Match Quality, Search, and the Internet Market for Used Books[1]

Glenn Ellison and Sara Fisher Ellison

Massachusetts Institute of Technology

September 2014
PRELIMINARY

[1]E-mail: gellison@mit.edu, sellison@mit.edu. We would like to thank David Liu, Hongkai Zhang, Vivek Bhattacharya, and Chen Lian for outstanding research assistance, Gabriel Carroll for comments on an earlier draft, and both the NBER Sloan Program on Digitization and the Toulouse Network for Information Technology for financial support.

Abstract

The paper examines Internet-related changes in the used book market. A model in which sellers wait for high-value consumers brings out two expected effects: improvements in the match-quality between buyers and sellers raise welfare (and may lead to higher prices); meanwhile increased competition brings down prices especially at the lower end of the distribution. The paper examines differences between offline and online prices in 2009 and between online prices in 2009 and 2012 and finds several features consistent with the model predictions. Most notably, online prices are higher than offline prices, suggesting a substantial match-quality effect. The paper develops a framework for structural estimation using the available price and quantity data. Structural estimates suggest that the shift to Internet sales substantially increased both seller profits and consumer surplus.

1 Introduction

The empirical literature on Internet pricing has found that online commerce did not have the dramatic price-lowering and law-of-one-price-reinforcing effects that some had forecast. Brynjolfsson and Smith (2000), for example, 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. We note here that the used book market appears to be an extreme example of these outcomes: online prices are typically higher than offline prices and are very dispersed.[1] The failure of the Internet to bring about low and homogeneous prices has often been seen (including in Ellison and Ellison (2009)) as an indication that Internet retail markets are not working as well as one might have expected. The absence of a price decline, however, can also have a much more agreeable cause: if the Internet allows consumers to find goods that are better matched to their tastes, then there is effectively an increase in demand, which would lead to higher prices in many monopolistic and competitive models. In this paper, we explore this improvement-in-match-quality idea in the context of the Internet market for used books. We develop some simple models of how reduced search costs would affect price distributions in a model with competition and match-quality effects, find that several predictions of the model are borne out in our data, discuss how a version of the model can be structurally estimated, and present estimates suggesting that Internet sales of used books may be providing substantial profit and welfare benefits.

A personal anecdote may help illustrate some of these ideas. 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![2] 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

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

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

more.

It seems natural that many used book dealers were Internet sales pioneers in the early to mid 1990's. First, many potential purchases of used or out-of-print books were presumably never consummated due to time cost of finding books in physical bookstores in the pre-Internet era. The Internet promised a solution to this inefficiency through search technologies. In addition, books were particularly amenable to these search technologies and remote sales because they are both easily describable and easy to ship. In the second half of the 1990's, Bibliofind, AbeBooks, Bookfinder, Alibris, and others developed web sites that aggregated listings from multiple bookstores, helping savvy consumers find the books they wanted and compare prices. 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. In the late 2000's there was a second substantial change after Amazon acquired AbeBooks. The 2008 acquisition had little immediate impact – Amazon initially left AbeBooks to operate as it had. But in 2010 they launched a program to allow used book dealers to have their AbeBooks listings also appear on Amazon. The addition of "buy used" links on Amazon could potentially have had a large effect on the number of consumers who viewed aggregated used book listings.

In practice we think that Internet has reduced search costs for used books in two different ways: Google will instantly find some online bookstore that has a copy of just about any title that a consumer types into its search box (which is a huge improvement over calling stores one-by-one on the phone until finding a copy); and consumers who are familiar with sites like Abebooks and Alibris (or who see their listings via Amazon) can now quickly find and compare prices from many online retailers. We begin in section 2 with some simple models that aim to help understand how such technological improvements might affect price distributions and sales patterns. We cast the selling of unique items as a waiting-time problem: the firm sets a price p for its item, and consumers with valuations drawn from a distribution F arrive at a known Poisson rate γ. In a monopoly model, we note that prices increase when the arrival rate increases and when the value distribution has a thicker upper tail (which we think of as possible consequences of reduced search costs of the first type). Heterogeneity in firm-specific arrival rates, e.g. from some firms appearing higher in Google's results, will naturally lead to price dispersion. We also provide a striking theorem on price levels and welfare that holds in one special case: when demand has the constant-elasticity form, then price changes and welfare changes are perfectly aligned. To

incorporate the second factor that we think is important – a technology that lets some consumers compare prices – we first discuss a complete-information oligopoly model and then a hybrid model along the lines of Baye and Morgan (2001) in which some "shoppers" see all prices and some "nonshoppers" only visit a single firm. We note that there can be a pure-strategy dispersed-price equilibrium if firms have heterogeneous nonshopper arrival rates, and we discuss the form of the first-order condition. Most importantly, we note that a shift to online sales would be expected to have different effects at different parts of the price distribution. Improvements in match quality will increase prices at the high end. Meanwhile increased price competition will pull down prices at the bottom end of the distribution.

In a final subsection of section 2 we highlight four testable implications of the model for how prices should be distributed. One of these concerns how offline and online price distributions should compare: although the net change in average prices is ambiguous (because the two effects work in opposite directions) we can say that the online market should have more high prices and that online prices should be more dispersed than offline prices. Two other predictions focus on how prices should compare across different types of books. In addition to the "standard" set of mostly out-of-print titles that fill most of the shelves at the stores we visited, we oversampled books that are of "local interest" to the area where the store is located, and we label a number of books found at a large number of Internet retailers as "popular." Our fourth prediction concerns how one would expect the distribution of online prices to have changed between 2009 and 2012 due to Amazon's incorporation of Abebooks's listings into its "Buy Used" option (which we think of as having increased the proportion of consumers who are "shoppers").

Section 3 presents the dataset. The data include information on 335 titles which we first found at brick and mortar used book stores in 2009. The set of titles was chosen to allow us to separately analyze three distinct types of books. We collected offline prices and conditions for these books in 2009. And we collected prices and conditions for copies of the same books listed online via AbeBooks.com in 2009, 2012, and 2013. The 2009 online data collection lets us compare contemporaneous online and offline prices. The 2012 data collection lets us examine how prices compare before and after Amazon's incorporation of the AbeBooks listings. The 2013 data collection provides something akin to demand data by letting us look at whether copies listed in November 2012 are no longer listed for sale two months later. In addition to presenting summary statistics, we also show the price distributions of three "typical" titles.

Section 4 presents general descriptive evidence on online and offline prices and examines the four predictions mentioned above. We begin with our basic observations that online prices are typically higher than offline prices and are very dispersed. The online-prices-are-higher result holds in a very strong sense: for more than half of our "standard" titles, the single offline price that we found for the title was below every one of the online prices we found. We then examine the additional predictions of the theory and find a number of striking facts that support the model's applicability. Among these are that the online price distribution for standard titles has a much thicker upper tail than the offline version, that offline and online prices are more similar for "local interest" titles (as if the Internet is making the market for all titles look more like the market for local interest titles), and that between 2009 and 2012 there has been growth in the number of sellers listing very low prices with strikingly little change in the upper tail.

Section 5 examines the demand for online used books in 2012. Again, the goal is both to present descriptive evidence on demand patterns and to assess the applicability of our model. Among the findings are that demand appears to be low and fairly price sensitive: the single lowest-priced listing has a substantial chance of being sold in less than two months, but an average-priced listing will not sell for years.

Observed price differences between offline and online books can be thought of as reflecting the net of two effects: improved match-quality online increases the effective demand for books, pushing prices up; and increased competition online, because prices across sellers are easier to compare, pulls prices down. Estimating profit and/or social welfare requires separately estimating the magnitudes of the two gross effects. In section 6 we develop a structural approach to provide such estimates. We begin by describing an econometric model along the lines of the theoretical model of section 2: shoppers and nonshoppers arrive at Poisson rates, firms are heterogeneous in the arrival processes they face, and products are sufficiently differentiated so that pure strategy equilibria exist. In a parsimonious model, we note that there is a one-to-one correspondence between prices and arrival rates. This 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. The structural estimates indicate that arrival rates 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 a (very small) inflow of much less price sensitive "nonshoppers." (These nonshoppers might arrive at a used book

seller's website directly through typing in a title on Google, for instance.) Our profit and welfare calculations indicate that both book dealers and consumers are benefitting from the shift to online sales: profits and consumer surplus are estimated to be substantially higher in the online environment than in the 2009 offline world. Per-listing profit levels do, however, appear to have declined by about 25% between 2009 and 2012, perhaps due to the increased use of price comparison tools.

