(p_i) + \gamma_0 D(p_i, p_{-i}))}{r + \gamma_i D^m(p_i) + \gamma_0 D(p_i, p_{-i})}$$

For the reason noted in the final remark after Proposition 5, this objective function would not be expected to satisfy increasing differences at all prices. And here the departures are consequential: the model will not have a pure strategy Nash equilibrium for some parameter values. Intuitively, if the oligopoly demand function is very price sensitive and two firms have nearly identical γ_i, then there cannot be an equilibrium where both firms set nearly identical high prices because each would then like to undercut the other. There also cannot be an equilibrium with nearly identical low prices because the firms would then gain from jumping up to the monopoly price to exploit their nonshoppers. For other parameters, however, there will be a pure strategy equilibrium in which firms with more nonshoppers

[12] Among the classic papers in this literature are Salop and Stiglitz (1977), Reinganum (1979), Varian (1980), Burdett and Judd (1983), Stahl (1989).

set a higher price. This will occur when the oligopoly demand is less price sensitive, the shopper population is relatively small, and/or when arrival rates γ_i of nonshoppers are farther apart.

When a pure strategy Nash equilibrium exists, the equilibrium prices p_i^* will satisfy the first-order conditions which can be written as:

$$\begin{aligned} 0 &= rp_i^*\gamma_i D^{m\prime}(p_i^*) + r\gamma_i D^m(p_i^*) + \gamma_i^2 D^m(p_i^*)^2 \\ &+ rp_i^*\gamma_0 \frac{\partial D}{\partial p_i}(p_i^*, p_{-i}^*) + r\gamma_0 D(p_i^*, p_{-i}^*) + \gamma_0^2 D(p_i^*, p_{-i}^*)^2 \\ &+ 2\gamma_0\gamma_i D^m(p_i^*) D(p_i^*, p_{-i}^*). \end{aligned}$$

Note that the first line of this expression is γ_i, the nonshopper arrival rate, times the expression from the monopoly first-order condition. If the monopoly demand function is single peaked, it is positive for $p < p^m$ and negative for $p > p^m$. The second line of the FOC is γ_0, the shopper arrival rate, times the first-order condition from the oligopoly model in which all consumers are shoppers. When the shoppers-only oligopoly game has single-peaked profit functions and increasing best responses, this term will be positive for the player i setting the lowest price p_i if p_i is less than the lowest equilibrium price of the full-information oligopoly game. The third term is everywhere positive. Hence, when the monopoly price p^m is above the equilibrium price in the shoppers-only oligopoly model, all solutions to this $N + 1$ population model will have firms setting prices above the shoppers-only oligopoly level.

Roughly, one can think of the solution as being that firms with γ_i large relative to γ_0 (with many nonshoppers relative to shoppers), will set prices close to $p^m(\gamma_i)$.[13] Meanwhile, firms with γ_i small will set prices somewhat above shoppers-only oligopoly level both because of the third term in the FOC and because some of their rivals are mostly ignoring the shopper population and pricing close to $p^m(\gamma_{-i})$.

Note that the mechanism behind the price dispersion is somewhat different from that of Baye and Morgan's (2001) gatekeeper model. In Baye and Morgan's model price dispersion is a mixed strategy outcome made possible by the fact that there is a positive probability that no other firms will be listed with the clearinghouse. We have modified the model in two ways to get dispersion as a pure strategy phenomenon: we add product differentiation in the shopper segment to eliminate the discontinuity in demand; and we add exogenous firm heterogeneity (in the consumer arrival rates) to make asymmetric pricing natural. Given

[13]Prices may be lower than $p^m(\gamma_i)$ because of the oligopoly demand, but may also be higher because the shoppers also constitute an increase in the arrival rate.

that arrival rates can be thought of as creating different opportunity costs of selling the good, the model can be thought of as more akin to that of Reinganum (1979) which first generated dispersed price equilibria via heterogeneous costs.

2.3 A numerical example

Figure 1 illustrates how one might think of the difference between offline and online prices in light of this model. We think of prices as differing because of two effects. First, differences in the flow of nonshoppers may make online *monopoly* prices higher than offline monopoly prices.[14] Second, online prices will be pulled down by the competition effect as firms (especially those with low arrival rates of nonshoppers) compete to attract the shopper population.

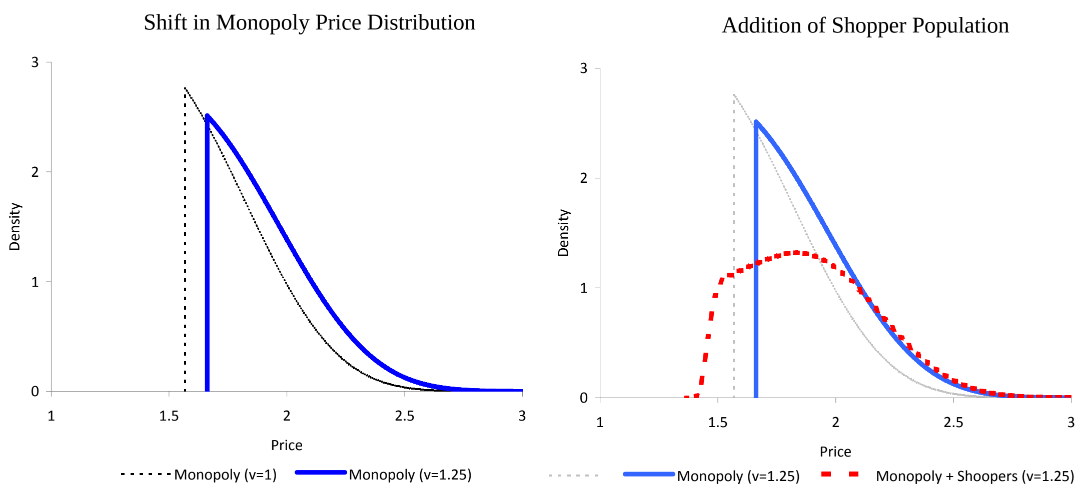

Figure 1: Numerical example: Effects of increasing valuations and adding shoppers

The left panel of Figure 1 illustrates the first effect. The thinner dashed line graphs the distribution of offline monopoly prices for one specification of the demand/arrival process. Each consumer j arriving at firm i is assumed to get utility $1 - p_i + \epsilon_{ij}$ if he purchases from firm i and ϵ_{0j} if he does not purchase, where the ϵ_{ij} are independent type 1 extreme value random variables. The heterogeneous arrival rates γ_i, which lead firms to set different prices, are assumed to be exponentially distributed with mean 1. The thicker solid line is the distribution of monopoly prices that results if we shift the distribution of consumer valuations upward: we assume the utility of purchasing is now $1.25 - p_i + \epsilon_{ij}$. We think

[14] We can decompose this effect into two components. Selection into searching may result in the distribution of searchers' values being higher, and the reduced cost of searching may make the customer arrival rate higher.

of this as the online monopoly price distribution. The gap between the two distribution illustrates how the higher valuations in the online population would lead to higher prices if retailers retained their monopoly power.

The right panel illustrates the competition effect. The thick solid line is the online monopoly distribution from the left panel. The thick dashed line is the distribution of equilibrium prices in a nine-firm oligopoly model.[15] Each firm in this model faces a non-shopper arrival process identical to the online monopoly process. But in addition there is also a population of shoppers who arrive at Poisson rate $\gamma_0 = 2$, see the prices of all firms, and buy from the firm that provides the highest utility if it is greater than the utility of the outside good (with random utilities as in the online monopoly model). Note that the competition effect (the difference between the thick solid blue and thick dashed red lines) is different at different parts of the distribution. There is essentially no competition effect on the upper part of the distribution: firms with high nonshopper arrival rates essentially ignore the shopper population.[16] But at the lower end of the distribution it is powerful and oligopoly prices are substantially below monopoly levels: firms with low nonshopper arrival rates compete aggressively for shoppers.

The comparison between the the light gray line and the dark dashed red line illustrates how the two effects in combination increase price dispersion. The match quality effect pushes up the high end of the price distribution. Meanwhile the competition effect pulls down (or at least offsets some increases in) the lower tail.

3 The Market for Used Books

In this section we provide some background on the used book market and discuss more nuanced predictions of our theoretical model in this context that we will examine in section 5 to assess its applicability.

3.1 Background

Our analysis of the used book market will focus on the run-of-the-mill titles that for many years lined the shelves of physical used bookstores. It includes biographies, literary fiction, detective novels, histories, former best sellers, and a wide variety of others, most of which

[15]To generate this figure, we repeatedly took nine independent draws from the assumed distribution of nonshopper arrival rates and solved for the pure strategy equilibrium prices. The graphed distribution is a kernel smoothing of the prices obtained from many such draws.

[16]Online oligopoly prices are actually slightly higher than the online monopoly prices due to the extra shopper demand.

were out of print. Many used book dealers began also selling online in the early to mid 1990's. In that pre-Google era, however, it was probably unlikely that a consumer looking for an uncommon title could find a copy just by typing the name into a search engine. In the second half of a 1990's a number of sites including AbeBooks, Alibris, Biblio, Bibliofind, and Bookfinder pursued the business model of aggregating listings and making referrals. AbeBooks, which initially just aggregated listings of physical stores in Victoria B.C., grew to be the largest aggregator (in part by acquiring Bookfinder) with 20 million listings by 2000 and 100 million by 2007. Alibris is of comparable size. Amazon acquired AbeBooks om 2008. Initially, this had little impact as Amazon left AbeBooks to operate as it had. But in 2010 Amazon launched a program to have AbeBooks listings also appear underAmazon's "buy used" button. This may have substantially increased the number of consumers who viewed aggregated listings.

People find used books online in multiple ways, which we think fit well with the model's assumption that there are both "shoppers" and "nonshoppers". Shoppers would include both regular purchasers of used books familiar with Abebooks and Alibris and those who search on Amazon and click on a buy used link. Such customers will usually find a list of merchants offering the book ordered on the shipping-inclusive price. Others look for books via a variety of other methods. Customers who go directly to their favorite online used bookstore would be most similar to the nonshoppers in our model.[17] The behavior of the many consumers who type a book's title into Google is probably also fairly well described by our nonshopper model. For example, figure 2 shows the top part of the page returned when we searched for one of the books in our sample using both the first part of the book's title and the phrase "used book". Some consumers might be led by the first two links to become searchers and look on AbeBooks or Amazon. But some others would just buy copy prominently displayed in the Thrift Books ad, figuring that $5.69 is a fair price and they have better things to do than engage in a search odyssey to try to save a dollar or two. And others who are comfortable with eBay would presumably click on the bottom organic link to buy the copy that AwesomeBooks is selling through eBay. Note that the nonshopper channels will naturally lead to different stores having different arrival rates: some stores will have many more repeat customers than others; and some are much more likely than others to appear near the top of a Google search for a particular title.

Several features of the used book market make it attractive as a setting in which to

[17] According to Alexa.com's traffic estimator Thriftbooks.com is among the 4000 sites most visited by US consumers and Powells.com and Betterworldbooks.com are both among the top 20,000.

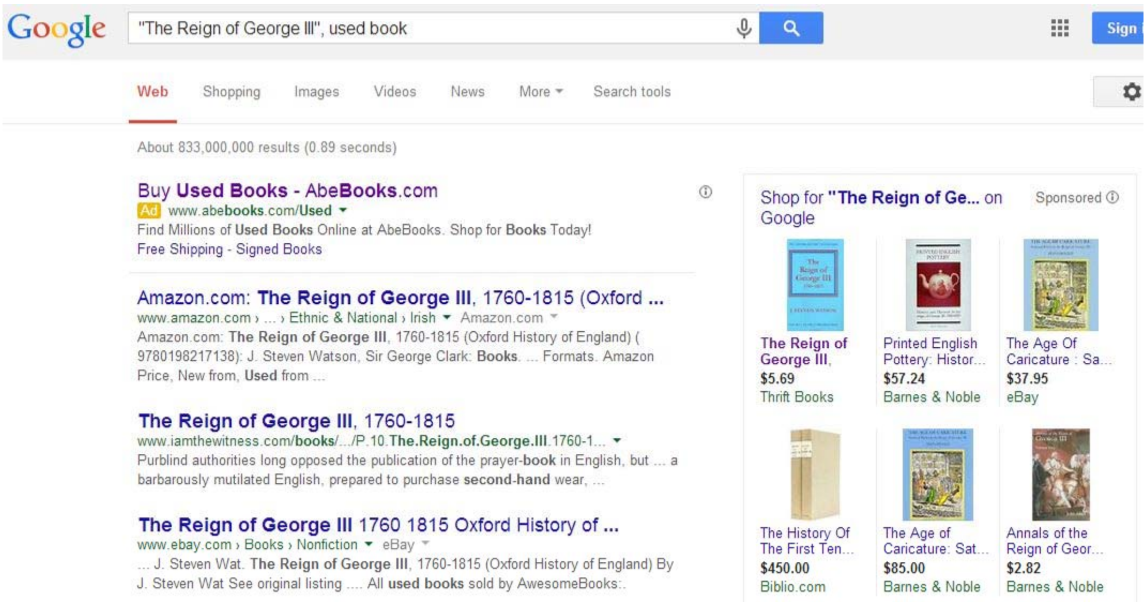

Figure 2: A sample Google search for a used book

study the effect of the Internet on match-quality markets. First, books are well-defined products and AbeBooks scale makes it feasible to find the prices at which thousands of different sellers are (or are not) offering each product. Second, demand will naturally differ for different types of books, allowing us to examine cross-sectional predictions of our model. Third, Amazon's incorporation of the AbeBooks listings presumably increased the number of searchers in the market, which provides us with an opportunity to also examine cross-time predictions of the model. Finally, we think the market is of interest both because its experience may be shared by many other markets for collectibles and similar items, and in its own right because it was one of the earliest markets to thrive on the Internet and because Amazon's entry raises questions of potential regulatory concern.