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 the degree of online price dispersion. One early study, Bailey (1998), collected prices from samples of online and offline retailers in 1997 and reported that online prices for software, books, and CDs were 3% to 13% higher on average than offline prices. Later studies like Brynjolfsson and Smith (2000), Clay, Krishnan, and Wolff (2001), Baye, Morgan, and Scholten (2004), and Ellison and Ellison (2009), however, report that online prices are lower than offline prices. All of these studies note that there is substantial dispersion in online prices. 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.

On the theory side, we have not found an exact antecedent in the literature for our simple initial analysis of monopoly pricing with Poisson arrivals. The model could be thought of as the simplest special case of the dynamic inventory model of Arrow, Harris, and Marschak (1951), and similar stopping time problems where the price offers are made by buyers rather than sellers can be found going back at least to 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., Our subsequent

consideration of oligopoly pricing is influenced by the 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).[3] 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 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 (2013). Relative to this literature, we will not try to estimate search costs to rationalize demand – instead we focus just on estimating a consumer arrival process from price distributions (and some quantity data) in a model that allows for substantial firm-level heterogeneity, in contrast to much of this literature which assumes firms are identically situated. Broadly speaking, 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 simple monopoly and duopoly models that can be used to think about the problems faced by traditional and online used book dealers. The models provide some predictions about offline and online prices tested in section 4 and motivate the structural model estimated in section 6.

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

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 bookstore or an Internet store serving customers who visit only its website. It will also serve as a starting point for our subsequent analysis of an oligopoly model in which some consumers also search across stores.

Suppose that a monopolist has a single unit of a good to sell. Consumers randomly arrive at the monopolist's store according to a Poisson process with rate γ. The value v of the 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 the consumer will buy is $D(p) = 1 - F(p)$. We assume that $\lim_{p\to\infty} pD(p) = 0$ to ensure that optimal prices will be 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 $\gamma D(p)$, which we will call the "net arrival rate."[4] 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

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

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

$$p^m = \operatorname*{argmax}_p \pi(p) = \operatorname*{argmax}_p \frac{\gamma pD(p)}{r+\gamma D(p)}.$$

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.

[4] We will use the term "net arrival" to refer to the arrival of potential customers who *actually* purchase whereas we will still use the non-modified term "arrival" to refer to the arrival of potential customers regardless of their willingness to pay.

The second way to think about the dynamic profit maximization problem is as a dynamic programming problem. Let π^* (which depends on γ, 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

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

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

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

Proof

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

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 fairly tractable example is a truncated constant elasticity demand curve: $D(p) = \min\{1, hp^{-\eta}\}$. Here the monopoly price is

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

One may be interested in comparative statics results on this dynamic monopoly price. If, for instance, one thinks that a difference between offline and online used bookstores is that more consumers may visit online stores, it would be interesting to know how the monopoly price varied with arrival rate.

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

Proof

As noted above, the monopoly price can be defined by

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

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

Remarks:

1. When the arrival rate γ is small, the monopolist's problem is approximately that of a standard monopolist with zero costs, so the monopoly price approaches the maximizer of $pD(p)$.

2. The behavior of the monopoly price in the $\gamma/r \to \infty$ limit 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 $\gamma/r \to \infty$. 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.

3. 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: p^m

remains bounded above by one as $\gamma/r \to \infty$ and converges to this upper bound at just a $1/\sqrt{\gamma/r}$ rate. 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 1, the monopoly price is almost proportional to the arrival rate. But when the tail is thinner, i.e., when η larger, the monopoly price increases more slowly. When demand is exponential, $D(p) = e^{-\gamma p}$, the monopoly price is bounded above by a constant times $\log(\gamma/r)$.

In addition to having different arrival rates of potential consumers, 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 for the book, whereas consumers who are browsing in a physical bookstore will be equally likely to come across titles for which they have relatively low and high valuations. One way to capture such an effect would be to assume that offline searchers' valuations are random draws from the full population $f(v)$ whereas online searchers are more likely to have valuations drawn for the upper part of the distribution. In particular, the likelihood that a consumer with value v searches online for a title is an increasing function $q(v)$ of his or her valuation for that title. The density of valuations in the online searcher population will then be $g(v) = af(v)q(v)$ for some constant a. Note that 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 $p^m(\gamma/r, G) \geq p^m(\gamma/r, F)$.*

Proof

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

The monopoly model with constant elasticity demand also has an interesting welfare property that will come up in our empirical implementation. Given the formula we saw earlier, $p^m = \left(\frac{h}{\eta-1}\right)^{1/\eta}\left(\frac{\gamma}{r}\right)^{1/\eta}$, any observed price can be rationalized by a variety of (γ, h, η) combinations, e.g. a high price could indicate that demand is very elastic and arrival rates are high or that demand is less elastic but arrival rates are low. It turns out, however, that social welfare is identical across all such combinations.

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 expected social welfare in the model is given by* $E(W) = p^m$.

Proof

With constant elasticity demand, $E(v-p|v>p) = \frac{p}{\eta-1}$. Hence,

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

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 $\gamma D(p^m) = r\eta - r$. The result then follows from simplifying the formula for welfare given above. QED

On the positive side, the result can be seen as a powerful tool for estimating social welfare: if we are willing to assume that firms are profit maximizing and demand has the constant elasticity form, then there is no need to try to estimate arrival rates of consumers or the distributions of their valuations: we can immediately infer expected social welfare just from observing a firm's price. Obviously, however, assuming that demand has the constant elasticity form is a strong assumption and welfare would not be parameter-independent if demand belonged to a different family. Here, the result can also be thought of as providing the cautionary observation that estimated welfare will be approximately equal to the price of the good unless demand estimation is sufficiently flexible so that $E(v-p^m|v>p^m)$ is not approximately equal to what it would be with a constant elasticity demand curve that matched the estimated demand elasticity at the observed price.

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 could suddenly become more or less intense as additional sellers enter or current sellers sell their goods. Optimal pricing in such a setting could be an interesting topic to explore, but in this paper we will explore 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.[5] 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 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.*

Proof

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

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

[5] 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.

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

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

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

Remarks:

1. The first-order condition for the equilibrium in the full information oligopoly model is similar to that for the monopoly price in the monopoly model:

$$r p^* \frac{\partial D_i}{\partial p_i}(p^*, p^*) + r D_i(p^*, p^*) + \gamma_0 D_i(p^*, p^*)^2 = 0.$$

2. 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.

3. A more precise statement of the previous remark on comparative statics is that the set of Nash equilibrium prices increases in γ_0 in the strong set order. This is relevant because 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.

4. 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.[6] 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 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 it should be noted that the modified models will not globally have the increasing differences property.

In practice, we know that there is a great deal of price dispersion in online used book prices. 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.[7] 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

[6] 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.

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

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 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:

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

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, a lot of nonshoppers relative to shoppers, will set prices close to $p^m(\gamma_i)$.[8] 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 $p^m(\gamma_{-i})$.

[8] 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.

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 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.[9] 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.

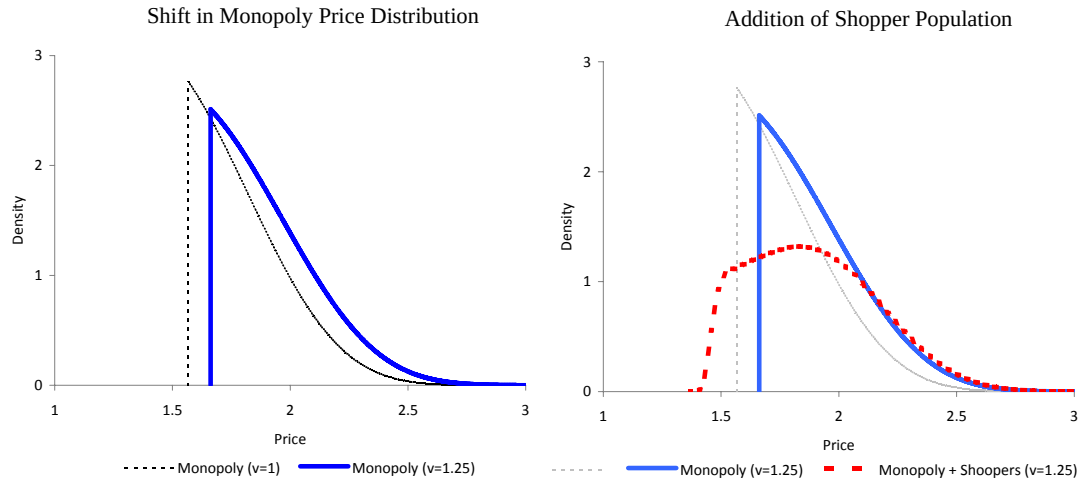

Figure 1: Numerical example: Effects of increasing valuations and adding shoppers

The left panel of Figure 1 illustrates the first effect. The thinner dashed line graphs the

[9] 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.