One set of books that we will include in our cross-sectional analysis is books that are of "local interest". Some examples in our sample are the short story collection, *Christmas in Georgia*, a history book by a Boston College professor emeritus, *Boston Catholics: A History of the Church and Its People*, and Indiana native Booth Tarkington's novel *The Turmoil.* We think these books could be interesting to analyze because physical used bookstores may have already been very well positioned to match such books with high willingness-to-pay owners, and, therefore, there may be little added benefit from listing many such books on the Internet. One instructive example is *The Mount Vernon Street Warrens: A Boston Story, 1860-1910.* Beacon Hill's narrow cobbled streets flanked by 19th century row houses attract tourists as well as well-heeled and history conscious homeowners. It seems plausbile that someone who sees this book on a used bookstore shelf shortly after having walked down Mount Vernon Street might decide that reading the book would be an enjoyable way to learn more about the area's history and have a high willingness to pay. Given that the book achieved limited commercial success–it has just a single review on Amazon–it seems likely that there are few consumers searching directly for it on the Internet. Moreover, even if the consumer in our example did find the book via an Internet search, the *Publishers Weekly* review,

> *The fortune of the Warren family ... enabled five siblings to grow up in the elite society of Boston's Beacon Hill in the early 1900s. In telling the stories of those children who became notable for eccentricity and philanthropy, Green ... focuses on Ned Warren, a homosexual and mover in the international movement of aestheticism, who was determined to lead a "grand but blighted life." ... This somewhat jumbled tale of a family's sundering through greed and suicide is enlivened with anecdotes ....*

might well have convinced him or her that some other book might make for lighter reading.

Another cross-sectional difference is that some books are much more popular than others. A critical feature of our opening anecdote about searching for a book on the pharmaceutical industry is the book's obscurity. If instead the book was Michael Dibdin's 1994 detective novel, *Dead Lagoon*, or Ron Chernow's 2004 best-selling inspiration for the Hamilton musical, *Alexander Hamilton*, the story would have ended differently. It would not be so unlikely to find the book at any random used bookstore and, futhermore, many, many copies would be available online, resulting in a completely different degree of competition for that title. Also, willingness to pay would be different – if the cheapest available used copy was still $20, the story might have ended instead with a recognition that the paperback was still in print and cheaply available on Amazon, or perhaps with a decision to instead choose among the many other used detective novels available for a dollar or two.

3.2 Testable implications of our model

Our model provides a very simple potential explanation for the facts we began with. That online prices are higher than offline prices can be explained as a consequence of the Internet having boosted customer flows. And online price dispersion can a consequence of the different stores' facing different consumer arrival rates (due to asymmetric prominence, etc.), with the lower tail also pulled down by the presence of a shopper population. But, of course, these facts could also be explained in other ways. In this section we discuss several more nuanced implications of the model that can be examined to help assess its relevance to the used book market.

To begin consider the comparison between offline and online price distributions. In the model we would think of this as comparing the light gray and the thick dashed red lines in the right panel of Figure 1. The offline world where prices differ only because different stores have different arrival rates would correspond to the light gray line. The online world where arrival rates are higher, but firms also are in competition for shoppers would correspond to the red line. Two differences between the curves reflect effects that should be found robustly across parameter values. First, we would expect to see more high prices online than offline because a more favorable nonshopper arrival process pushes prices up and the competition effect is negligible at the high end of the price distribution. Second, the online price distribution would be expected to be more dispersed than the offline price distribution because the competition effect pulls down prices at the low end of the distribution while having little effect on the upper tail. These effects are summarized in the following prediction:

Prediction 1: The online price distribution should be more dispersed and have more mass in the upper tail than the offline price distribution.

We noted above that we think there may be two important differences between standard used books and those that are of "local interest." Local interest books may already have high arrival rates of high-value consumers at physical bookstores. And listing such a book online may do little to boost the arrival rate of high willingness-to-pay consumers. This leads to our second prediction:

Prediction 2: Offline prices should be relatively high for local interest books. Offline prices for these books may have some similarities to online prices for standard titles. The offline-online price gap should be smaller for local interest titles than for standard titles.

We would put our comments about popular books in the context of our model in two ways. First, it may be that popular titles tend to be more substitutable across titles (think detective novels or romance novels). Even if this is not the case, the availability of other outside options (including in-print paperbacks), would result in the distribution of valuations for used copies of many popular books to have thinner upper tails than do the distributions for more unusual books. Recall, that the effect of an increase in the consumer arrival rate on prices is smaller when the distribution has a thinner upper tail. Second, popularity will imply that the books are available from a larger number of online sellers, which will make the competition effect more powerful. The combination leads to our third prediction:

Prediction 3: The online-offline price gap should be smaller for popular books. The upper tail of the online price distribution should be thinner.

Finally, recall that Amazon incorporated AbeBooks's listings under its "buy used" button between 2009 and 2012. In terms of our model, this could be captured by assuming that the size of the searcher population increased between the two dates. The right panel of Figure 1 illustrated the potential effect of a similar change, involving introducing a shopper population where there had not been one. Increasing the size of the shopper population should have a similar effect, pulling down the lower end of the price distribution (because firms with few nonshoppers focus on competing for the shoppers) but leaving the upper part of the distribution largely unchanged (because firms with many nonshoppers prefer selling to them at high prices to setting prices low enough to be competitive in the shopper submarket). This leads to our fourth prediction:

Prediction 4: The 2012 online price distribution should include more listings at very low prices than the 2009 online price distribution. The upper tails of the 2009 and 2012 distributions, however, should be similar.

4 Data

Our dataset construction began with a sample of books found at physical used book stores in the spring and summer of 2009. One of the authors and a research assistant visited three physical used book stores in the Boston area, one store in Atlanta, and one store in Lebanon, Indiana, and recorded information on a quasi-randomly selected set of titles. The information recorded was title, author, condition, type of binding, and the presence of any special attribute, such as author's signature.

We then collected online prices, shipping charges, and conditions for the same set of titles from www.AbeBooks.com at three points in time: first in the fall of 2009, then in

November of 2012, and again in January of 2013.[18] In most of our analyses we will use a price variable defined as listed price plus shipping charges minus two dollars to reflect the money received by the seller from the sale (with $2 being a rough estimate of the actual cost to the seller of shipping the book). The online collection was restricted to books with the same type of binding as the offline copy. For most of the titles the online data include the complete set of listings on www.AbeBooks.com.[19] But for some titles with a large number of listings we only collected a subset of the listings.[20]

Most of our analyses will be run on the set of 335 titles that satisfy three conditions: the copy found in a physical bookstore was not a signed copy, at least one online listing was found in 2009, and at least one online listing was found in November of 2012. We will often divide the sample into three subsamples which we refer to as "popular", "nonpopular local interest," and "standard" titles. We classify titles as popular if the number of copies found in our 2009 online search is at least fifty. Examples include Chernow's *Alexander Hamilton*, the detective novel *Dead Lagoon*, and Jeff Smith's cookbook *The Frugal Gourmet*. Paperback versions of the first two were still available new on Amazon in 2012 whereas the latter was out of print. We intentionally oversampled books of "local interest," which we define to include histories of a local area, novels set in the local area, and books by authors with local reputations. Most local interest books were drawn from shelves labeled as being of local interest in the bookstores, but some were swept up in our general collection and may be "displaced" local interest books that are of interest to some very distant location. For all local interest books we constructed a measure of distance between the locus of interest and the particular bookstore. For example, if a history of the state of Maine were being sold in a Cambridge, Massachusetts bookstore, the distance measure would take on the value of the number of miles between Cambridge and Maine's most populous city, Portland. Finally, we refer to books which are neither poplar nor of local interest as standard titles. Many of these books never achieved much commercial success and have been out of print for some time.

[18]The latter two data collections were primarily conducted on November 3, 2012 and January 5, 2013, respectively.

[19]By 2012, new copies of some formerly out-of-print books had again become available via print-on-demand technologies. We remove any listings for new print-on-demand copies from our 2012 and 2013 data.

[20]In the 2009 data collection, we collected every nth listing if a title had more than 100 listings, with n chosen so that the number of listings collected would be at least 50. In the 2012 and 2013 collections we collected all listings if a title had at most 300 listings, but otherwise just collected the 50 listings with the lowest shipping-inclusive prices plus every 5th or 10th listing thereafter.

4.1 Summary statistics on title-level data

Table 1 reports summary statistics describing offline and online prices.[21] The average offline price (in 2009) of the books in our sample is $11.29. In 2009 the average across the same titles of the lowest online price for each title is a little lower at $9.27. Statistics on the median and maximum online price give some sense for the magnitude of the price dispersion in the data. The average across titles of the within-title median price, $17.77, is almost twice as high as the average of the minimum prices, and the most expensive online copy is often very expensive. Statistics on the 2012 online price distributions are similar,with perhaps a little more dispersion. To give a sense of where each offline price would rank in the online price distribution for each title, we constructed the variable *PlaceInDistribution* giving the fraction of 2009 online listings for each title which are at prices below the price of the copy found in a physical store. The mean for this variable, 0.26, indicates that on average, 2009 offline copies were around the 26th percentile of the contemporaneous online price distribution.

The table gives a few additional statistics for two subsamples. 100 of the 355 titles are "standard" titles. Like the pharmaceutal book in our opening story, many are somewhat obscure. In the current online world, saavy consumers interested in any of them will usually have a number of options. On average we found 17.4 online listings in 2009 and 22.7 online listings in 2012. But given the number of bookstores in the US, these numbers suggest that it would be quite unlikely that one would come across any one of them in any particular online or offline bookstore. Our sample contains 158 nonpopular local interest titles. They are, if anything, slightly more obscure on average with averages of 14.5 online listings in 2009 and 17.5 online listings in 2012. Note that 20% of these books were found at a bookstore that is more than 100 miles from their location of interest. One statistic on the 77 popular titles is that the mean number of online listings in 2012 was 204.[22] The markets for these titles are quite different.

[21]To avoid having our statistical analyses influenced by a small number of listings at extremely high (and probably mostly unrealistic) prices, we drop all listings with prices above $500 in our comparisons of means, variance decompositions, regression analyses, and these summary statistics. In 2009 this drops 19 listings with prices ranging up to $8252. In 2012 it drops 34 listings with prices ranging up to $17,498. We do include these listings in the rightmost bar showing the number of listings priced above $200 when we present price histograms.

[22]Due to our data collection protocol we cannot provde a comparable statistic for 2009.

Variable	Mean	St Dev	Min	Max
Title-Level Variables Full Sample: 335 Titles				
OfflinePrice09	11.29	21.11	1.00	250.00
MinOnlinePrice09	9.27	22.34	1.89	351.50
MedOnlinePrice09	17.77	23.86	2.95	351.50
MaxOnlinePrice09	67.16	83.04	5.00	495.50
PlaceinDist09	0.26	0.28	0	1
MinOnlinePrice12	8.63	21.34	1.01	302.00
MedOnlinePrice12	17.64	25.46	1.95	302.00
MaxOnlinePrice12	78.77	94.07	2.05	498
Standard Subsample: 100 Titles				
NumberListed09	17.41	11.99	1	49
NumberListed12	21.15	29.52	1	270
Nonpopular Local Interest: 158 Titles				
NumberListed09	14.53	11.36	1	49
NumberListed12	17.49	34.33	1	380
Over100Miles	0.20	0.40	0	1
Popular Subsample: 77 Titles				
NumberListed12	204.35	269.35	41	1608
2012 Online Listing-Level Dataset 5284 Listings for 318 Titles				
Price12	13.90	36.06	1.01	1256
Disappear	0.15	0.36	0	1
Condition	3.97	1.12	1	7
StoreTitles	23.06	26.42	1	108

Table 1: Summary statistics

4.2 "Quantity" data

We have no sales data, but we did obtain a proxy for sales based on our gathering of data on multiple dates close together in 2012 and 2013: whether a listing in our November 2012 sample has disappeared by January of 2013.[23] Of course this measure is only a proxy and probably overstates true sales because listings can be removed for reasons other than a book being sold. For instance, the seller could have exited the AbeBooks platform or just stopped listing this particular copy. In the interest of interpreting this measure conservatively, we treat our *Disappear* variable as missing in several circumstances where we think it is particularly likely to have reliability problems.[24] The resulting dataset has a total of 5284 listings across 318 titles.

The bottom panel of table 1 provides summary statistics on this subset of our listing-level data. The average price of the listings in this sample is $13.90. The mean disappearance rate is 15.2%. This should be thought of as the sum of true sales over a two month period (which includes the Christmas season) for listings which are among the 50 lowest-priced listings for their title plus the rate at which such titles are removed without being sold. Listings can be described as being in one of seven conditions, but in practice this has limited variation: 85% of the listings are within one step of "very good".[25] About half of the roughly 2000 online retailers in our data have listings for just one or two of our 318 titles. They account for 18% of the listings in our dataset. At the other extreme, some sellers apparently have many, many listings: one has listings for 108 of the 318 titles in our dataset and fifteen others have listings for at least 50. Given that we selected titles pseudo-randomly, oversampled local interest titles, and have dropped the 20 titles with the largest number of 2012 listings, these sellers must have enormous numbers of titles available.

To investigate whether the disappearance data may provide a reasonable proxy for demand, we used the data to estimate standard logit demand models. The first column of

[23] Listings do not have a permanent identifier, so what we observe more precisely is whether the seller no longer lists a copy of the same title in the same condition.

[24] We treat the variable as missing in four circumstances. (1) If the listing was not unique by title-seller in the 2012 data. (2) If the title was one of the 20 for which we collected only a subset of the listings in 2012/2013. (3) If the seller listed copies of three or more of the titles in our dataset in 2012 and all of the seller's listings were removed before the 2013 collection. (Of 203 sellers listing exactly three titles only two removed two of their three listings but 14 removed all three so we presume most of the 14 are due to sellers' exiting the platform.) (4) Listings which were not among the 50 lowest priced listings for a title in the 2012 data collection. (We worried that some high price listings might be due to something other than sellers having a copy they wish to sell at the high price and thought other motives might overwhelm true sales at such noncompetitive prices.) The fact that we drop data for 7% of the sellers with exactly three listings could be thought of as an estimate of the background disappearance rate due to exit that might be expected among sellers with one or two listings. Sales can also be underestimated if a seller has multiple copies of a title in the same condition and only lists the second after selling the first.

[25] We code poor as 1, fair as 2, and so on. The modal "very good" is 4.

Table 2 presents estimates of a logit model in which the probability that a listing will disappear is a function of the listing's rank, price, condition, and some additional variables.[26] The coefficients on a listing's price and on its rank in a price-sorted list are both highly significant. Together they indicate that disappearance displays the strong price-sensitivity that one would expect of demand in this environment. For example, a seller that raised its price from \$4 to \$5 and fell from first to third on the list would see its probability of disappearance drop by over one third. The *log(Rank)* variable appears more important quantitatively than that price variable, which fits with a world where low-priced firms usually sell to price-sensitive shoppers while more the sales of higher-priced firms are to less price sensitve nonshoppers. The right panel reports similar estimates from a model with title-fixed effects, dropping titles for which no or all listings sell. The estimates are fairly similar with the price coefficient becoming somewhat larger and the rank coefficient somewhat smaller.