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 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.[10] Each firm in this model faces a nonshopper 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 blue and dashed red lines) is different at different parts of the distirbution. There is essentially no competition effect on the upper part of the distribution: firms with high nonshopper arrival rates essentially ignore the shopper population.[11] 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.

In this numerical example the competition effect is sufficiently powerful so that the online oligopoly distribution has more low prices than even the offline monopoly distribution (the light gray line). This, however, is not a robust prediction. Comparisons of the lower tails and comparisons of average prices will both be parameter dependent: when the competition effect dominates, the online price distribution will feature more low prices; and when nonshopper arrival effect dominates, there will be fewer low prices online.

[10] 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.

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

2.4 Implications for the market for used books

Our model provides very simple potential explanations for the facts that motivated our study. The fact that online prices are higher than offline prices can be explained as a consequence of the Internet having increased the flow of consumers who visit used book stores. And the fact that online prices are dispersed can be explained as a consequence of different stores' facing different consumer arrival rates (which could in turn come from some stores being higher in Google's search results, or being more prominent, etc.). But, of course, the basic facts are things we knew before writing down the model and 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. In particular, we focus on predictions about the distributions of online and offline prices.

Recall that we think of the Internet as having affected the used book market via two channels. First, the Internet makes it less costly for consumers to visit used book stores. It thereby may have increased both the number of consumers who come across each available copy and the distribution of valuations of these consumers. These consumers include, among others, consumers who type a title into a search box and find a listing for the title at some book store, consumers who browse online bookstores they know, and consumers who click on an online advertisement for a used bookstore. Second, the Internet provides a technology that lets some consumers find and compare prices for a given title from many online bookstores. These are consumers who use AbeBooks, Alibris, or other similar sites and, more recently, Amazon's "buy used" button. The first effect is reflected in the model as a higher flow rate of arrival of nonshoppers and a higher distribution of nonshopper valuations online versus offline. The second effect is reflected in the fact that the online market includes a population of shoppers where the offline world does not.

To begin consider the comparison between offline and online prices. In the model we would think of this as comparing the light gray and the dashed red lines in the right panel of Figure 1. 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. (Recall that this occurs because firms with few nonshoppers focus on attracting shoppers whereas firms with many nonshoppers largely

ignore the shoppers and price to their captive customers.) 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.

An additional observation crucial for our empirical analysis is that the strength of the two effects noted above will differ depending on characteristics of the particular title. This difference will allow us to make additional predictions about how price distributions will vary across titles. One characteristic of titles, immediately apparent when visiting many brick-and-mortar used book stores, is that some books are of "local interest": each of the stores we visited had a bookcase or section labeled as containing such books. Some examples of such books 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*. Our primary thought about such books is that physical used book stores may already have been very good at matching them to their high-value readers. For example, person willing to pay the most for a copy of *Boston Catholics* may be a Catholic tourist who is visiting Boston and who has spent the day visiting churches and traditionally Irish/Italian Catholic neighborhoods. A second thought is that such books may not benefit much from Internet listing. For example, the potential high-value consumer of *Boston Catholics* imagined above may only have learned of its existence by coming across it on a shelf and would not have typed its title into an Internet search engine. In terms of our model, the former could be captured by assuming that the offline arrival rate and/or the distribution of values of offline arriving consumers is more favorable for local interest titles than for standard titles, perhaps being comparable to the arrival rates/distributions that standard titles receive when listed online. The latter could be captured by assuming that the difference between offline and online arrival rates is smaller for local interest titles than it is for standard titles. We then have the following 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.

The story we told in the introduction of searching for an obscure book on the pharmaceutical industry depended importantly on the book's obscurity. If, instead, the book we

had sought was a copy of Michael Dibdin's 1994 detective novel *Dead Lagoon*, the story would have been different. For one thing, our willingness-to-pay would have been lower: if the cheapest copy we could find was $20 the story might have been that we decided to buy a different detective novel instead, or that we realized it was only the hardcover that is out of print and decided to buy a new paperback copy from Amazon. Competition would also have come into the story differently: finding a copy at the neighborhood used book store would not have been so unlikely; and an online search would have netted many available copies for far less than $20. In terms of our model, we can think of there being a few main differences between the markets for such "popular" titles and the more obscure long out-of-print titles we have been thinking about as "standard": the distribution of consumer valuations for a popular title will probably have a much thinner upper tail (because consumers are willing to substitute to an alternate title, to substitute to a new paperback copy, or to look for a copy in a physical bookstore); and the number of online sellers will be larger for popular books. As we noted in section 2.1 the rate at which prices increase when the consumer arrival rate increases depends on the thickness of the upper tail of the consumer value distribution. These observations give us 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.

A comparison between 2009 and 2012 is also potentially informative. We think of Amazon's incorporation of AbeBooks's listings under its "buy used" button as the most salient change in the market in between the two dates. 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 can also be thought of as illustrating the potential effect of such a change: the price distribution would shift further in the direction of the change from the thick blue line to the thick dashed red line. Recall that we noted there that adding searchers to the model pulls down the lower end of the price distribution (because firms with few nonshoppers focus on competing for the shoppers) but leaves 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 observation 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 two distributions, however, should be similar.

We will examine each of these predictions in section 4.

3 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 a number of 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 and shipping charges 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.[12] AbeBooks' default sort is on the shipping-inclusive price, which makes sense due to the heterogeneity in how sellers use shipping charges – some sellers offer free shipping, others have shipping fees in line with costs, and others have very high shipping fees. In most of our analyses we will use a price variable defined as listed price plus shipping charges minus two dollars. This price is designed to reflect the money received by the seller from the sale (assuming $2 is 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, but includes books in a variety of conditions. We collected information on the condition of each online copy and control for condition differences in some analyses. For most of the titles the online data include the complete set of listings on www.AbeBooks.com.[13] But for some titles with a large number of listings we only collected a subset of the listings. 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.

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.

The quasi-random set of titles selected was influenced by a desire to have enough books

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

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

of different types to make it feasible to explore how price distributions vary with the title characteristics. First, we intentionally oversampled books of "local interest." We define this category 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 selected by oversampling from shelves in the bookstores labeled as having books of local interest and hence are of interest to the bookstore's location, but that it not always the case: some are what we call displaced local interest books which were randomly swept up in our general collection but are of local interest to some other 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.

Second, we collected data on a number of "popular" books. We define this subsample formally in terms of the number of copies of the book we found in our 2009 online search: we classify a book as popular if we found more than 50 copies online in 2009. Some examples of popular books in our sample are Jeff Smith's cookbook *The Frugal Gourmet*, Ron Chernow's 2004 best-selling biography *Alexander Hamilton*, and Michael Dibdin's detective novel *Dead Lagoon*. Informally, we think of the popular subsample as a set of common books for which there is unlikely to be much of an upper tail of the consumer valuation distribution for two reasons: many can be purchased new in paperback on Amazon, which puts an upper bound on valuations[14]; and many potential consumers may be happy to substitute to some other popular book in the same category.

Table 1 reports summary statistics for title-level variables. Recall that for each title, we have one offline copy found at a used book store and matched online copies that give us information on, among other things, online price distributions. The average offline price (in 2009) for the books in our sample is \$11.29. One half of the titles were deemed to be of local interest to some location. The mean of the *Close* variable indicates that in a little more than three quarters of those with local interest, the location of interest is within 100 miles of the bookstore location. About 23% of titles are classified as *Popular*.

The table also provides some summary statistics on the online price distributions. In the contemporaneous 2009 data, the median online price for a title is on average well above the offline price we had found: the average across titles of the median online price is \$17.80

[14]Of the books mentioned above, Chernow and Dibdin's books are in print in paperback, whereas *The Frugal Gourmet* is not.

or a little more than 50% above the average offline price. But there is also a great deal of within-title price dispersion. The average minimum online price is just $9.27 and range between the maximum and minimum online price averages almost $100. To give more of a sense of how online and offline prices compare, the *PlaceinDist* variable reports where in the empirical CDF of online prices the offline price would lie. The average of 0.26 says that on average the offline price would be in the 26th percentile of the online distribution. Median online prices have not changed much between 2009 and 2012. There is, however, a moderate but noticeable decline in the lowest available online price and a substantial increase in the online price range.

Variable	Mean	St Dev	Min	Max
OfflinePrice09	11.29	21.11	1.00	250.00
LocalInterest	0.50	0.50	0	1
Close	0.74	0.44	0	1
Popular	0.23	0.41	0	1
MinOnlinePrice09	9.27	22.84	1.89	351.50
MedOnlinePrice09	17.80	23.87	2.95	351.50
MaxOnlinePrice09	108.16	488.38	5.00	8252
PlaceinDist09	0.26	0.28	0	1
NumList09(Max50)	23.55	17.72	1	50
MinOnlinePrice12	8.63	21.34	1.01	302.00
MedOnlinePrice12	17.67	25.46	1.95	302.00
MaxOnlinePrice12	184.23	1059.92	2.05	17,498
NumList12(Max50)	23.53	18.31	1	50

Note: Most variables each have 335 observations. *Close* is defined only for the 168 local interest titles.

Table 1: Summary statistics

3.1 Within-title price distributions

We would like to offer information on what these price distributions look like beyond just summary statistics. Showing 335 graphs would be impractical, so we have chosen three "typical"-shaped[15] online price distributions to examine more closely.

[15]To identify "typical"-shaped price distributions, we first performed a cluster analysis which divides the titles into three groups in such a way so that the characteristics of each title, in our case, the shape of the price distribution of each title, is closer to the mean for its group than to the mean for any other group. We then chose the title that was closest to the mean for each group as our "typical" title. We wanted to cluster titles by the basic shape characteristics of their price distributions, so we chose the following variables as the characteristics to cluster on: the log of the lowest, 10th, 20th, ..., 90th, and maximum prices, and the ratios of the 10th, 20th, and 30th percentile prices to the minimum price. The estimation uses Stata's "cluster kmeans" command which is a random algorithm, not a deterministic one. The number of elements in each cluster varied somewhat on different runs, but the identity of the "typical" element of each cluster appears to be fairly stable. We included the ratios in addition to the percentiles to increase the focus of

The left panel of figure 2 is a histogram of our first typical price distribution, that for *An Introduction to Massachusetts Birds*, a 32 page paperback published by the Massachusetts Audubon Society in 1975 that has long been out of print. There are ten online listings for this title. Most are fairly close to the lowest price – the distribution starts $3.50, $4, $5, $6, and $6.75 – but there is not a huge spike at the lower end. The upper tail is fairly thin with a single listing at $22.70 being more than twice as high as the second-highest price of $11.30.

The center panel of figure 2 is the histogram for the second typical price distribution, that for the hardcover version of *Alexander Hamilton* by Ron Chernow.[16] For this title there is a fairly tight group of six listings around the lowest price: the first bin in the figure consists of copies offered at $2.95, $2.95, $4.14, $4.24, $4.24, and $4.48. But beyond these, the distribution is more spread out with the largest number of offers falling in the $10 to $12.49 bin. There is also an upper tail of prices including six between $20 and $30 and four between $30 and $45. There is some correlation between price and condition: the four most expensive copies are all in "fine" or "as new" condition. But most of the upper tail does not seem to be attributable to differences in the conditions of the books. For example, the six copies between $20 and $30 include two copies in "poor" condition, two in "very good", one in "fine", and one in "as new", whereas five of the six copies offered at less than $5 are "very good" copies.

The right panel of figure 2 presents a histogram for the third typical distribution, that for *The Reign of George III, 1760 - 1815*, a hardcover first published by the Oxford University Press in 1960 as part of its Oxford History of England series.[17] Here again we see a concentration of listings close to the lowest price: one in very good condition at $10.99, a good copy at $12.03, and two poor copies at $12.09. But the lowest price is not nearly as low as it is for our other two typical distributions, and the rest of the distribution is also more spread out. There is a clear correlation between price and condition – the eight most expensive listings are all in very good condition or better. None are signed copies, but some might be distinguished by other unobserved characteristics such as being a first edition or having an intact dust jacket.

---

the clustering on the shape of the lower tail. Only titles with at least four online prices were included. The cluster analysis divided the titles into three groups containing 125, 109, and 69 titles.

[16] This book was a best-seller when released in 2004 and a paperback version was released in 2005. Both are still in print with list prices of $35 and $20, respectively.

[17] In 2013 it is again available new – presumably via a print-on-demand technology – at a very high list price of $175, but is currently offered by Amazon.com and BN.com at just $45. The rise of print-on-demand is a recent phenomenon and we assume few, if any, of the books in our sample were available via print-on-demand in 2009.

The fact that the third price distribution still includes a small group of sellers with prices very close to the lowest price suggests that some firms are competing to attract a segment of shoppers. The variation in condition among the low-priced listings suggests that book condition may not be very important to these consumers. The upper tail has some relation to condition, but mostly appears to be another example of price dispersion on the Internet for fairly homogeneous products.[18]

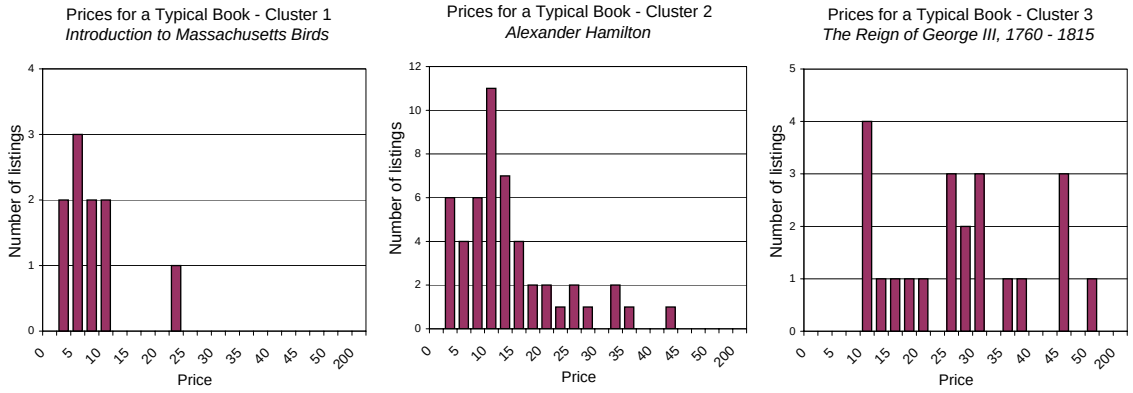

Figure 2: Online prices for three "typical" titles in 2009

4 Used Book Prices: Facts and Tests of Model Predictions

The purpose of this section is two-fold. We want both to present general descriptive evidence on prices of used books, establishing basic facts about this market, and also to perform reduced form tests of some more subtle predictions of our models, as enumerated earlier. The first subsection establishes the basic facts concerning price levels and price dispersion offline versus online. The next four subsections, comparing 2009 offline and online prices for the three different types of books and comparing 2009 and 2012 online prices, roughly correspond to the four empirical predictions we enumerated. 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.

[18] Although the three typical distributions we have examined here are of different types – the first is a nonpopular local interest book, the second is a popular book, and the third is a standard book – the clusters based on shape are not closely aligned with title types. For example, cluster 1 includes 42 standard titles, 41 nonpopular local interest titles, and 42 popular titles.

4.1 Basic facts

The introduction mentions a number of papers that have established somewhat counter-intuitive results 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 little 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 2 compares the mean of the offline prices with the average across titles of the mean online price for each title.[19] 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	11.21	Average across titles of average price	21.90
		Average within-title coefficient of variation	0.73
Variance of prices	448.59	Average of within-title variance	627.61
		Variance of mean prices across titles	670.81

Notes: 2009 online data were used for this table, and only those titles that had more than one online copy. Also, any copies listed at prices of $500 or more were omitted for these calculations.

Table 2: 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 0.73, i.e. the standard deviation is on average about 73 percent of the mean price.

We cannot directly compare offline and online price dispersion because our sample contains just a single offline price for each title.[20] 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

[19] We drop all listings with prices above $500 to avoid having the results heavily influenced by outliers.The calculation also omits nine titles for which we found only a single online price and hence cannot compute a within-title standard deviation.

[20] Our data set has this structure out of necessity – many of our books are relatively obscure and even if we had visited dozens of brick-and-mortar book stores we probably would not have found another copy.

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 less than 450 in our sample. Hence, the first term on the RHS, $\frac{1}{N}\sum_i \sigma_i^2$, must be smaller than 450 and is probably substantially smaller than this. The right column of the third row indicates that the average within-title variance in the offline data is over 600. Hence, this calculation suggests that in 2009 there was substantialy less price dispersion offline than online.

4.2 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 15.3. 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 more than four times as high 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.[21] 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.[22]

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

[21] 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.

[22] 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.