Holding price fixed, listings that are in better condition are significantly more likely to disappear. This also fits with expectations about demand. The maginitude of the condition effect is not very large compared to the rank effects. It indicates that a "good" condition copy ranked third would have about the same expected probability of disappearance as a "very good" copy ranked fourth. The median price difference between the third- and fourth-ranked listings is less than 50 cents. The two title-level controls, the lowest price at which any copy of the tile is offered and the number of listings for the title (which we have been taking as a proxy for popularity) both also have effects of the expected sign.

The final listing-level covariate, *log(StoreTitles)*, which reflects the number of our 318 titles for which the seller is listing at least one copy, does not have the effect one would expect in our fully rational model. By itself it might be expected to be a positive predictor of demand – sellers with more listings could be websites with greater nonshopper traffic which will sell more copies in equilibrium. But in our model price is supposed to be a sufficient statistic for a store's unobserved traffic level, in which case the variable would have no effect in a regression that includes price. Two potential reasons for a departure in this direction are that some small sellers might fail to live up to the standard of fully informed profit-maximizing behavior assumed by our model and not realize that they should be pricing relatively aggressively given their low arrival rates; or that some large sellers may set lower

[26] The regressions also included two unreported variables: a constant and an indicator for the condition variable being missing. We construct the rank variable by sorting listings for each title on shipping inclusive rank, but it need not correspond with how consumers would have seen the listings presented both because we do not know how AbeBooks broke ties and because consumers can choose to sort on something other than the default setting.

prices than the model implies because they have other (nonlisted) copies in inventory and hence have a lower opportunity cost of making a sale than the model assumes.

Overall, we take this look at disappearance patterns to support our hope that disappearance can provide a useful proxy for demand.

	Dependent Variable: *Disappear*			
Variable	Est.	St.Err	Est.	St.Err
log(Rank)	-0.43	(0.06)	-0.28	(0.09)
log(Price/MinPrice)	-0.54	(0.10)	-0.97	(0.14)
Condition	0.13	(0.04)	0.17	(0.04)
log(StoreTitles)	0.31	(0.04)	0.25	(0.04)
log(MinPrice)	-0.51	(0.19)		
log(TitleListings)	0.24	(0.07)		
Title fixed effects	No		Yes	
Number of obs.	5284		4332	
Pseudo R2	0.141		0.157	

The table reports coefficient estimates and standard errors from logit regressions with an indicator for whether a listing was removed between November 3, 2012 and January 5, 2013 as the dependent variable. The sample includes the (up to 50) lowest priced listings for 318 titles.

Table 2: Logit estimates of the listing disappearance process

5 Used Books: Facts and Tests of Model Predictions

The purpose of this section is two-fold: to present general descriptive evidence on prices of used books and to perform reduced form tests of some more subtle predictions of our models. The first subsection establishes the basic facts concerning price levels and price dispersion offline versus online. The next four subsections roughly correspond to the four sets of empirical predictions we enumerated earlier. The final subsection, a regression analysis of offline-online price differences, can be thought of as providing significance tests for some of the observations we make informally in the earlier subsections.

5.1 Basic facts

The introduction mentions a number of papers with somewhat counterintuitive findings about what happens to prices when markets move online. First, previous research has found that online markets continue to display substantial price dispersion. Second, previous research has found some evidence that price levels are lower online than offline, but perhaps not as much lower as some had expected. In this section we establish a similar dispersion fact for the online market for used books. And we establish a more concrete and extreme version of a stylized fact about price levels: online used book prices are higher on average than offline used book prices, sometimes much higher. These facts will resonate throughout the section, but we establish them first here in their most aggregated form.

The first row of Table 3 compares the mean of the offline prices with the average across titles of the mean online price for each title.[27] The fact that online prices are higher again comes through clearly: the average online price is about twice as high.

2009 Offline Prices		2009 Online Prices	
Average price	10.60	Average across titles of average price	20.30
		Average within-title coefficient of variation	0.72
Variance of prices (LHS)	281.14	Average of within-title variance (RHS Term 1)	621.84
		Variance of mean prices across titles (RHS Term 2)	670.81

Note: The table presents statistics on 2009 online and offline prices. All listings priced at \$500 or more are omitted and the set of titles is limited to those that had more than one online copy in 2009.

Table 3: Comparison of Price Dispersion Offline and Online

The second row quantifies the observation that there is a great deal of price dispersion online. The coefficient of variation of the set of prices at which a title is listed online averages 0.72, i.e. the standard deviation is on average 72% of the title's mean price.

We cannot directly compare offline and online price dispersion because our sample contains just a single offline price for each title.[28] Nonetheless, a simple calculation indicates that there must have been substantially less within-title price dispersion offline. Suppose that our offline prices, $p_1, p_2, \ldots, p_N$, are independent random draws from distributions $H_1, H_2, \ldots, H_N$ and that the distribution, H_i, of offline prices for title i has expectation μ_i and variance σ_i^2. If we set $\overline{p} = \frac{1}{N}\sum_{i=1}^{N} p_i$, then we would have a standard variance decomposition:

$$E\left(\frac{1}{N}\sum_{i=1}^{N}(p_i - \overline{p})^2\right) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^2 + \frac{1}{N}\sum_{i}(\mu_i - \mu)^2,$$

where $\mu = \frac{1}{N}\sum_{i=1}^{N}\mu_i$. The left column of the bottom row of the table indicates that the LHS of this expression is 281 in our 2009 offline sample. Hence, the first term on the RHS,

[27] For this calculation we omit nine titles for which we found only a single online price in 2009 and hence cannot compute a within-title standard deviation.

[28] Our data set has this structure out of necessity – many of our books are very obscure relative to the size of a typical physical used bookstore. Even if we visited hundreds of brick-and-mortar book stores and spent hours scanning the shelves of each of them, we probably would have failed to find another copy of many of our titles.

$\frac{1}{N}\sum_i \sigma_i^2$, must be smaller than 281 for offline listings and is probably substantially smaller than this given that the online data suggest that the titles are quite different from each other. The right column of the third row indicates that the average within-title variance in the offline data is over 600. Hence, there was substantially less price dispersion offline than online in 2009.

5.2 Offline and online prices in 2009: standard titles

In this section we examine offline and online prices for "standard" titles, which we define to be titles that have no particular local interest and are not offered by sufficiently many merchants to meet our threshold for being deemed "popular." We find patterns that are strikingly consistent with Prediction 1: the online price distribution has many more high-priced listings and online prices are much more dispersed.

We have 100 standard titles in our sample. Most are out of print. The mean number of 2009 online listings for these books was 17.4. Our basic fact about online prices being higher than offline prices is true to an extreme in the standard title subsample. The average offline price for the standard titles in our sample is $4.27. The average across titles of the average online price is 315% higher at $17.74.

Figure 3 provides a more detailed look at online vs. offline prices. The left panel contains the distribution of prices at which we found these titles at offline bookstores. Twenty of the books sell for less than $2.50. Another 74 are between $2.50 and $7.50. There is essentially no upper tail: only 6 of the 100 books are priced at $7.50 or more with the highest being just $20. The right panel presents a comparable histogram of online prices.[29] The upper tail of the online distribution is dramatically thicker: on average 27% of the listings are priced at $20 or higher including 6% at $50 or more.[30]

The contrast between the upper tails is consistent with Prediction 1: the upper tail of online price distribution should be thicker if online consumers arrive at a higher rate and/or have valuations drawn from a higher distribution. The other part of prediction 1 – that online prices should be more dispersed – is also very clear in the comparison. A comparison of the left tails of the price distribution suggests that the competition effect is not strong enough to make the online distribution have a thicker lower tail.

To provide a clearer picture of the lower-tail comparison, the left panel of figure 4

[29]To keep the sample composition the same, the figure presents an unweighted average across titles of histograms of the prices at which each title is offered.

[30]To show the full extent of the distribution we have added three extra categories – $50-$100, $100-$200, and over $200 – at the right side of the histogram. The apparent bump in the distribution is a consequence of the different scaling.

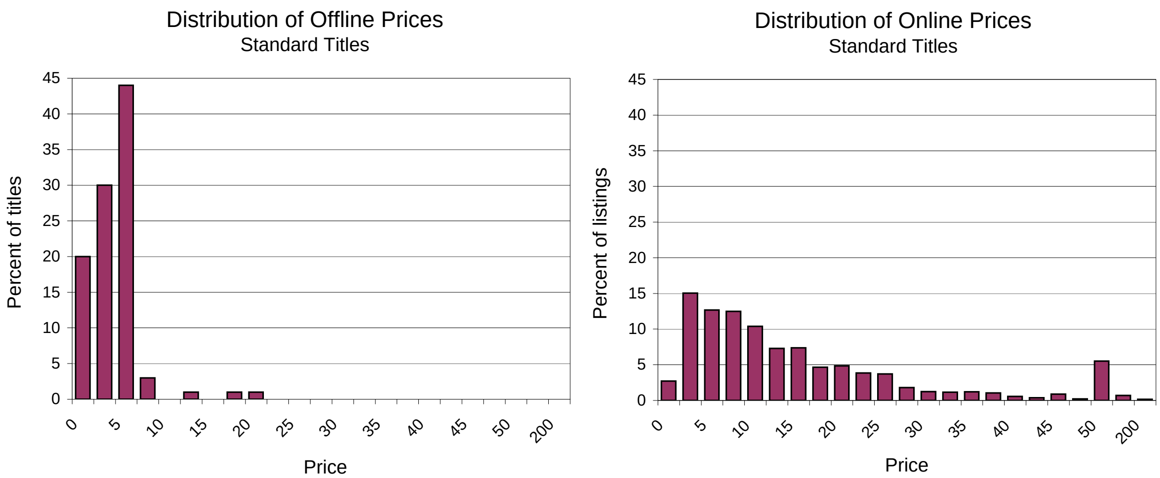

Figure 3: Offline and online prices for standard titles in 2009

presents a histogram of the *PlaceInDist* variable. (Recall that this variable is defined as the fraction of online prices that are below the offline price for each title.) The most striking feature is a very large mass point at 0: for 54% of the titles, the price at which the book was found in a physical bookstore was lower than every single online price! (This occurs despite the fact that we had found on average 15.3 online prices for these 54 titles.) Beyond this, the pattern looks roughly like another quarter of offline books are offered at a price around the 20th percentile of the online price distribution and the remaining 20% spread fairly evenly over the the upper 70 percentiles of the online distribution. Overall, the patterns suggest that the match quality effect is much more important than the competition effect for these titles.

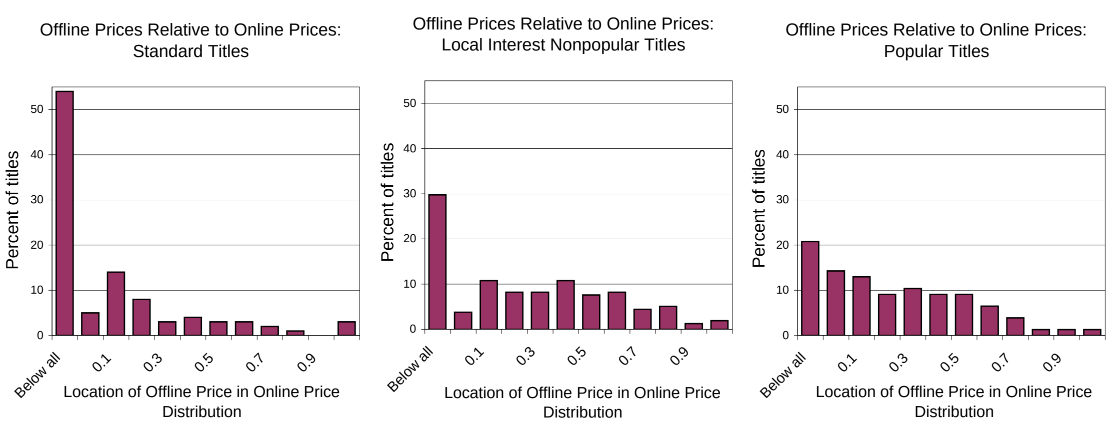

Figure 4: Offline prices relative to online prices for the same title

5.3 Offline and online prices for local interest titles in 2009

In this section we examine price distributions for local interest books. We note that the offline prices for these titles look very different from those of other types of books and are very much in line with our Prediction 2.

Recall that Prediction 2 was that offline prices for local interest books should be relatively high and that differences between the offline and online price distributions should not be as pronounced. We argued that the highest-value match for a title like *The Mount Vernon Street Warrens: A Boston Story, 1860-1910* might well be a tourist who has just walked into a Boston used bookstore looking for something to read that evening. Consistent with this presumption, we will show here that offline prices for local interest titles look more like the online prices we saw in the previous section. Accordingly, offline stores may already have been achieving levels of match quality comparable to what the Internet makes possible for standard books, and the incremental gains from also listing books on the Internet may not be so large.

Our sample contains 158 titles which we classified as being of "local interest" and which did not meet our threshold for being labeled as "popular." The mean offline price for these titles is \$18.86. Average online prices are again higher, but the proportional increase in prices is much smaller: the mean across titles of the mean online price is \$27.82. A nearly 50% increase is substantial, but much less than a 315% increase. Figure 5 provides striking visual evidence of the predicted patterns. The distribution of offline prices in the left panel looks very different from the the corresponding figure for standard titles (in the left panel of Figure 3). Instead, the distribution shares several features with the distribution of online prices for standard titles: the largest number of prices fall in the \$7.50-\$9.99 bin; and there is a substantial upper tail of prices including 26 books with prices from \$20 to \$49.99, and nine books with prices above \$50. Comparing the left and right panels we see that the distribution of online prices does again have a thicker upper tail, but, consistent with Prediction 2, the online-offline difference is not nearly as large. The online distribution also has a slightly higher percentage of listings at prices below \$5, but there is nothing to suggest that the competition effect is very strong.

The middle panel of figure 4 includes a histogram showing where in the online price distribution for each title the offline copy falls. Here we see that about 30% of the offline copies are cheaper than any online copy. For the other 70% of titles, the offline prices look a lot like random draws from the online distribution, although the highest prices are a bit underrepresented.

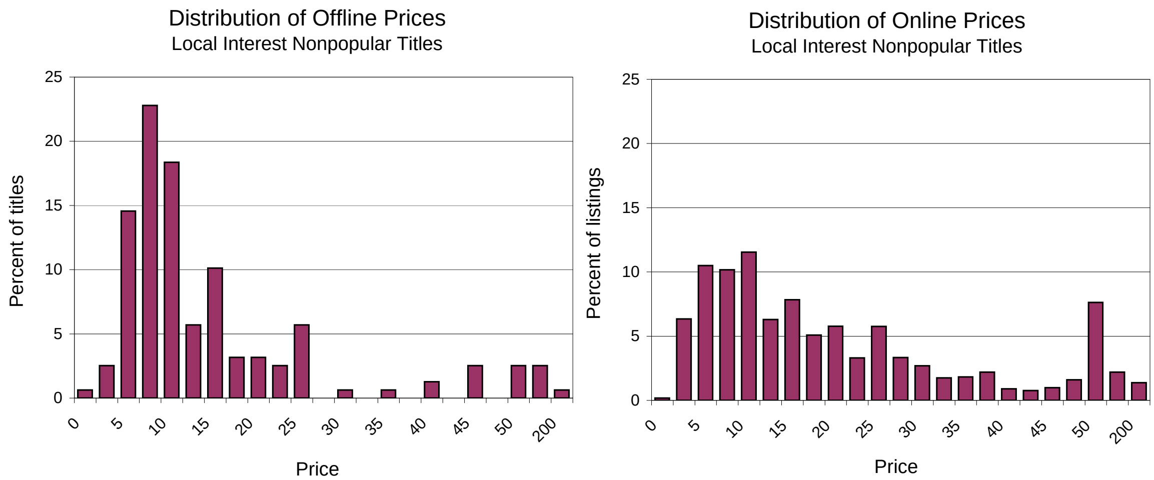

Figure 5: Offline and online prices for nonpopular local interest titles in 2009

5.4 Offline and online prices for popular titles in 2009

We now turn to the final subsample, popular books. Again, we will show that online-offline differences and comparisons to the earlier data on standard titles generally appear consistent with a match-quality model, Prediction 3, in particular.

Recall that we labeled 77 books as "popular" on the basis of there being at least 50 copies offered through AbeBooks. Our prior was that two differences between these books and standard titles would be most salient. First, the greater number of shoppers (and sellers) might make the competition effect more important. Second, the distribution of consumer valuations might have less of an upper tail because consumers sometimes have the option of simply buying a new copy of the book in paperback (and may, to the extent that these titles are more substitutable across title, be quite willing to switch from one detective novel to another). Prediction 3 was that as a result the growth of the upper tail in an offline-online comparison, which was so dramatic for standard titles, should be less pronounced here.

The left panel of Figure 6 shows that popular book prices are fairly similar to standard book prices at offline bookstores: the mean price is \$4.89 (whereas the mean price for standard titles was \$4.27). Fourteen percent of these books are selling for below \$2.50 with the vast majority (70%) being between \$2.50-\$7.49. None is priced above \$18. Comparing the offline prices with the online prices shown in the right panel of Figure 6 we once again note that our most basic prediction about how offline and online prices should compare (Prediction 1) is again validated: online prices for popular books are more dispersed and have a thicker upper tail than offline prices.

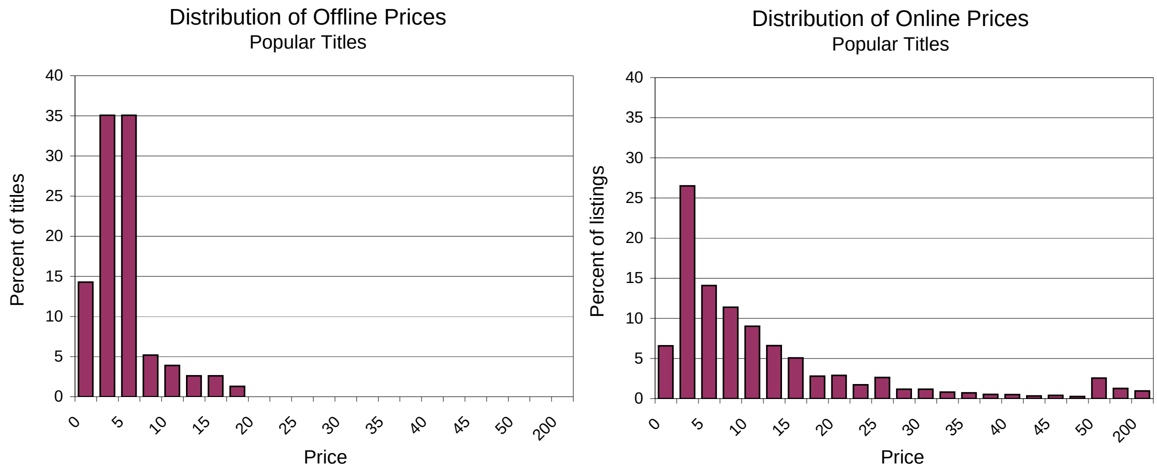

Figure 6: Offline and online prices for standard titles in 2009

Comparing this figure to the corresponding figure for standard titles (Figure 3) provides support for Prediction 3. The upper tail of the online distribution is less thick than that for standard titles: on average 18% of listings are priced above $20 whereas the comparable figure for standard titles was 27%. And the online-offline difference is less dramatic.

One other difference between popular and standard titles is that the online distribution for popular books has a larger concentration of low prices: about one-third of the listings are priced below $5. The more pronounced lower tail is consistent with the hypothesis that the competition effect may be more powerful for these titles.

The right panel of figure 4 shows that for about 20% of titles, the offline price we found was below *all* online prices. This number is smaller for popular books then for the other two types, but it is still a strikingly large number, given that each title had at least 50 online listings. Meanwhile the remaining prices look like they are mostly drawn from the bottom two-thirds of the online price distribution for the corresponding title. A comparison of the left and right panels provides another illustration of the data's consistency with prediction 3: the online-offline price gap is narrower here than it was for standard titles.

5.5 Online prices: 2009 and 2012

In this section we compare online prices from 2009 and 2012 and note changes in the price distributions in line with Prediction 4.

Recall that prediction 4 was that the 2012 online price distributions may have more low-priced listings than 2009 distributions, but a similar upper tail. The reason for the prediction is that AbeBooks' integration into Amazon, which may have substantially increased the number of shoppers who viewed AbeBooks' listings, occurred between 2009 and 2012. In our theory section, specifically the right panel of Figure 1, we had noted that introducing shoppers pulls down prices at the lower end of the distribution and may lead several firms to price below the former lower bound of the price distribution. But in the upper part of the distribution it should have almost no impact (as firms setting high prices mostly ignore the shopper segment).

The upper left panel of figure 7 illustrates how prices of standard titles changed between 2009 and 2012. The gray histogram is the histogram of 2009 prices we saw previously in Figure 3. The outlined bars superimposed on top of this distribution are a corresponding histogram of prices from November 2012. At the low end of the distribution we see a striking change in the distribution of the predicted type: there is a dramatic increase in the proportion of listings below \$2.50. Meanwhile (and perhaps even more striking), the upper tail of the distribution appears to have changed hardly at all. We find this consistency somewhat amazing given that there is a three-year gap between the collection of these two data sets. The other two histograms in the figure illustrate the changes in the price distributions for local interest and popular books. In each case we again see an increase in the proportion of listings priced below \$2.50. The absolute increase is a bit smaller in the local interest case, although it is large in percentage terms given that almost no local interest books were listed at such a low price in 2009. In both cases we also again see little change in the upper part of the distribution. This observation is particularly true for the popular histogram in which almost all of the growth in prices below \$2.50 seems to come out of the \$2.50-to-\$5 bin. We conclude that the pattern of the lower tail having been pulled down while the upper part of the distribution changes less is consistent across the different sets of titles. This is very much in line with what we would expect if Amazon's integration of used book listings increased the size of the shopper population.

5.6 Regression analysis of offline-online price differences

In the preceding sections we used a set of figures to illustrate the online-offline price gap for standard, popular, and local interest books and noted apparent differences across the different groups of books. In this section we verify the significance of some of these patterns by regressing the $PlaceInDist$ variable on book characteristics.

The first column of Table 4 presents coefficient estimates from an OLS regression. The second column presents estimates from a Tobit regression which treats values of zero and

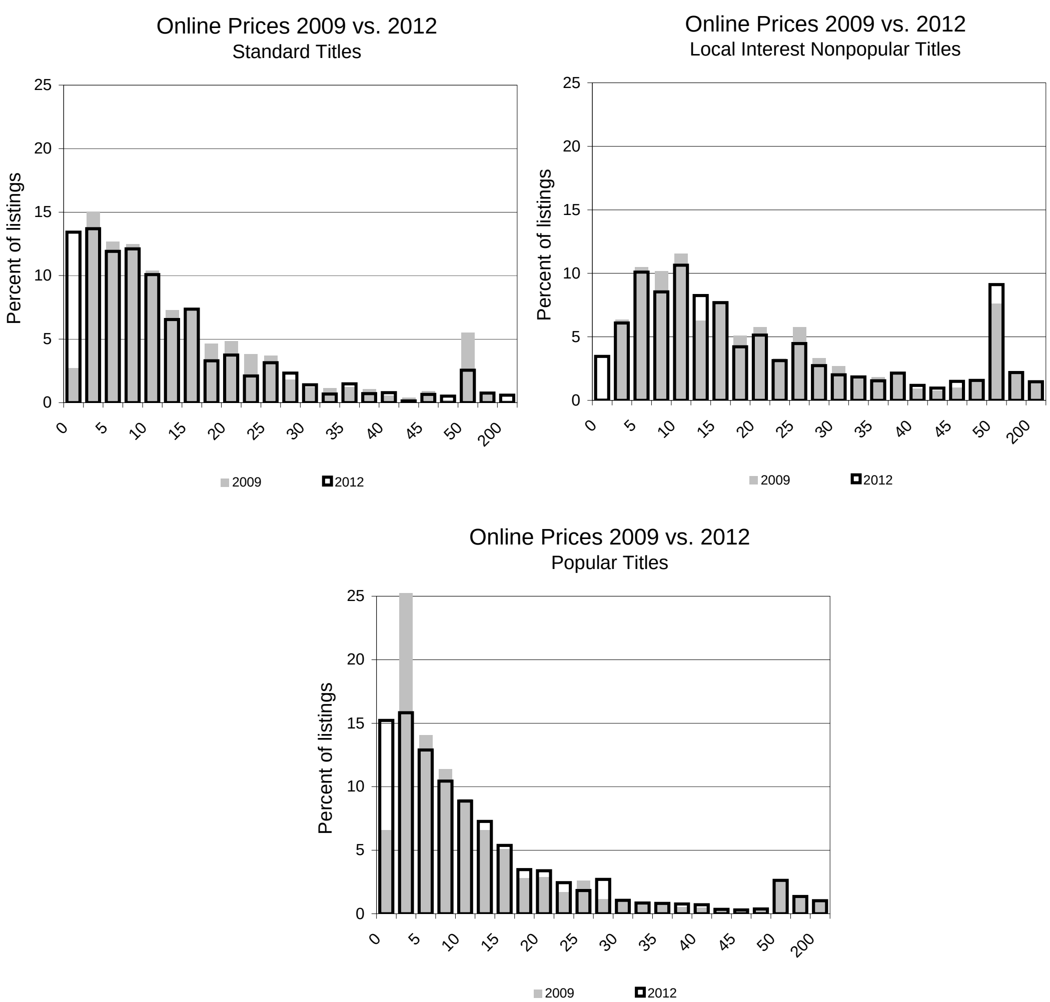

Figure 7: Comparison of 2009 and 2012 online prices

one as censored observations. With regard to our prediction that the Internet may not have as large of an effect on the upper tail of the price distribution for popular books, the coefficient estimate of 0.11 in the first column indicates that offline prices are indeed higher in the online price distribution for popular titles. The estimate of the coefficient on *Popular* from the Tobit model is larger at 0.21, i.e. offline prices are 21 percentiles higher in the online distribution for popular books on average, and even more highly significant.

	Dependent variable: *PlaceInDist*			
	OLS		Tobit	
Variable	Coef.	SE	Coef.	SE
Popular	0.11	(0.04)	0.21	(0.06)
$LocalInt \times Close$	0.17	(0.04)	0.27	(0.06)
$LocalInt \times Far$	0.08	(0.05)	0.17	(0.08)
Constant	0.16	(0.03)	-0.01	(0.04)
Num. Obs.	335		335	
R^2 (or pseudo R^2)	0.07		0.06	

Table 4: Variation in offline-online prices with book characteristics

Local interest books located in physical bookstores close to their area of interest may have both a relatively high arrival rate of interested consumers and a relatively high distribution of consumer valuations. Again, this should lead to relatively small online-offline price gap, i.e. to relatively high offline prices. The 0.17 coefficient estimate on the $LocalInterest \times Close$ variable indicates that this is true for local interest books in used bookstores within 100 miles of the location of interest. The Tobit estimate, 0.27, is again larger and more highly significant.

One would not expect misplaced local interest books to benefit in the same way. Here, the regression results are less in line with the model. In the OLS estimation the coefficient on $LocalInterest \times Far$ is about half of the coefficient on $LocalInterest \times Close$, and the standard error is such that we can neither reject that the effect is zero, nor that it is as large as that for local interest books sold close to their area of interest. In the Tobit model, however, the estimate is a bit more than 60% of the size of the estimated coefficient on $LocalInterest \times Close$ and is significant at the 5% level. This suggests that a portion of the differences between local interest and other books noted earlier may be due to other unobserved book characteristics.

6 Welfare Estimates from a Structural Model

In our model the welfare effects of online sales depend on the various primitives and the induced equilibrium effects: the change in consumer arrival rates; the change in valuation distributions; the equilibrium delay before sales occur; etc. In this section we note that features of our theoretical model make it amenable to structural estimation and develop a version that can be estimated with our limited dataset. Our estimates suggest that Internet sales led to substantial increases in both profits and consumer surplus.

6.1 A structural framework

In this section we discuss an empirical model closely related to the theoretical model of section 2 and note that aspects of the model facilitate estimation.

Consider a model similar to that of section 2 in which $I_k + 1$ populations of consumers shop for title k at stores $i = 1, 2, \ldots, I_k$. One of these is a population of shoppers who arrive at Poisson rate γ_{0k}. Shoppers observe all prices and purchase from store i at the instant at which they arrive with probability $D_k^o(p_{ik}, p_{-ik}; X_{ik}, \Lambda, \xi_{ik})$, where X_{ik} is a vector of store-title characteristics, Λ is a vector of parameters to be estimated, and ξ_{ik} is a vector of unobserved characteristics that may vary at the title or title-store level. Assume that the arrival rate γ_{0k} and the unobserved ξ_{ik} are draws from distributions that may depend on the parameter vector Λ.