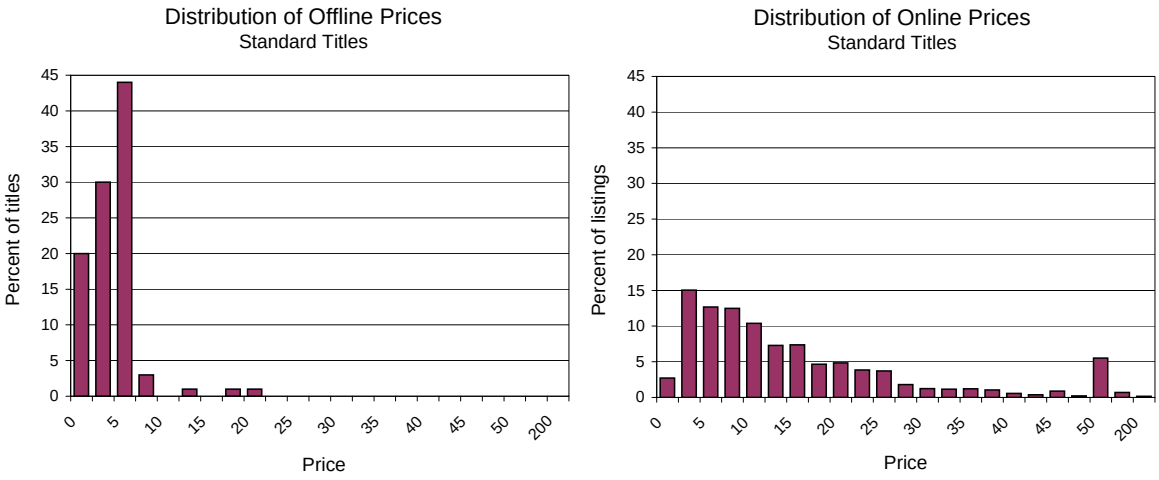

Figure 3: Offline and online prices for standard titles in 2009

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. The low ends of the price distributions suggest 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 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 each standard title.) 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 suggests that the match quality effect is much more important than the competition effect for these titles.

4.3 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 our 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

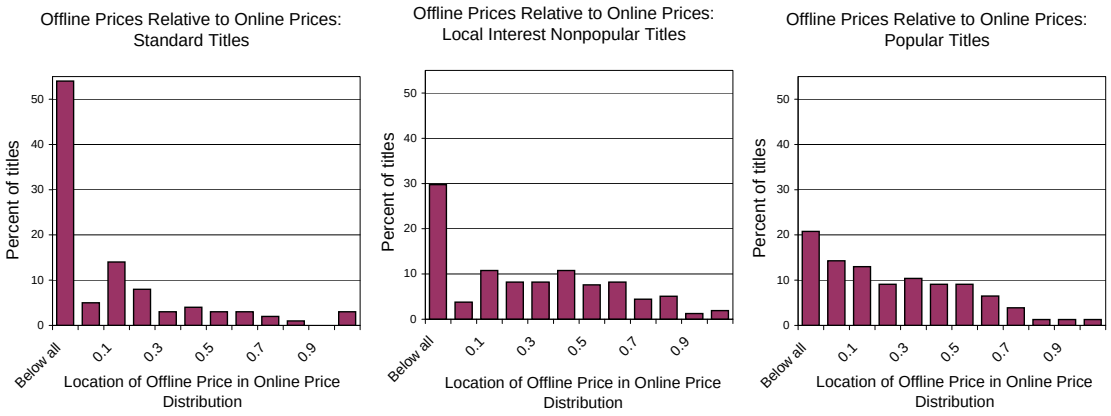

Figure 4: Offline prices relative to online prices for the same title

not be as pronounced. Intuitively, we argued that the highest-value match for a title like *The Mount Vernon Street Warrens: A Boston Story, 1860-1910*, *New England Rediscovered*, or *Boston Catholics: A History of the Church and Its People* might 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 difference is much smaller: the mean across titles of the mean online price is \$28.40. 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.

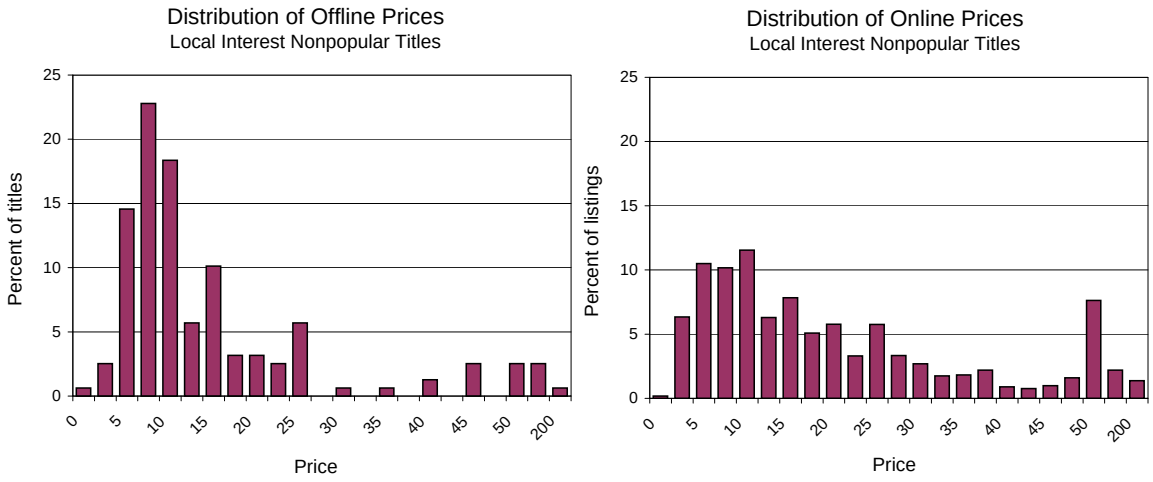

Figure 5: Offline and online prices for nonpopular local interest titles in 2009

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.

4.4 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 may be quite willing to switch from one detective novel and also sometimes have the option of simply buying a new copy of the book in paperback. 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.

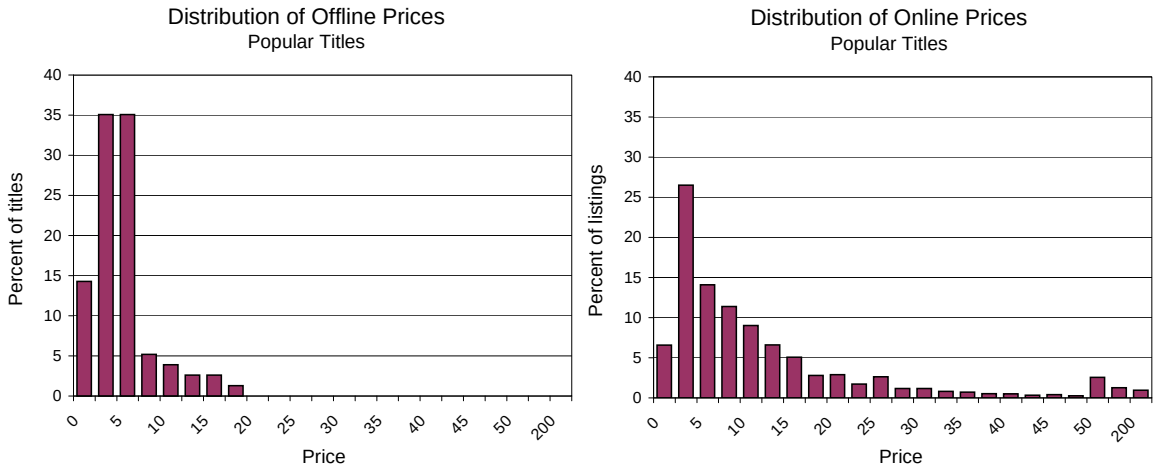

Figure 6: Offline and online prices for standard titles in 2009

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 prediciton 3: the online-offline price gap is narrower here than it was for standard titles.

4.5 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, e.g. Figure 1, we had noted that an increase in the number of 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 are mostly ignoring 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 of 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 fairly 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.