The other I_k populations are nonshoppers who do not compare prices across sellers: nonshoppers from population i arrive at store i at Poisson rate γ_{ik}. Suppose that they purchase upon arrival with probability $D_{ik}^m(p_{ik}; X_{ik}, \Lambda, \xi_{ik})$. The γ_{ik} and ξ_{ik} are unobserved random variables with a distribution that may depend on Λ.

Assume that stores choose the prices that would maximize expected profits in a stationary dynamic model like that of section 2, i.e. assume that p_{ik} is chosen to maximize

$$\pi_i(p_{ik}, p_{-ik}) \equiv \frac{p_{ik}(\gamma_{ik} D_k^m(p_i) + \gamma_{0k} D_k^o(p_{ik}, p_{-ik}))}{r + \gamma_{ik} D_k^m(p_{ik}) + \gamma_{0k} D_k^o(p_{ik}, p_{-ik})},$$

where we have omitted the characteristics and parameters from the arguments for readability.

Suppose that we are given data on a set of titles $k = 1, 2, ..., \ldots, K$. These data will take two distinct forms. For some titles we observe just the vector of prices $(p_{1k}, p_{2k}, \ldots, p_{I_k k})$. For other titles we observe both prices and an indicator for whether the title sells in store i in a given time period: $(p_{1k}, \ldots, p_{I_k k}, q_{1k}, \ldots, q_{I_k k})$. We wish to estimate the parameter vector Λ.

One feature of this model that facilitates estimation is that the first order condition for store i's title k price to be optimal,

$$\begin{aligned} 0 &= rp_{ik}\gamma_{ik}D^{m\prime}(p_{ik}) + r\gamma_{ik}D^m(p_{ik}) + \gamma_{ik}^2 D^m(p_{ik})^2 \\ &\quad + rp_{ik}\gamma_{0k}\frac{\partial D^o}{\partial p_i}(p_{ik}, p_{-ik}) + r\gamma_{0k}D^o(p_{ik}, p_{-ik}) + \gamma_{0k}^2 D^o(p_{ik}, p_{-ik})^2 \\ &\quad + 2\gamma_{0k}\gamma_{ik}D^m(p_{ik})D^o(p_{ik}, p_{-ik}), \end{aligned}$$

is a quadratic function of γ_{ik} once one fixes γ_{0k}, the parameters affecting $D^m(p_{ik})$ and $D^o(p_{ik}, p_{-ik})$, and values for the random coefficients. Specifically, this FOC is of the form $a\gamma_{ik}^2 + b\gamma_{ik} + c = 0$ for

$$\begin{aligned} a(p_k, X_{ik}; \Lambda, \xi_{ik}) &= D^m(p_{ik})^2 \\ b(p_k, X_{ik}; \Lambda, \xi_{ik}) &= rp_{ik}D^{m\prime}(p_{ik}) + rD^m(p_{ik}) + 2\gamma_{0k}D^m(p_{ik})D^o(p_{ik}, p_{-ik}) \\ c(p_k, X_{ik}; \Lambda, \xi_{ik}) &= rp_{ik}\gamma_{0k}\frac{\partial D^o}{\partial p_i}(p_{ik}, p_{-ik}) + r\gamma_{0k}D^o(p_{ik}, p_{-ik}) + \gamma_{0k}^2 D^o(p_{ik}, p_{-ik})^2 \end{aligned}$$

Under some conditions ($b > 0$, $c < 0$), only the larger root of this quadratic will be positive. When this occurs, we can calculate the conditional likelihood of each price observation p_{ik} (conditional on the parameters, X_{ik}, and random coefficients) by backing out the unique γ_{ik} which makes p_{ik} optimal and then computing the likelihood via

$$L(p_{ik}|\gamma_{0k}, X_{ik}, \Lambda, \xi_{ik}) = L(\gamma_{ik}|\gamma_{0k}, X_{ik}, \Lambda, \xi_{ik})\frac{1}{\frac{\partial g}{\partial \gamma}(\gamma_{ik})},$$

where g is the best-response pricing function with $g(\gamma_{ik}) = p_{ik}$. By implicitly differentiating the FOC we find that

$$\frac{\partial g}{\partial \gamma_{ik}}(\gamma_{ik}) = -\frac{2\gamma_{ik}a(p_{ik}) + b(p_{ik})}{\frac{\partial a}{\partial p_{ik}}\gamma_{ik}^2 + \frac{\partial b}{\partial p_{ik}}\gamma_{ik} + \frac{\partial c}{\partial p_{ik}}}.$$

Another aspect of our model that facilitates estimation is that the one-to-one correspondence between the observed p_{ik} and unobserved γ_{ik} also makes it easy to account for endogeneity when using the demand data.[31] Given the observed p_{ik} and an inferred γ_{ik}, the arrival rate of consumers who would buy book k from store i is

$$d_{ik} \equiv \gamma_{0k}D^o(p_{ik}, p_{-ik}) + \gamma_{ik}D^m(p_{ik})$$

Hence the probability that the book will be sold in a Δt time interval is

$$E(q_{ik}|p_{ik}, p_{-ik}, \gamma_{0k}, \Lambda, \xi_{ik}) = 1 - e^{d_{ik}\Delta t}.$$

[31] Normally, one would need an instrument for p_{ik} in a logit demand model because of its correlation with the unobserved product attributes. Here, we know the relationship between p_{ik} and the single unobservable γ_{ik}.

The joint likelihood of observed pairs (p_{ik}, q_{ik}) is simply the product of this expression and our earlier expression for the likelihood of p_{ik}.

Together these two observations suggest a simple procedure for simulated maximum likelihood estimation. Given any potential parameter vector Λ, we take random draws for the unobservables γ_{0k} and ξ_{ik}. Conditional on each random draw, we compute the joint likelihood of each observed price vector $(p_{1k}, \ldots, p_{I_k k})$ and of each observed price/quantity vector $(p_{1k}, \ldots, p_{I_k k}, q_{1k}, \ldots, q_{I_k k})$ using the above formulae. Summing across the draws of the random coefficients gives the unconditional likelihood. Parameter estimates are obtained by maximizing this likelihood over the parameter space.

6.2 Empirical specification

Given the limitations of our dataset, we implement a parsimonious version of the model with a limited number of parameters and random unobservables. The model can match the main features of the dataset and includes coefficients that give it some flexibility in estimating most (but not all) of the underlying effects that matter most for welfare.

The rate γ_{ikt} at which nonshoppers interested in title k arrive at store i in year t is allowed to vary systematically with the title's popularity, the year, whether store i is online or offline, and whether the title is of local interest (for offline stores only). We assume that it also varies idiosyncratically across store-titles – a store's price is an increasing function of the rate at which it is visited by nonshoppers, and it is through the random variation in γ_{ik} that the model can account for each observed price as a best response. Formally, we assume that the arrival rate is

$$\gamma_{ikt} = \overline{\gamma}^{ns}_{g(ikt)} \frac{Pop_{kt}^{\gamma^{ns}_{Pop}}}{\sum_\ell Pop_{\ell t}^{\gamma^{ns}_{Pop}}} \tilde{z}_{ikt},$$

where the first term in this expression is a set of estimated fixed effects giving the mean arrival rates for a group $g(ikt)$ of store titles, the second term allows for the possibility that the arrival rate at each online store may vary with the popularity of the title, and $\tilde{z}_{ikt}$ are exponential random variables with mean 1 assumed to be drawn *i.i.d.* across store-titles to create the heterogeneity in store-specific arrival rates. The fixed effects allow the mean rate to vary depending on whether ikt belongs to one of four groups: 2009 offline standard, 2009 offline local interest, 2009 online, and 2012 online. The popularity term is intended to reflect that there may be more people searching online for more popular titles, but that the number of nonshoppers who end up at any one online store may also be reduced when there are more stores offering the title that any given consumer might find via a Google search or other means. We include it only for online stores, and take Pop_{kt} equal to the number of online listings for title k in year t in our sample.

We assume that each nonshopper's demand curves has the constant elasticity form. Specifically, we assume that

$$D^m_{kt}(p_{ikt}) = \delta_k p_{ikt}^{-\eta},$$

where η is the price elasticity of demand which will be estimated and δ_k is an unobserved demand shifter which will allow the model to account for heterogeneity across titles in average price levels and in the fraction of listings which are sold. We assume the δ_k are independent draws from a lognormal distribution with mean 1 and a variance σ_δ which is to be estimated. In utility terms, our demand specification is equivalent to assuming that a nonshopper j considering buying book k from store i gets utility

$$u_{ikj} = \begin{cases} v_{jk} - p_{ik} & \text{if he buys} \\ 0 & \text{if he does not} \end{cases},$$

where the v_{jk} are heterogeneous across consumers and for title k are independent draws from the density $f_k(v) = \delta_k \eta v^{-\eta-1}$ on $[\delta_k^{1/\eta}, \infty]$. [32]

We similarly assume that the rate γ_{0kt} with which shoppers interested in title k arrive in year t and see all available online prices is the product of a year-specific fixed effect, a popularity effect, and a title-level unobservable $\tilde{w}_{kt}$:

$$\gamma_{0kt} = \overline{\gamma}^s_t \frac{Pop_{kt}^{\gamma^s_{Pop}}}{\sum_\ell Pop_{\ell t}^{\gamma^s_{Pop}}} \tilde{w}_{kt}.$$

Here, we estimate just two fixed effects, one for 2009 and one for 2012, and the coefficient related to the effect of a title's popularity on the searcher rate, γ^s_{Pop}, is unambiguously expected to be positive because γ_{0kt} is the total arrival rate of shoppers, not a per-listing arrival rate. We assume that the $\tilde{w}_{kt}$ are gamma-distributed random variables with mean 1 and shape parameter γ^s_{shape}.

We assume that shoppers have pure logit demands that do not vary over time:

$$D^o_{kt}(p_{ikt}; p_{-ikt}) = \frac{e^{-\alpha\delta_k^c p_{ikt}}}{e^{-\alpha\delta_k^c p_{0kt}} + \sum_\ell e^{-\alpha\delta_k p_{\ell kt}}}.$$

Here, α is a parameter to be estimated that will reflect consumer price senstivity when choosing among listings for a title, c is a coefficient to be estimated which allows the

[32]Note that we must have $\eta > 1$ for the monopoly price not to be infinite, and that the density justification for the demand curve is only technically valid if the $\delta_k > 0$ is such that that all observed prices are in the support of the value distribution.

cross-listing price sensitivity to be higher or lower for titles that have higher nonshopper demand. We parsimoniously introduce an outside option that is equally relevant across titles by assuming that its "price," p_{0kt}, is Δp^{out} higher than the lowest price at which title k is available online in year t and estimating a single parameter Δp^{out}. Note that in this part of the specification we are not flexibly estimating an effect that is important to welfare: the degree to which shoppers are willing to substitute between inside and outside goods. We would like to have estimated a nested logit or some other such model, but found that we did not have the power to estimate the relevant parameter with any precision. Given this, we have chosen to just estimate the simple logit model, and then to make an ad hoc adjustment when we compute welfare.[33] Another limitation, driven by the fact that we only have demand data for 2012 and not for 2009, is that we are assuming that the shopper price sensitivity is the same in both years.[34]

The model as described above is (as are many structural models) somewhat heroic in its reliance on firm optimization. In our case this would result in the model having trouble fitting two types of behavior: it is hard to rationalize why any firm would list a price that is much below the second-lowest price at which the same title is offered; and extremely high prices can only be rationalized by the firms' having unreasonably high customer arrival rate. As noted earlier, our reduced form sales regressions also suggest that either small or large sellers are not always pricing in a manner that is fully consistent with our model assumptions. Accordingly, we allow for some degree of firm irrationality by assuming that each firm chooses the optimal price with probability $1 - R^p$ and with the complementary probability R^p instead chooses its price as an i.i.d. random draw from a Gamma distribution with mean μ^R and shape parameter s^R. R^p, μ^R, and s^R are additional parameters to be estimated.

We also remarked earlier that some books that disappear in between our 2012 and 2013 data collections disappear because they were sold, but others will disappear for other reasons, e.g. they were listed by small firms that withdrew from the AbeBooks platform or had their listing taken down for other reasons. To account for this, we assume that there is a constant background disappearance rate R^q so that the total probability that a listing will disappear in between the two periods is $R^q + (1 - R^q)q_{ikt}$, where q_{ikt} is the disappearance probability calculated from the arrival/demand model. The background disappearance rate

[33] As described below we assume that shopper valuations for inside versus outside goods have the same distribution as that we have estimated for nonshoppers.

[34] We also do not include book condition effects in the demand specification. Our reduced-form regressions suggest that they are not large in magnitude, and in preliminary investigations we found that we could not estimate these with any precision within our structural model.

R^q is another parameter to be estimated.

This specification yields a model with eighteen parameters to be estimated: $(\overline{\gamma}^{ns}_{09,of,std}, \overline{\gamma}^{ns}_{09,of,local}, \overline{\gamma}^{ns}_{09,on}, \overline{\gamma}^{ns}_{12,on}, \gamma^{ns}_{Pop}, \eta, \sigma_\delta, \overline{\gamma}^{s}_{09}, \overline{\gamma}^{s}_{12}, \gamma^{s}_{Pop}, \gamma^{s}_{shape}, \alpha, c, \Delta p^{out}, R^p, \mu^R, s^R, R^q)$.

6.3 Estimates of arrival rates and demand

We estimate the model described above on a dataset containing 236 books which include almost all titles in our "standard" and "nonpopular local interest" subsamples.[35] We do not include the "popular" titles because we worry that our quantity data is not sufficiently reliable for many of them and that our model assumptions are not as appropriate.[36] The estimation roughly follows the simulated maximum likelihood procedure described above with the integration over the unobservables dealt with via importance sampling as in Ackerberg (2009). Standard errors were obtained via a bootstrap approach. Appendix B provides more details.

While the structural model is stylized, some of the estimated coefficients will provide insights on the nature of demand in the market. Table 5 presents the complete set of coefficient estimates, with coefficients related to the nonshopper arrival processes in the top panel, followed by coefficients related to the shopper arrival process, followed by those related to the departures we have allowed from full rationality.