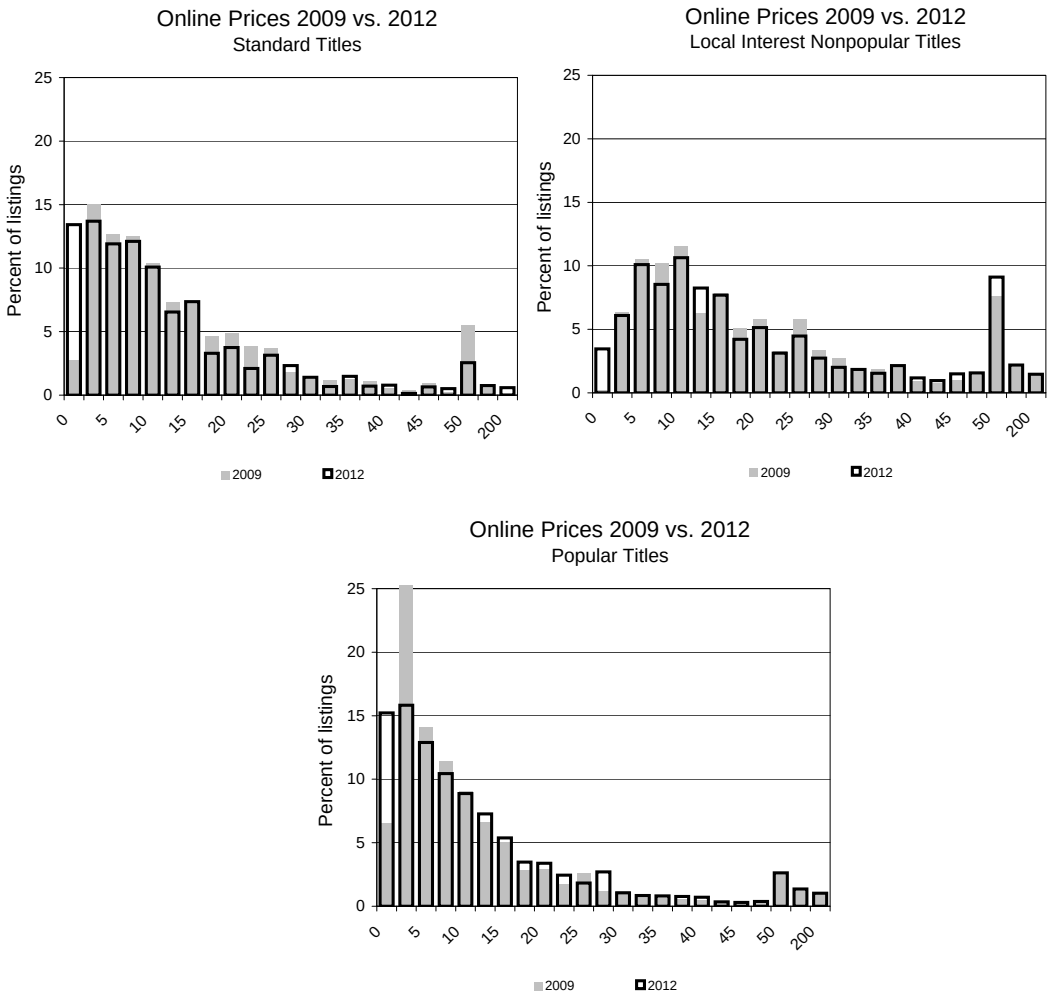

Figure 7: Comparison of 2009 and 2012 online prices

4.6 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 3 presents coefficient estimates from an OLS regression. The second column presents estimates from a Tobit regression which treats values of zero and 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, i.e. online prices are not as far above offline prices, for popular titles. The estimate from the Tobit model is larger at 0.21, i.e. offline prices are 21 percentiles higher in the online distribution 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)	0.06		0.06	

Table 3: 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 unobserved book characteristics.

5 Online Demand

In this section we present evidence on the demand for used books from online merchants. Our analysis in this section is descriptive, but we will note that some patterns we see appear consistent with online stores serving both shopper and nonshopper populations.

Although we do not have any sales data, we are able to see a proxy: whether a listing in our November 2012 sample is removed by January of 2013.[23] We presume that most books that disappear do so either because the book was sold (through AbeBooks or otherwise) or because the seller decided to withdraw from listing on AbeBooks. The latter is hard for us to detect when sellers have just one or two listings, but fortunately this is just a small fraction of listings, and sales rates are such that we can identify fairly well whether larger sellers exited simply by looking at whether all of their November 2012 listings have disappeared by January of 2013.[24] We drop all listings by the 32 firms with three or more listings who have all of their listings disappear. We presume that listings that are very far down on AbeBooks lists (i.e., have very high prices) are unlikely to sell through the website, and hence restrict our statistical analyses to a dataset of listings that were among the 50 lowest-priced listings on AbeBooks in 2012. This leaves a final estimation sample of 5282 listings for 318 titles. Summary statistics indicate that our disappearance rates for very small-scale sellers probably reflect a in part an exit rate which we have not been able to clean out. Mean disappearance rates for listings by sellers with just one or two listings in our sample are 10% and 12%, whereas the disappearance rate for listings by sellers with three to ten listings is about 6%.

[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. Given this matching strategy, we drop from this analysis all instances in which the same seller had multiple copies in a title-condition cell in November 2012, all print-on demand listings, all signed copies, and the very popular titles for which we did not collect all listings in our 2012 & 2013 data collections.

[24] For example, of the 203 sellers with exactly three listings in our November 2012 dataset, 168 have none of their copies disappear, 19 have one disappear, 2 have two disappear, and 14 have all three disappear. Given that only 2 firms sold two of their three listings, we presume that all or almost all of the 14 firms that had all listings disappear left the AbeBooks platform (or changed their name). Similarly, 11 firms with four or five listings have all of their listings disappear, whereas only one firm with four or five listings has all but one listing disappear.

Figure 8 presents histograms illustrating the relationship between listing-removal and order in which listings would appear when one sorts on shipping-inclusive price.[25] The left panel illustrates the relationship between listing-removal and item prices for standard titles. The x-axis gives the rank within the title of the shipping-inclusive price. The height of the bars indicates the average rate of disappearance for listings at that rank.[26] The figure suggests that sales rates are substantial for the lowest-price listings and that sales rates are quite price/rank sensitive. Listings in the top two positions disappear in the two-month span about one-third of the time. Listings in positions 3-5 disappear about 25% of the time. Disappearance rates fall to about 15% in the lower part of the top 10, and then appear to be 5-10% for titles in the second 10, although the rank-specific means are quite noisy by this point. The fact that demand drops of substantially after the first few ranks yet remains nontrivial even for the firm with the 20th lowest price seems consistent with our model's supposition that the firms serve both a price sensitive shopper population and a less price sensitive nonshopper population. Note also that a feature of the data collection may make lead us to underestimate price sensitivity at the lowest ranks: we only observe a firm's price rank on one particular date. If firms are churning through the low ranks, then firms with slightly higher prices on that date may well have slighly lower price at other times in the two-month period.

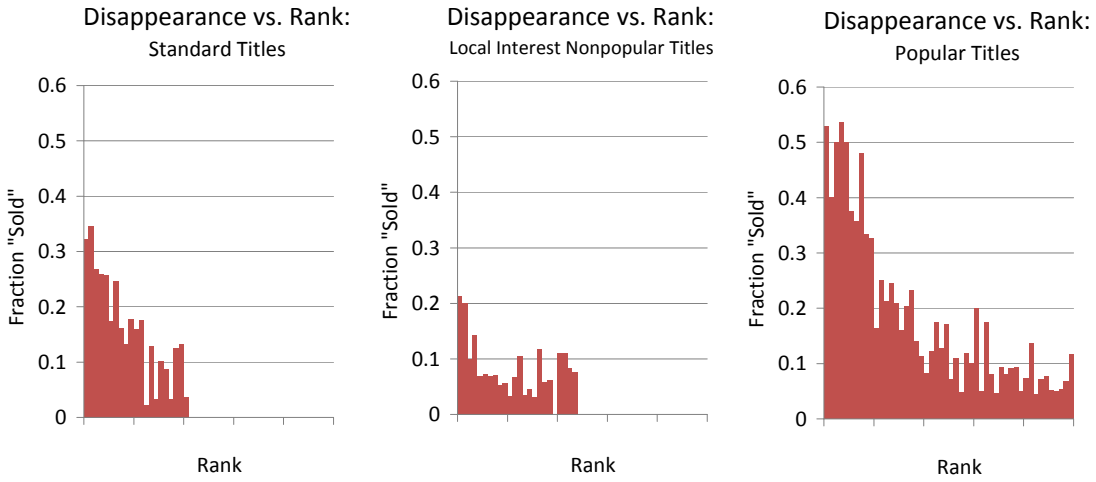

Figure 8: A demand proxy: removals of online price listings

[25] This is the default ordering on AbeBooks, but others are possible as well. Our orderings may not match what a consumer would have seen when prices (in dollars and cents) are identical.

[26] We have cut off the figures where the sample size falls below 25 because the estimates of the (small) disappearance probabilities become very noisy by that point. In this case, the figure presents data for ranks 1 through 21.

The middle panel presents a similar histogram for nonpopular local interest titles. Sales rates are somewhat lower for these titles. Disappearance rates are about 20% at the top two ranks, about 10-15% at the next two ranks, and then fall to 5-10% for the remainder of the ranks. In a shopper-nonshopper model these patterns might be interpreted as suggesting that there are not as many shoppers and only a few of the firms are competing for their business.

The right panel presents results for popular titles. Here, disappearance rates are substantially higher. They start at close to 50% for listings in the top 5 positions. Beyond this point they drop off fairly sharply: they are around 30-40% in ranks 6-10, a little over 20% in the next 10, and a little over 10% for listings with ranks in the 21-30 range. Sample sizes do not drop off as rapidly here, so we have extended this figure out through rank 50. The somewhat odd shape of the leftmost part of the graph – being roughly flat for the first few ranks and then dropping off rapidly – may agan reflect the difference between our ranks and what the typical consumer saw. We do not know how listings with identical prices were sorted by AbeBooks, and there also may be a great deal of churning in the price order among the top-ranked titles.[27] But again the fact that sales rates level off beyond some rank suggest that some sales are being made to a less price sensitive subpopulation.

One descriptive observation is that large professional sellers play a big role on AbeBooks.com. The majority of the 2067 online retailers in our November 2012 dataset have listings for just one or two of our 318 titles, but these small-scale sellers only account for 15% of the total listings.[28] At the other extreme, 45 firms we will refer to as "power sellers" have listings for more than 25 of the titles, including 16 firms that have listings for more than 50. The power sellers play a particularly big role at the top (i.e., lowest prices) of AbeBooks.com's lists. In our November 2012 data, 55% of the titles have a power seller in the top position and the proportion of power sellers remains above 30% in each of the top 25 slots. Power sellers sell books at a substantially higher rate – over 30% of their listings disappear – but in part this reflects that they sell many popular books and set low prices.

Table 4 presents statistical analyses that provide additional detail on the patterns. The first column presents estimates of a logit model in which the probability that a listing will disappear is a function of the listing's rank and condition. (It also includes two title-level

[27] Another potential bias is that disappearances could underestimate sales if a seller replaces a sold copy of a book with another copy in the same condition.

[28] These statistics are for only 318 of the millions of published titles, so even sellers that appear to be small-scale merchants in our data may have many, many listings on AbeBooks. We examine a smaller set of titles here than in previous sections because we omit very popular titles (those with more than 300 listings) for which we did not collect complete listings data.

controls and one seller-level covariate, the number of distinct titles in our full dataset for which the seller has at least one listing.) The coefficient on $\log(Rank)$ is highly significant. Its magnitude implies that increasing $\log(Rank)$ by one unit, e.g. moving from rank 1 to rank 2.7 or rank 2.7 to rank 7.3, is associated with a 58% decrease in the probability of disappearance. This result suggests that demand is highly elastic when prices are tightly bunched. In the second column we include both the rank of the listing and the shipping-inclusive price. The two are fairly collinear, but, nonetheless, each is highly significant in this regression. The magnitude of the rank effect is somewhat smaller – increasing $\log(Rank)$ by one unit is now associated with a 39% decrease in the disappearance rate. But this is augmented by the (fairly modest) price effect – a 10% increase in price is associated with a 4.6% decrease in the disappearance rate. The modesty of the price effect may, of course, reflect that this is not a structural demand estimation. The slope of this curve should be less than that of a structural demand curve because disappearances reflect both sales through the search engine and outside of it, and firms would be expected to choose relatively high prices when they have a high outside sales.

The *Condition* variable is also significant in both versions. It indicates that a listing that is one condition better, e.g. from "good" to "very good," will have a about a 10% higher disappearance rate. The coefficient on $\log(Storetitles)$ is also significant, reflecting that power sellers appear to sell many more copies even after we control for rank/price differences and the title popularity.[29]

The third and fourth columns present estimates from models which instead use title fixed effects to control for differences across titles.[30] The estimated coefficients on the price effects are similar, with the point estimate on the price effect a little larger and that on the rank effect a little smaller.

The final three columns reestimate this model on three subsamples: standard, nonpopular local interest, and popular titles. In the smaller samples it is harder to separately identify the price and rank effects, and not all coefficients are significant. The disappearance-price rank relationship seems to be somewhat weaker, and book condition appears to matter less for standard titles. Power sellers do better in all three samples, but the coefficient on $\log(Storetitles)$ is no longer significant in the local interest subsample.

Importantly, these demand patterns are broadly consistent with our priors and intuition about this market as well as predictions of our model. In particular, a market with two

[29] These specifications use just two controls for popularity: the number of listings for the title in the full dataset and the log of the lowest price at which the title is offered.

[30] 121 of the 318 titles in the previous regression are dropped because none or all of the listings disappear.

types of consumers, such as described in our model, would likely display a high degree of price sensitivity at low ranks, where shoppers operate, and a sharply decreased degree of price sensitivity at higher ranks, where nonshoppers would often operate.

	Dependent Variable: *Disappear*						
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log(Rank)	-0.67	-0.44	-0.77	-0.32	-0.10	-0.51	-0.27
	(0.05)	(0.06)	(0.05)	(0.09)	(0.17)	(0.21)	(0.12)
log(Price)		-0.53		-0.92	-0.74	-0.55	-1.54
		(0.10)		(0.15)	(0.23)	(0.31)	(0.25)
Condition	0.10	0.13	0.12	0.17	0.01	0.17	0.29
	(0.04)	(0.04)	(0.04)	(0.04)	(0.07)	(0.10)	(0.07)
log(StoreTitles)	0.39	0.31	0.34	0.25	0.30	0.13	0.22
	(0.04)	(0.04)	(0.04)	(0.04)	(0.08)	(0.09)	(0.07)
Constant	-2.90	-2.10					
	(0.34)	(0.37)					
log(MinPrice)	-0.40	0.02					
	(0.09)	(0.12)					
log(TitleListings)	0.43	0.25					
	(0.06)	(0.07)					
Title fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Titles included	All	All	All	All	Stnd.	Local	Popular
Number of obs.	5282	5282	4330	4330	1119	1137	2074
Pseudo R2	0.135	0.142	0.143	0.157	0.133	0.122	0.203

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 4: Logit estimates of the listing disappearance process

6 A Structural Model of the Market

In the theoretical model of section 2, online prices differ from offline prices for two offsetting reasons: increases in consumer arrival rates and distributions of valuations tend to increase online prices; and increased competition brings online prices down. The welfare effects of online sales depend on the magnitudes of the underlying gross effects: the increase in consumer arrival rates; the increase in consumer valuations; the equilibrium delay before sales occur; etc. In this section we develop and estimate a structural model which includes all of these effects. We note that features of the model make it amenable to estimation via simulated maximum likelihood. Our estimates indicate that customer arrival rates were substantially higher online than offline by 2009 and 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 $I+1$ populations of consumers shop for each title k at stores $i = 1, 2, \ldots, I$.[31] 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 $D_k^o(p_{ik}, p_{-ik}; X_{ik}, \Lambda, \beta_{ik})$, where X_{ik} is a vector of characteristics of store i and title k, Λ is a vector of parameters to be estimated, and β_{ik} is a vector of unobserved random coefficients. Assume that the arrival rate γ_{0k} and random coefficient vector β_{ik} are draws from distributions that may depend on the parameter vector Λ.

The other I 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 $D_{ik}^m(p_i; X_{ik}, \Lambda, \beta_{ik})$. Again, the γ_{ik} and β_{ik} are 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 $(p_{1k}, p_{2k}, \ldots, p_{Ik})$. For other titles we observe both prices and an indicator for whether the title sells in a given time period: $(p_{1k}, q_{1k}, \ldots, p_{Ik}, q_{Ik})$. We wish to estimate the parameter vector Λ.

One observation about this model that will facilitate estimation is that the first order condition for store i's title k price to be optimal,

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

[31] The number of sellers varies by title, but we omit the k subscript on I in this description for simplicity.

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

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

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

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

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. Given the observed p_{ik} and an inferred γ_{ik}, the "net" 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 probabilty that the book will be sold in a Δt time interval is

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

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 any random coefficients (which may vary by title or by listing). Given these random coefficients, we compute the joint likelihood of each observed price vector $(p_{1k}, \ldots, p_{Ik})$ and of each observed price/quantity vector $(p_{1k}, q_{1k}, \ldots, p_{Ik}, q_{Ik})$ 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

To estimate arrival rates and demand in the used book market and explore how they have changed with the shift to online sales, we implement a parsimonious version of the model including only as many parameters and random coefficients as was necessary to estimate quantities of interest and match the main features of the data.