Our most basic observation about the arrival rates at offline bookstores is that they are quite low. The estimated arrival rate for a standard title at an offline bookstore is 0.65 customers per year. Given the estimated demand function $D(p) = p^{-1.87}$, each potential customer who "arrives" would only be willing to buy the book even at a price of \$5 with probability $5^{-1.87} \approx 0.05$, so the arrival rate of customers willing to pay \$5 or more is just $0.65 \times 0.05 \approx 0.03$. This calculation suggests that copies of standard titles will typically sit on shelves of offline bookstores for quite a number of years before being purchased. The estimated parameters imply that the probability of selling a standard title within a year at \$10 is less than 1 percent. This is consistent with the model needing to rationalize that almost all standard titles at offline bookstores are priced below \$10. Arrival rates for local interest titles are estimated to be about 90% higher than those for standard titles. We also

[35]Within these subsamples we include all titles that had at least two listings priced at less than \$50 successfully scraped from AbeBooks.com in the 2009 and 2012 online collections.

[36]One difficulty is that some sellers will own multiple copies of highly popular books. We will miss sales if such a seller makes a sale and subsequently relists another copy of the same title. It would also make sellers' opportunity costs different from what is assumed in the model. Also, in 2012/2013 we only collected a subset of listings when the number of listings was very large. For such titles we will not know if a seller who is seen in the 2012 collection but not in the 2013 collection removed a listing or simply was not included in the second sample.

think these estimates are in line with our casual empiricism: books, especially unusual or niche titles, do sometimes languish on the shelves of used book stores for years.

Estimated nonshopper arrival rates at online bookstores are substantially higher. In 2009 the estimated arrival rate at online stores is estimated to be over twenty times the contemporaneous arrival rate at offline stores for standard titles. Nonshopper arrival rates are then estimated to have declined by about one half between 2009 and 2012. Increased use of online comparison tools following Amazon's incorporation of AbeBooks could be one factor that has reduced nonshopper arrival rates. Another factor that might lead to a decline in nonshopper arrival rates on a per-seller basis is that the nonshoppers who are in the market are being divided among a larger number of sellers: the number of listings in the 2012 dataset is 13% larger.

Shopper arrival rates are substantial and appear to have increased substantially between 2009 and 2012. The estimates are estimates are that the shopper arrival rate was 5.6 per year in 2009 and 14.9 per year in 2012. The estimated shopper demand function is such that most shoppers who "arrive" will purchase a book from some seller. Hence, the parameter estimates can be thought of as saying that the shopper population contributes substantially to aggregate sales totals. The estimated price coefficient for the shoppers suggests that consumer choice at this price comparison website is very price sensitive.

Note that one way in which the model is rationalizing the upper tail of prices in the online data is by assuming that firms are choosing prices randomly with probability about one-quarter, and that randomly-chosen prices chosen are drawn from a fairly diffuse distribution with a mean of 15.3 and a standard deviation of 11.6 ($\approx 15.25/\sqrt{1.73}$). As a result, the model is much more conservative in its estimates of online welfare than it would have been if we forced it to maintain the assumption that all sellers were pricing rationally as in the theoretical model, which would have implied that high-priced firms must have very high arrival rates. Intuitively, a feature of the data that helps drive the nonrational pricing estimates is the contrast between estimated disappearance rates for high-priced listings and the disappearance rates that would be expected under the arrival rates that are needed to rationalize the prices.

6.4 Estimating welfare

In this section we discuss how we estimate welfare given the structural model estimates.

Given an estimated parameter vector $\hat{\Lambda}$, the average per-listing welfare generated by the listings for title k in year t can be calculated by integrating over the posterior distribution

Parameter	Est.	SE
Nonshopper arrival and demand		
Mean arrival 2009 offline standard title ($\overline{\gamma}^{ns}_{09,of,std}$)	0.65	(0.17)
Mean arrival 2009 offline local interest ($\overline{\gamma}^{ns}_{09,of,local}$)	1.25	(0.37)
Mean arrival 2009 online ($\overline{\gamma}^{ns}_{09,on}$)	14.90	(3.61)
Mean arrival 2012 online ($\overline{\gamma}^{ns}_{12,on}$)	7.95	(2.00)
Popularity effect on arrival (γ^{ns}_{Pop})	-1.36	(0.19)
Nonshopper price elasticity (η)	1.87	(0.08)
StDev of title-level unobservable (σ_δ)	1.16	(0.05)
Shopper arrival and demand		
Mean arrival 2009 online ($\overline{\gamma}^{s}_{09}$)	5.65	(1.41)
Mean arrival 2012 online ($\overline{\gamma}^{s}_{12}$)	14.86	(3.10)
Popularity effect on arrival (γ^{s}_{Pop})	0.80	(0.09)
Arrival distribution shape parameter (γ^{s}_{shape})	0.32	(0.03)
Shopper price coefficient (α)	15.77	(1.28)
Outside option relative price (Δp^{out})	0.16	(0.02)
Effect of unobservable on price coefficient (c)	-0.91	(0.07)
Departures from fully rational model		
Probability of randomly chosen price (R^p)	0.26	(0.02)
Mean of randomly chosen price (μ^R)	15.25	(0.80)
Random price shape parameter (s^R)	1.73	(0.09)
Background disappearance rate (R^q)	0.07	(0.01)

Table 5: Estimates of structural model parameters

of the unobservables ξ_{ikt}:

$$E(W_{kt}|\hat{\Lambda}) = \frac{1}{I_{kt}} \int_{\xi} \left(E(CS_{kt}|\hat{\Lambda}, \xi) + \sum_i E(\pi_{it}(p_{ikt}, p_{-ikt})|\hat{\Lambda}, \xi) \right) f(\xi|p_{1kt}, \ldots, q_{I_{kt}kt}\} d\xi,$$

where I_{kt} is the number of listings for title k, CS_{kt}, is the total discounted consumer surplus generated by the eventual sales of all the listings and π_{it} is the discounted expected profit that the firm listing copy i will earn.

The profit term can be computed using the same approach we use in estimating the model. Given the estimated parameters, the price p_{ikt}, and values for the unobservables γ_{0kt} and δ_{kt}, firm i's profits are

$$E(\pi(p_{ikt}, p_{-ikt})|\hat{\Lambda}, \xi) = \frac{p_{ikt}(\gamma_{ikt} D^m(p_{ikt}) + \gamma_{0kt} D^o(p_{ikt}, p_{-ikt}))}{r + \gamma_{ikt} D^m(p_{ikt}) + \gamma_{0kt} D^o(p_{ikt}, p_{-ikt})}$$

We numerically integrate this function over the unobserved δ_{kt} and γ_{0kt}, treating the unobserved γ_{ikt} in two ways. With some posterior probability the model estimates that firm i chose its price rationally, in which case we plug in the backed out value of $\hat{\gamma}_{ikt}$ that rationalizes the observed prices. And with the complementary probability (corresponding to p_{ikt} having been randomly chosen) we simply plug in the estimated mean.

Consumer surplus from sales to nonshoppers can be calculated almost identically. Writing t_i for the random variable giving the time when listing i will be sold it is

$$\begin{aligned} E(CS^{ns}_{kt}) &= \sum_i E(e^{-rt_i}) \text{Prob}\{i \text{ sells to a nonshopper}\} E(v - p_{ikt}|v > p_{ikt}) \\ &= \sum_i \frac{\gamma_{ikt} D^m(p_{ikt}) + \gamma_{0kt} D^o(p_{ikt}, p_{-ikt})}{r + \gamma_{ikt} D^m(p_{ikt}) + \gamma_{0kt} D^o(p_{ikt}, p_{-ikt})} \frac{\gamma_{ikt} D^m(p_{ikt})}{\gamma_{ikt} D^m(p_{ikt}) + \gamma_{0kt} D^o(p_{ikt}, p_{-ikt})} \frac{p_{ikt}}{\eta - 1}, \\ &= \sum_i \frac{\gamma_{ikt} D^m(p_{ikt})}{r + \gamma_{ikt} D^m(p_{ikt}) + \gamma_{0kt} D^o(p_{ikt}, p_{-ikt})} \frac{p_{ikt}}{\eta - 1}, \end{aligned}$$

We used a logit demand function because we could not practically estimate something more flexible and felt that this functional form was a reasonable one for estimating how many highly price-sensitive consumers are in the market and how price senstive they are. It would not, however, be reasonable to estimate consumer surplus using the formula that comes from the standard discrete-choice justification for logit demands: the justification assumes that consumers are equally willing to substitute between purchasing and not puchasing as between purchasing from any two retailers, which is clearly unreasonable in this context. Accordingly, we instead estimate shoppers' welfare gains by assuming that the distribution of consumer surplus among shoppers who buy a book at a price of p matches the estimated distribution of consumer surplus among nonshoppers who buy at that same price. With this assumption,

$$\begin{aligned}
E(CS^s_{kt}) &= \sum_i E(e^{-rt_i})\text{Prob}\{i \text{ sells to a shopper}\}E(v - p_{ikt}|v > p_{ikt}) \\
&= \sum_i \frac{\gamma_{ikt}D^m(p_{ikt}) + \gamma_{0kt}D^o(p_{ikt}, p_{-ikt})}{r + \gamma_{ikt}D^m(p_{ikt}) + \gamma_{0kt}D^o(p_{ikt}, p_{-ikt})} \frac{\gamma_{0kt}D^o(p_{ikt}, p_{-ikt})}{\gamma_{ikt}D^m(p_{ikt}) + \gamma_{0kt}D^o(p_{ikt}, p_{-ikt})} \frac{p_{ikt}}{\eta - 1}, \\
&= \sum_i \frac{\gamma_{0kt}D^o(p_{ikt}, p_{-ikt})}{r + \gamma_{ikt}D^m(p_{ikt}) + \gamma_{0kt}D^o(p_{ikt}, p_{-ikt})} \frac{p_{ikt}}{\eta - 1},
\end{aligned}$$

Note that this assumption ignores the additional source of surplus in a standard discrete choice model that comes from being able to purchase from a retailer for which a consumer has a high idiosyncratic preference. We do not feel that this utility gain is at all important in our application, and given the estimated parameters it would only be on the order of a few cents if we did incorporate it.

Note that the equations above imply that our estimates will have the property that total surplus is split between merchants and consumers in the ratio of 1 to $\frac{1}{\eta-1}$. Hence, the estimate of the producer-consumer split is coming from the single estimated parameter η. Note also that our model for 2009 offline sales differs from the the simple isoelastic demand model in which we showed that expected social welfare is equal to the average price only in its allowance that some prices may be set randomly rather than in a profit-maximizing way. The models that we use for 2009 and 2012 online sales are more complicated multi-population models in which estimates of social welfare are affected by various other estimated parameters as well.

6.5 Welfare gains from Internet sales

In this section we present profit and welfare estimates calculated using the above methodology. Among our main findings are that profits and consumer surplus resulting from online sales are quite large and represent a substantial gain to market participants relative to the offline market for used books. Welfare appears to have remained roughly constant between 2009 and 2012.

The first row of Table 6 presents estimates of the expected gross profit per listing. More precisely, it is the average across titles of the average across listings of the estimated gross profit given the listing's price and our estimated demand parameters.[37] The first column indicates that average per-listing profits are estimated to have been fairly low in the offline

[37] These are "gross" profits in that they do not account for the acquisition cost of the books being sold.

world, just \$3.62 per listing. This is the product of the mean price for the titles, \$11.95, and a discount factor reflecting that sales occur probabilistically in the future. The estimated discount factor of 0.30 reflects our estimates that many books would take many years to sell.[38]

The second column gives comparable figures for the 2009 online listings. It illustrates the dramatic increase in profits from moving online: per listing gross profits are estimated to be over twice as large at \$7.86. The higher gross profits reflect both higher average prices and a higher estimated sales rate, which reduces the extent to which the eventual sales are discounted. Note, however, that the ten to twenty times higher arrival rate does not reduce the effective discount factor as much as a naive calculation would indicate: firms react to the higher arrival rate by increasing prices and high-priced listings take longer to sell; and the 5% discount rate we assume means that even selling a book after a decade is not such a bad outcome.

The final column presents estimates for the 2012 listings. The estimates indicate that per-listing profits are still well above the 2009 offline profit level, although not as high as the 2009 online profits. This reflects both that average prices have dropped and price-weighted waiting times have (slightly) increased. The 2012 data contain more listings offered at very low prices and it is primarily these titles that get purchased by the shoppers, so more expensive books are predicted to take longer to sell in 2012 than they did in 2009.

Average value per listing	2009 offline listings	2009 online listings	2012 online listings
Gross profit	**\$3.62** (0.27)	**\$7.86** (0.37)	**\$6.84** (0.41)
Price × Discounting	\$11.95 × 0.30	\$16.42 × 0.48	\$14.87 × 0.46
Consumer surplus	**\$4.15** (0.35)	**\$9.01** (0.75)	**\$7.84** (0.70)
Nonshoppers + Shoppers	\$4.15 + —	\$7.74 + \$1.27	\$6.50 + \$1.34
Welfare	**\$7.76** (0.53)	**\$16.87** (0.94)	**\$14.68** (0.95)

Table 6: Profit and welfare estimates

[38] It is a price-weighted average so the more expensive (and faster to sell) local interest titles are getting a higher weight in the calculation.

In thinking about the reliability of these estimates, it should be noted that the 2012 figures are estimated from better data – it is only in 2012 that we have a proxy for the rate at which listings are sold. The 2009 vs. 2012 online comparison reflects both the observed fact that average online prices were higher – recall that price distributions were otherwise similar but 2012 had more low-priced listings – and the structural model's inferences about demand. The 2009 offline profit estimates are also made without a sales proxy. We know prices were lower but are relying on the model's inference that firms set lower prices because they were facing lower demand.

The second row of the table presents estimates of consumer surplus. Consumers are estimated to receive the majority of the joint surplus.[39] Intuitively, the model rationalizes the substantial price dispersion and patterns of demand with a distribution of valuations that has a thick upper tail, This implies that the average consumer who purchases a book receives a great deal of surplus. The bottom part of the panel provides a breakdown into surplus accruing to shoppers and nonshoppers. This reflects both the relative number of purchases which the model attributes to the two populations and its inference that shoppers on average receive less surplus from their purchases. The latter effect is an implication of the assumed constant elastic demand curve that would not be present with some other specifications and could be leading us to underestimate shopper surplus.[40]

The bottom row of the table presents estimates of total welfare. Recall that our model of the 2009 offline market departs from the simplest model in which the welfare-equals-price theorem holds in its allowance for the possibility that some firms may be pricing randomly. The model estimates imply that there is a high posterior probability that some of the high prices were set randomly (as opposed to sellers having very high arrival rates for those titles) and results in estimated welfare being well below the average price. The models of the online markets also have a second difference – they are assumed to have shopper populations which have the effect of pulling down the lower tail of the price distribution. The model estimates imply that the ratio of welfare to average price is higher for this reason. The estimates are that per-listing welfare was more than twice as high in the 2009 online market than in the 2009 offline market. Recall also that the sample being used here involves an intentional oversampling of local interest titles – they constitute over 60% of the titles in the estimation sample – and we think that welfare gains were much smaller for

[39]Recall that given how we have specified the model, the ratio of profit to consumer surplus is assumed to be the same in all three markets.

[40]In the constant elasticity demand specification the expected difference between a consumer's valuation and p conditional on being willing to pay at least p is directly proportional to p.

these titles than for standard titles. Accordingly, we regard the results as indicating that the shift to online sales led to quite large welfare gains.

Per-listing welfare is estimated to be 13% lower in 2012 than it was in 2009, reflecting the increased fraction of sales to shoppers and the somewhat average lower prices. Note though that this is a reduction in per-listing welfare. The total number of listings in the dataset increases from 3758 in 2009 to 4229 in 2012 which is approximately a 13% increase.[41] Hence, our estimates suggest that total welfare was similar in 2009 and 2012.

7 Conclusion

A number of previous studies have noted that the Internet has not transformed retail markets as some forecast: price declines have been more moderate than revolutionary, and the "law of one price" has not come to pass. We began this paper by noting that the Internet market for used books shows these effects in the extreme: prices increased in a strong sense and there is tremendous price dispersion. We feel that these facts make the Internet market for used books a nice environment in which to try to gain insight into the mechanisms through which the Internet affects retail markets. Crucially, we emphasized that these basic facts do not necessarily indicate that the Internet has failed to live up to its promise. If Internet search allows consumers to find products that are much better matched to their tastes, then it leads to an increase in demand which can lead to higher prices in a variety of models (particularly for goods like out-of-print books for which supply is fairly inelastic).

The match-quality-increased-demand theory is very simple, so we devoted a substantial part of our paper to developing and empirically examining less obvious implications of the theory to help assess its relevance. We examined these implications using three sources of variation. First, we examined how price distributions – as opposed to just price levels – differ between the online and offline markets. Here, our primary observation was that the online price distribution for standard titles has a thick upper tail where the offline distribution had none. Second, we examined how price distributions differed for different types of used books. Here, we noted that there was already an upper tail of prices for local interest books in physical bookstores (which one would expect if physical bookstores were already an effective institution for matching such books to high-value owners) and that the

[41]Note also that this statistic underestimates listing growth because we have taken a sample of titles all of which had at most 49 copies in 2009 and then either (1) truncated the 2012 sample to only include the 50 lowest priced copies; or (2) dropped the title entirely if the number of listings had grown to over 300 by 2012.

growth of the upper tail is less dramatic for popular books (which one would expect if the valuation distribution had less of an upper tail). Our favorite characterization of the former result is that it appears as if the Internet has made all books of local interest. Third, we examined how online price distributions changed between 2009 and 2012. Here, we noted that the Amazon-induced increase in viewing of aggregated listings would be expected to increase the number of sellers offering very low prices but have little impact on the upper part of the price distribution, and found that this was strikingly true in the data. Our demand analysis also revealed patterns that seem consistent with the model's assumption that demand includes a shopper segment that is very sensitive not only to price, but also to differences in the order in which firms appear on a price-ordered list.

The structure of our model – in particular the use of one-dimensional unobserved heterogeneity and the assumption that firms maximize relative to steady-state beliefs – makes it relatively easy to estimate structurally. The one-to-one mapping between unobserved consumer arrival rates and observed prices makes it easy to control for endogeneity in demand and to estimate the model via simulated maximum likelihood. Our implementation of the model suggests that there were substantial increases in both profits and consumer surplus from the move to online sales of used books. We estimate that per-listing welfare-was about 80% higher online than offline in 2009, and feel that this is a very conservative estimate because local interest books are overrepresented in our sample and we would argue (and our reduced-form results suggest) that local interest books have much lower welfare gains from being sold online. Amazon's subsequent incorporation of used-book listings may have slightly reduced the per-listing profits of used book dealers, but the precision of our estimates is such that it is difficult to make precise comparisons, and an overall assessment would also need to consider Amazon's effect on the number of listings which are now available.

Our analysis has a number of limitations that could provide opportunities for future research. On the theory side it would be interesting to analyze a similar dynamic pricing problem without the steady-state beliefs we have imposed in the model: there could be interesting swings in pricing as duopolists hold off on selling in hopes of becoming a monopolist and then lower prices substantially when entry occurs and the potential for monopoly profits becomes less relevant. On the empirical side we think that the combination of assumptions we have used could make other analyses tractable as well, but think that it would also be worth exploring generalizing our model in other ways and allowing for multidimensional heterogeneity among firms. With regard to used books, we think that among the most important elements we have not incorporated is a relation between market prices and the flow of used books into used book dealers. As is, the model is most relevant to markets in which the supply response to a price increase is limited. This could include markets for collectibles and services like home repair contracting in which small businesses play an important role, the latter because a proprietor's time constraint can limit the number of jobs that a small firm can accept. Building out the model to incorporate a supply response, however, would be useful both in the used book context and in improving our understanding of how different markets have been affected by the Internet.

We hope that this work may also spur further attention to match-quality gains in other markets. In many other retail and personal services markets, there is a great deal of heterogeneity in consumer valuations, and an important welfare consideration is the extent to which each good or service ends up being purchased by the correct consumer. The recent literature on market design has explored this issue in depth in a number of applications in which it had not historically been a prominent concern – medical residents matching, school choice, kidney allocation, etc. – and it may be that greater insights can be obtained in many other industrial organization and labor applications as well.

References

Ackerberg, Daniel (2009): "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation," *Quantitative Marketing and Economics* 7, 343-376.

Armstrong, Mark, and Jidong Zhou (2016): "Search Deterrence," *Review of Economic Studies* 83, 26-57.

Arrow, Kenneth J., Theodore Harris, and Jacob Marschak (1951): "Optimal Inventory Policy," *Econometrica* 19, 250-272.

Baye, Michael R., Babur De los Santos, and Matthijs R. Wildenbeest (2015): "Searching for Physical and Digital Media: The Evolution of Platforms for Finding Books," in Shane Greenstein, Avi Goldfard, and Catherine Tucker (eds.) *Economics Analysis of the Digital Economy*, University of Chicago Press, 2015, 137-165.

Baye, Michael R., and John Morgan (2001): "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets," *American Economic Review* 91, 454-474.

Baye, Michael R., John Morgan, and Patrick Scholten (2004): "Price Dispersion in the Small and the Large: Evidence from an Internet Price Comparison Site," *Journal of Industrial Economics* 52, 463-496.

Baye, Michael R., John Morgan, and Patrick Scholten (2006): "Information, Search, and Price Dispersion," in Terence Hendershott (ed.) *Handbook on Economics and Information Systems*. Amsterdam: Elsevier.

Brown, Jeffrey R., and Austan Goolsbee (2002): "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry," *Journal of Political Economy* 110, 481-507.

Brynjolfsson, Eric, Yu Jeffrey Hu, and Michael D. Smith (2003): "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers," *Management Science* 49, 1580-1596.

Burdett, Kenneth and Kenneth L. Judd (1983): "Equilibrium Price Dispersion," *Econometrica*, 51, 955-969.

Butters, Gerard (1977): "Equilibrium Distributions of Sales and Advertising Prices," *Review of Economic Studies*, 44, 465-491.

Celerier, Claire, and Boris Vallee (2017): "Catering to Investors Through Security Design: Headline Rate and Complexity," *Quarterly Journal of Economics* 132, 1469-1508.

Chevalier, Judith, and Austan Goolsbee (2003): "Measuring Prices and Price Competition Online: Amazon vs. Barnes and Noble," *Quantitative Marketing and Economics* 1, 203-222.

Chioveanu, Ioana, and Jidong Zhou (2013): "Price Competition with Consumer Confusion," *Management Science* 59, 2450-2469.

Clay, Karen, Ramayya Krishnan, and Eric Wolff (2001): "Prices and Price Dispersion on the Web: Evidence from the Online Book Industry," *Journal of Industrial Economics* 49, 521-539.

Cohen, Adam (2002): *The Perfect Store: Inside eBay.* Boston: Little, Brown, and Company.

De los Santos, Babur, Ali Hortacsu, and Matthijs R. Wildenbeest (2012): "Testing Models of Consumer Search Behavior Using Data on Web Browsing and Consumer Purchases," *American Economic Review*, 102(6), 2955-2980.

Ellison, Glenn (2005): "A Model of Add-on Pricing," *Quarterly Journal of Economics*, 120(2), 585-637.

Ellison, Glenn, and Sara Fisher Ellison (2009): "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica* 77(2), 427-452. 10570.

Ellison, Glenn and Alexander Wolitzky (2012): "A Search Cost Model of Obfuscation," *Rand Journal of Economics* 43, 417-441.

Grubb, Michael D. (2015): "Consumer Inattention and Bill-Shock Regulation," *Review of Economic Studies* 82, 219-257.

Hortacsu, Ali, and Chad Syverson (2004): "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry," *Quarterly Journal of Economics* 119, 403-456.

Hong, Han, and Matthew Shum (2006): "Using Price Distributions to Estimate Search Costs," *RAND Journal of Economics* 37 (2), 257-275.

Kaplan, Greg and Guido Menzio, “The Morphology of Price Dispersion,” *International Economic Review* 56 (4), 1-41.

Karlin, Samuel (1962): “Stochastic Models and Optimal Policy for Selling an Asset,” in Kenneth J. Arrow, Samuel Karlin, and Herbert Scarf (eds.) *Studies in Applied Probability and Management Science*. Stanford: Stanford University Press.

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2010): “Online Demand Under Limited Consumer Search,” *Marketing Science* 29, 1001-1023.

Koulayev, Sergei (2013): “Search for Differentiated Products: Identification and Estimation,” *Rand Journal of Economics* 45, 553-575.

McCall, J. J. (1970): “Economics of Information and Job Search,” *Quarterly Journal of Economics* 84, 113-126.

McDonald, Stephen, and Colin Wren (2013): “Multiple Price Posting and Consumer Search Obfuscation: Evidence from an Online Market,” *International Journal of Industrial Organization*

Moraga Gonzalez, Jose Luis, and Matthijs R. Wildenbeest (2008): “Maximum Likelihood Estimation of Search Costs,” *European Economic Review* 52, 820-848.

Moraga Gonzalez, Jose Luis, Zsolt Sandor, and Matthijs R. Wildenbeest (2013): “Semi-Nonparametric Estimation fo Consumer Search Costs,” *Journal of Applied Econometrics* 28, 1205-1223.

Moraga Gonzalez, Jose Luis, Zsolt Sandor, and Matthijs R. Wildenbeest (2017): “Prices and Heterogeneous Search Costs,” *Rand Journal of Economics* 48, 125-146.

Piccione, Michele, and Ran Spiegler (2012): “Price Competition Under Limited Comparability,” *Quarterly Journal of Economics* 127, 97-135.

Reinganum, Jennifer (1979): “A Simple Model of Equilibrium Price Dispersion,” *Journal of Political Economy*, 87, 851-858.

Salop, Stephen and Joseph Stiglitz (1977): “Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion,” *Review of Economics Studies*, 49, 3-13.

Smith, Michael D. and Erik Brynjolfsson (2001): “Consumer Decision-Making at an Internet Shopbot: Brand Still Matters," *Journal of Industrial Economics* 49, 541-558.

Spiegler, Ran, "The Market for Quacks," *Review of Economic Studies*, 73, 1113-1131.

Stahl, Dale O. (1989): "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review* 79, 700-712.

Stiglitz, Joseph E. (1976), "Monopoly and the Rate of Extraction of Exhausitble Resources," *American Economic Review* 66, 655-661.

Talluri, Kalyan and Garrett van Ryzen (2004), *The Theory and Practice of Revenue Management*, New York: Springer.

Varian, Hal (1980): "A Model of Sales," *American Economic Review*, 70, 651-659.

Wilson, Chris M. (2010): "Ordered Search and Equilibrium Obfuscation," *International Journal of Industrial Organization* 28, 496-506.

Appendix A

Proof of Proposition 1

The first expression is the standard Lerner index formula for the optimal monopoly markup. The second can be derived from the first by substituting $\frac{\gamma p^m D(p^m)}{r+\gamma D(p^m)}$ for π^* and solving for ϵ. It also follows directly from the first order condition for maximizing $\pi(p)$:

$$rp^m D'(p^m) + rD(p^m) + \gamma D(p^m)^2 = 0.$$

QED

Proof of Proposition 2

As noted above, the monopoly price can be defined by

$$p^m = \underset{p}{\operatorname{argmax}}(p - \pi^*(\gamma/r))D(p).$$

The function $\pi^*(\gamma/r)$ is increasing because $\pi(p, \gamma/r)$ is increasing in γ/r for all p. Hence, the function $(p - \pi^*(\gamma/r))D(p)$ has increasing differences in γ/r and p and the largest maximizer is increasing in γ/r. QED

Proof of Proposition 3

Let $k = \frac{1-F(p^m(\gamma/r,F))}{1-G(p^m(\gamma/r,F))}$ be the ratio of demands under the two distributions when the firm charges $p^m(\gamma/r, F)$. The desired result follows from a simple two-step argument:

$$p^m(\gamma/r, G) \geq p^m(k\gamma/r, G) \geq p^m(\gamma/r, F).$$

The first inequality follows from Proposition 2 because $k \leq 1$. (This follows because $k \leq \frac{1-F(0)}{1-G(0)} = 1$.) The second holds because $\pi(p; k\gamma/r, G)$ and $\pi(p; \gamma/r, F)$ are identical at $p^m(\gamma/r, F)$ and their ratio is increasing in p. Hence for any $p < p^m(\gamma/r, F)$ we have $\pi(p; k\gamma/r, G) \leq \pi(p; \gamma/r, F) \leq \pi(p^m(\gamma/r, F); \gamma/r, F) \leq \pi(p^m(\gamma/r, F); k\gamma/r, G)$. QED

Proof of Proposition 4

With constant elasticity demand expected consumer surplus of the consumer who purchases the good is directly proportional to the monopolist's profits, $E(v - p|v > p) = \frac{p}{\eta-1}$. Result (i) follows immediately: the maximizer of the sum of profits and consumer surplus is the same as the maximizer of profits.

Welfare is given by

$$E(W) = \int_0^\infty \left(p^m + \frac{p^m}{\eta - 1}\right) e^{-rt}\gamma D(p^m)e^{-\gamma D(p^m)t}dt = \frac{\gamma D(p^m)}{r + \gamma D(p^m)} p^m \left(1 + \frac{1}{\eta - 1}\right).$$

The FOC for profit maximization, $rp^m D'(p^m)+rD(p^m)+\gamma D(p^m)^2 = 0$, can be manipulated to show that $r+\gamma D(p^m) = r\eta$ and $\gamma D(p^m) = r\eta - r$. Result (ii) then follows from simplifying the formula for welfare given above. QED

Proof of Proposition 5

Let $\overline{V}(p_{-i}) \equiv \max_p \pi(p_i, p_{-i})$ be firm i's profit when it plays a best response to p_{-i}. The best response correspondences satisfy

$$BR_i(p_{-i}) = \underset{p_i}{\operatorname{argmax}}(p_i - \overline{V}(p_{-i}))D(p_i, p_{-i}).$$

This will be montone increasing if the function on the RHS has increasing differences in p_i and p_{-i}. Writing $\tilde{\pi}(p_i, p_{-i})$ for the function and differentiating twice we see

$$\frac{\partial^2}{\partial p_i \partial p_{-i}} \tilde{\pi}(p_i, p_{-i}) = \frac{\partial^2}{\partial p_i \partial p_{-i}} \left((p_i - c)D(p_i, p_{-i})\right)\big|_{c=\overline{V}(p_{-i})} - \frac{\partial \overline{V}}{\partial p_{-i}} \frac{\partial D}{\partial p_i}.$$

The first term on the right is nonnegative by the assumption about the static profit function because we only need to consider prices above $\overline{V}(p_{-i})$ (because demand is a probability and hence less than one). The second is positive because firm i's demand is decreasing in p_i and the value function are increasing in p_{-i}. As in Milgrom and Roberts (1990), this also suffices to guarantee equilibrium existence. QED

Appendix B

To estimate the model by maximum likelihood we choose the parameter vector Δ to maximize the likelihood of observing $p_1, \ldots, p_N$ given Δ, i.e. $\sum_i \log(L(p_i|\Delta))$. To avoid computational difficulties that arise in finding exact solutions to first order conditions and evaluating the change of variables derivative $\frac{1}{\frac{\partial g}{\partial \gamma}(\gamma_{ik})}$ that appears in the likelihood formula, we treat each observed price as if we are observing that the price lies in an interval with width of 5 cents centered at the observed price. The likelihood then contains two parts. The first part is the likelihood assuming the firm optimally sets the price in $[p-0.025, p+0.025]$. Note that holding fixed the parameters Δ we can regard optimal price $p_i = g(\gamma_i; p_{-i}, \Delta)$ as a function of γ, so we calculate the value of $\overline{\gamma}_{ik} = g^{-1}(p+0.025)$ and $\underline{\gamma}_{ik} = g^{-1}(p-0.025)$, and let the likelihood be $F_{\gamma_{ik}}(\overline{\gamma}_{ik}) - F_{\gamma_{ik}}\left(\underline{\gamma}_{ik}\right)$, according to the parameterized CDF of γ_{ik}.[42] The second part is the likelihood assuming the firm sets price randomly in $[p-0.025, p+0.025]$, so the likelihood is $f(p_{ik}) \times 0.05$ according to the parameterized CDF f of random price setting. To summarize,

$$L(p_i|\Delta)) = (1-q)\left(F_{\gamma_{ik}}(\overline{\gamma}_{ik}) - F_{\gamma_{ik}}\left(\underline{\gamma}_{ik}\right)\right) + R^p f(p_{ik}) \times 0.05$$

Note for many of the combinations of Δ, $\overline{\gamma}_{ik}$ or $\underline{\gamma}_{ik}$ is complex number or negative for at least one p_{ik}. If $\underline{\gamma}_{ik}$ is complex or negative, we take $\underline{\gamma}_{ik} = 0$ instead; if $\overline{\gamma}_{ik}$ is complex or negative, the first item is regarded as zero.

To account for the random coefficients we consider a fixed grid of possible (γ_{0k}, δ_k) constructed by crossing $\left\{\gamma_{0k}^{(n)}\right\}_{n=1}^{N}$ and $\left\{\delta_k^{(m)}\right\}_{n=1}^{M}$. Each point is assigned an importance weight according to the parametrized distribution,

$$\begin{aligned} I\left(\gamma_{0k}^{(n)}, \delta_k^{(m)}\right) &= \left(F\left(\frac{\gamma_{0k}^{(n+1)} + \gamma_{0k}^{(n)}}{2}\right) - F\left(\frac{\gamma_{0k}^{(n-1)} + \gamma_{0k}^{(n)}}{2}\right)\right) \\ &\quad \times \left(F\left(\frac{\delta_k^{(m+1)} + \delta_k^{(m)}}{2}\right) - F\left(\frac{\delta_k^{(m-1)} + \delta_k^{(m)}}{2}\right)\right) \end{aligned}$$

with adjustments $F\left(\frac{\gamma_{0k}^{(0)}+\gamma_{0k}^{(1)}}{2}\right) = F\left(\frac{\delta_k^{(0)}+\delta_k^{(1)}}{2}\right) = 0$ and $F\left(\frac{\gamma_{0k}^{(N+1)}+\gamma_{0k}^{(N)}}{2}\right) = F\left(\frac{\delta_k^{(M+1)}+\delta_k^{(M)}}{2}\right) = 1$. Conditional on each (γ_{0k}, δ_k) evaluate the likelihood of the full set of prices for the titles, i.e. we maximize

$$\begin{aligned} \sum_k \log(L(p_{1k}, \ldots, p_{n_k k}|\Delta)) &= \sum_k \log\left(\sum_{n,m} L(p_{1k}, \ldots, p_{n_k k}|\gamma_{0k}^{(n)}, \delta_k^{(m)}, \Delta) I\left(\gamma_{0k}^{(n)}, \delta_k^{(m)}\right)\right) \\ &= \sum_k \log\left(\sum_{n,m}\left[\prod_i (L(p_{ik}|\gamma_{0k}^{(n)}, \delta_k^{(m)}, \Delta)\right] I\left(\gamma_{0k}^{(n)}, \delta_k^{(m)}\right)\right) \end{aligned}$$

[42] We have done several things to ensure at $\bar{\gamma}_{ik}$, $p+0.025$ is indeed profit-maximizing. First, we check the second order condition to rules out that it is a minimum. Second, we compare the profit, assuming $\bar{\gamma}_{ik}$ and $p+0.025$, to the profit assuming the firm just optimally choose price with $\bar{\gamma}_{ik}$ and no shoppers. If the latter is higher, $\bar{\gamma}_{ik}$ is reduced to the level such that $p+0.025$ is the firm's optimal choice assuming no shoppers.

For many of the (γ_{0k}, δ_k) the likelihood will be zero for at least one p_{ik}. We simply set the product to zero for all such (γ_{0k}, δ_k).

For some observations we observe an indicator q_{ik} for whether the book is sold alongside the price p_{ik}. When this occurs we compute the joint likelihood of (p_{ik}, q_{ik}) given the parameters and a value for the random coefficients as a sum of two terms. The first comes from the possibility that the price was set at the solution to the first-order condition for optimal pricing. Here, we set $\tilde{\gamma}_{ik}$ be the mean of the upper and lower bounds on the arrival rate justifying prices in the 5 cent interval[43] and multiply the likelihood of the price observation by

$$(R^q + (1-R^q)(1-e^{-d_{ik}\Delta t}))^{wI(q_{ik}=1)}(1-R^q-(1-R^q)(1-e^{-d_{ik}\Delta t}))^{wI(q_{ik}=0)},$$

where d_{ik} is the arrival rate of consumers willing to buy the book,

$$d_{ik} = \gamma_0 D^o(p_{ik}, p_{-ik}) + \tilde{\gamma}_{ik} D^m(p_{ik}).$$

The second term is a corresponding term reflecting the probability that the price was set to p_{ik} randomly. We evaluate the joint probability that (p_{ik}, q_{ik}) arose in this manner as the product of the term we used earlier, $R^p f_q(p_{ik}) \times 0.05$, the probability that the good was sold assuming that its arrival rate matches the sample mean. The inclusion of the weighting parameter w in the exponents effectively makes the log-likelihood we compute a weighted sum of the log likelihoods of the price and quantity data. In preliminary investigations we found that the quantity data was sometimes poorly fit without such a weight and adopted a weight of 3.

To calculate expected profit and consumer surplus for a listing we first need a posterior weight on the $\left(\gamma_{0k}^{(n)}, \delta_k^{(m)}\right)$, computed by

$$I\left(\gamma_{0k}^{(n)}, \delta_k^{(m)} | p_{1k}, \ldots, p_{nk}, q_{1k}, \ldots, q_{nk}\right) = \frac{I\left(\gamma_{0k}^{(n)}, \delta_k^{(m)}\right) \prod_i (L(p_{ik} q_{ik} | \gamma_{0k}^{(n)}, \delta_k^{(m)}, \Delta)}{\sum_{n,m} \left\{ \left[I\left(\gamma_{0k}^{(n)}, \delta_k^{(m)}\right) \prod_i (L(p_{ik} q_{ik} | \gamma_{0k}^{(n)}, \delta_k^{(m)}, \Delta) \right] \right\}}$$

Average per-listing welfare for title k can then be computed as

$$\sum_{n,m} I\left(\gamma_{0k}^{(n)}, \delta_k^{(m)} | p_{1k}, \ldots, p_{nk}, q_{1k}, \ldots, q_{nk}\right) \frac{1}{N_k} \left(E(CS_k) + \sum_i E(\pi_{ik} | p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}) \right),$$

where N_k is the number of listings for title k, CS_k, is the total discounted consumer surplus generated by the eventual sales of all the listings and π_{ik} is the discounted expected profit that the firm listing copy i will earn. Both CS_k and π_{ik} have two parts. The first part is calculated assuming the firm chose price optimally. We first derive the posterior probability of random price setting to be

$$\tilde{q} = \frac{R^p \left(F\left(p + 0.025\right) - F\left(p - 0.025\right)\right)}{L(p_i | \Delta))}$$

[43]In practice the upper and lower bounds are quite close together so we simply set $\tilde{\gamma}_{ik}$ equal to the interval mean rather than worrying about integrating over the small interval.

The profit assuming optimal price setting can be computed using the same profit functions we use in estimating the model. Given the price p_{ik} and any pair of the random coefficients, $\gamma_{0k}^{(n)}, \delta_k^{(m)}$, for which the price can be rationalized given the the other estimated parameters, we back out a value for γ_{ik} again as an average of upper and lower bounds and compute expected profits as

$$E(\pi_{ik}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}, \text{optimal price}) = \frac{p_{ik}(\gamma_{ik} D^m(p_{ik}) + \gamma_{0k} D^o(p_{ik}, p_{-ik}))}{r + \gamma_{ik} D^m(p_{ik}) + \gamma_{0k} D^o(p_{ik}, p_{-ik})}$$

Expected profit conditional on random price setting is computed by assuming γ_{ik} equals its expected value given the estimated parameters[44] and plugging it into

$$E(\pi_{ik}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}, \text{random price}) = \frac{p_{ik}(E[\gamma_{ik}] D^m(p_{ik}) + \gamma_{0k} D^o(p_{ik}, p_{-ik}))}{r + E[\gamma_{ik}] D^m(p_{ik}) + \gamma_{0k} D^o(p_{ik}, p_{-ik})}.$$

So the total expected profit is

$$\begin{aligned} E(\pi_{ik}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}) &= (1-\tilde{q})\, E(\pi_{ik}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}, \text{optimal price}) \\ &\quad + \tilde{q} E(\pi_{ik}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}, \text{random price}). \end{aligned}$$

As noted in the text, we calculate consumer surplus of the shoppers as if they have the same distribution of valuations as do the shoppers who purchase at the observed price. With this observation, total consumer surplus is simply $\frac{1}{\eta-1}$ times the profit. Accordingly, once we have estimated profit we simply multiply the estimate by this fraction (given the estimated $\hat{\eta}$) to obtain an estimate of consumer surplus. To divide this into portions attributable to shoppers and nonshoppers, we compute the nonshopper portion as the weighted sum

$$\begin{aligned} E(CS_k^{ns}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}) &= (1-\tilde{q})\, E(CS_k^{ns}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}, \text{optimal price}) \\ &\quad + \tilde{q} E(CS_k^{ns}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}, \text{random price}). \end{aligned}$$

Each term is computed analogously to how we computed profits.

$$\begin{aligned} & E(CS_k^{ns}|\{p_{ik}\}_i, \gamma_{0k}^{(n)}, \delta_k^{(m)}, \text{optimal price}) \\ = & \sum_i E(e^{-rt_i}) \text{Prob}\{i \text{ sells to a nonshopper}\} E(v - p_{ik}|v > p_{ik}) \\ = & \sum_i \frac{\gamma_{ik} D^m(p_{ik}) + \gamma_{0k} D^o(p_{ik}, p_{-ik})}{r + \gamma_{ik} D^m(p_{ik}) + \gamma_{0k} D^o(p_{ik}, p_{-ik})} \frac{\gamma_{ik} D^m(p_{ik})}{\gamma_{ik} D^m(p_{ik}) + \gamma_{0k} D^o(p_{ik}, p_{-ik})} \frac{p_{ik}}{\eta - 1} \\ = & \sum_i \frac{p_{ik}\gamma_{ik} D^m(p_{ik})}{r + \gamma_{ik} D^m(p_{ik}) + \gamma_{0k} D^o(p_{ik}, p_{-ik})} \frac{1}{\eta - 1} \end{aligned}$$

and again, $E(CS_k^{ns}|p_{ik}, \gamma_{0k}^{(n)}, \delta_k^{(m)}$, random price) is computed by replacing γ_{ik} with its expectation given the estimated parameters.

To compute standard errors, we resampled at the title level and reestimated the model. Our current standard errors were computed with 200 resamplings.

[44]Here, the approximation error relative to integrating over the space of possible values of γ_{ik} may be larger.