We assume that the arrival rate of shoppers varies only with the "popularity" of a title and the year (2009 versus 2012):

$$\gamma_{0k} = \gamma_0 Popularity_k^{\gamma_0^N} \Delta\gamma_0^{12\ I(t=2012)}.$$

The $Popularity_k$ variable is the ratio between the count of listings for title k in the 2012 online data and the mean of this count across listings. We assume that shoppers have logit-style preferences: consumer j is gets utility

$$u_{ijk} = \begin{cases} X_k\Lambda - \alpha\delta_k p_{ik} + \epsilon_{ijk} & \text{if } j \text{ purchases title } k \text{ from store } i \\ X_k\Lambda + \beta_{0k} + \epsilon_{ijk} & \text{if } j \text{ does not purchase title } k \end{cases},$$

where the ϵ_{ijk} are independent random variables with a type 1 extreme value distribution. The demand for firm i's offering is then

$$D_k^o(p_{ik}; p_{-ik}) = \frac{e^{-\alpha\delta_k p_{ik}}}{e^{\beta_{0k}} + \sum_\ell e^{-\alpha\delta_k p_{\ell k}}}.$$

The parameters δ_k let the price-sensitivity vary across titles, which will help the model fit data in which price levels vary substantially across titles. We will adopt a random coefficients specification which allows the unobserved outside good utilities β_{0k} to be normally distributed across titles. This feature helps to fit data in which the fraction of listings that sell varies substantially across titles and also helps the model explain why one store sometimes substantially undercuts all other stores.

The arrival rate of nonshoppers is similarly allowed to vary with popularity, year, and whether a store is online/offine. We assume that it also varies randomly 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, the arrival rate is

$$\gamma_{ik} = z_{ik} Popularity_k^{\gamma_i^N} \Delta\gamma_i^{12\ I(t=2012)} \Delta\gamma_i^{\text{off}\ I(\text{offline})},$$

where the z_{ik} are *i.i.d.* gamma-distributed random variables with mean μ_{γ_i} and standard deviation σ_{γ_i}. We assume that nonshopper demand curves have the constant elasticity

form. In utility terms, this amounts to assuming that a nonshopper j considering buying book k from store i gets utility

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

where the v_j are heterogeneous across consumers with density $f(v_j) = h\eta v^{-\eta-1}$ on $[h^{1/\eta}, \infty]$. We assume that $\eta > 1$ (otherwise the monopoly price is infinite) and that $h > 0$ is sufficiently small so that all observed prices are in the support of the value distribution. With this assumption the probability that a shopper purchases at the observed price is

$$D^m(p_{ik}) = h(\delta_{ik} p_{ik})^{-\eta}.$$

The model description above has more parameters than we are able to estimate. There are two issues of identification: the arrival rate γ_{ik} of nonshoppers cannot be separately identified from the multiplicative constant h in nonshopper demand; and all profit functions only depend on ratios γ/r. Also, as a practical matter, we are unable to estimate the large number of title fixed effects δ_k that shift the level of equilibrium prices for each title. Accordingly, we have chosen to fix some parameters in our estimation. We fix $r = 0.05$ so that arrival rates should be thought of as arrivals over a period of one year. We set the constant h in the shopper demand equation to one.[32] And, finally, we implicitly fix the δ_k at a different value for each book by scaling the prices for each book so that the lowest 2009 online price is equal to one and then setting each δ_k to one.

These assumptions produce a model with twelve parameters to be estimated: ($\gamma_0, \Delta\gamma_0^{12}$, $\gamma_0^N, \alpha, \mu_\beta, \sigma_\beta, \mu_{\gamma_i}, \sigma_{\gamma_i}, \Delta\gamma_i^{12}, \gamma_i^N, \eta, \Delta\gamma_i^{\text{off}}$). We estimate these parameters on a dataset containing 313 books which had valid listings in all four of our data collection waves: they were found in an offline store in 2009 and had listings successfully scraped from AbeBooks.com in all three online collections.[33] The demand data q_{ik} are inferred by comparing the November 2012 and January 2013 listings, and are used only for books for which our 2013 data collection was complete. The estimation follows the procedure noted above with the "outside options" β_{0k} as the only random coefficients.[34]

[32] We treat demand as being $p^{-\eta}$ even when this expression is greater than one and hence inconsistent with the demand being a probability of purchase given arrival. Note, however, that the same equations could always have be made consistent with the probability interpretation simply by choosing a smaller value of h and scaling up the consumer arrival rate while keeping their product constant.

[33] For our structural analysis we also drop all listings priced above $50 and any title that does not have at least two listings.

[34] With just one random coefficient relevant to each title, the "simulated" maximum likelihood becomes just a numerical integration over the unknown coefficient. We perform this integration by evaluating the likelihood for 50 values of the random coefficient spaced evenly in CDF space. For some parameters the model

6.3 Estimates of arrival rates and demand

Table 5 presents estimates of the parameters of our structural model. The left half of the table reports estimates related to the nonshoppers. The most basic finding is that the "net" arrival rate of nonshoppers (arrival rate of a nonshopper who actually purchases) at online stores appears to be quite low. The $\hat{\mu}_{\gamma_i} = 0.054$ estimate indicates a firm with a typical listing should expect that a nonshopper willing to purchase the good at the lowest price in 2009 online data for that title will arrive approximately once every 18.5 years. The $\hat{\sigma}_{\gamma_i} = 0.04$ estimate indicates that there is variation across stores, with some having essentially no nonshoppers and others having substantially more. The $\hat{\gamma}_i^N = 0.14$ estimate indicates that arrival rates are higher for more popular titles, but it is not a very big effect. For example, a title that has twice as many listings as average has about a ten percent higher (per listing) arrival rate of nonshoppers. The year dummy indicates that nonshopper arrival rates did not change between 2009 and 2012. All of these estimates are fairly precise. Intuitively, the 2012 quantity observations for high-priced firms provide a lot of information about nonshopper arrival rates, and the fact that the upper parts of the price distributions are so similar in 2009 and 2012 drives the estimate that nonshopper arrival rates must be similar in the two years.

Parameter	Coef.Est.	SE	Parameter	Coef.Est.	SE
Nonshopper arrival			Shopper arrival		
μ_{γ_i}	0.054	(0.003)	$\ln(\gamma_0)$	3.70	(2.68)
σ_{γ_i}	0.04	(0.003)	$\Delta\gamma_0^{12}$	1.06	(0.49)
$\Delta\gamma_i^{12}$	1.02	(0.04)	γ_0^N	1.23	(0.20)
γ_i^N	0.14	(0.02)			
Offline arrival 2009			Shopper utility		
$\Delta\gamma_i^{\text{off}}$	0.36	(0.03)	μ_β	-8.21	(2.31)
Nonshopper utility			σ_β	3.80	(0.34)
η	1.15	(0.01)	α	13.34	(1.42)

Table 5: Estimates of structural model parameters

Although nonshopper arrival rates at online stores are modest, the coefficient directly below this indicates that they are still substantially higher than arrival rates at offline stores. The 0.36 estimate for $\hat{\Delta\gamma_i^{\text{off}}}$ indicates that 2009 offline arrival rates were only about

cannot rationalize some observations (in which one firm sets a price substantially below the second lowest price) using any positive γ_{ik}). When this happens we use a penalty function, $L(p_{ik}|X_{ik}, \Lambda) = e^{-10+10|\gamma_{ik}|}$ that increases in the distance between the (negative) γ_{ik} that would make the first-order condition for profit-maximization satisfied and the nearest value that is in the support of the distribution (which is zero).

36% of online arrival rates, or about 0.02 consumers per year for the typical listing.[35] Such a difference in arrival rates would lead to substantial differences in the expected welfare gains produced by an eventual sale. Of course, it should be kept in mind that, unlike in the online world, these arrival rates are estimated without any quantity data and instead reflect just that lower offline prices can be rationalized in our model only by a lower consumer arrival rate. Despite that caveat, we do not find it at all implausible that a lot of used books sat on the shelves of used book dealers for years waiting for an interested buyer to happen by.

The final estimate on the left side is an estimate of the elasticity of demand of the nonshoppers. The estimated elasticity of demand is -1.15. A constant elasticity demand curve with this elasticity has a very thick upper tail. Such a situation will lead to estimates that each sale to a nonshopper generates a great deal of consumer surplus.

The right half of the table reports estimates related to the population of online shoppers. When interpreting arrival rates, keep in mind that "net arrival rates" will be lower than the arrival rate coefficient because some consumers prefer the outside good. We have allowed for unobserved heterogeneity across titles in the net shopper arrival rate by introducing heterogeneity in the utility of the outside option. The estimated mean outside good utility turns out to be substantially higher than the mean utility provided by the lowest-priced listing for each title, so for the median title, only a little over 1% of arriving shoppers will actually purchase from anyone. The estimated variance of the outside good utility indicates that there is a great deal of unobserved heterogeneity across titles. About 9% of titles have outside good utilities that are below the mean utility provided by the lowest-priced listing, in which case most shoppers will purchase. At the other extreme, a large number of titles have essentially no shoppers willing to purchase at the observed prices.

The estimated shopper arrival coefficient ($e^{3.70} \approx 40$) indicates that some titles of will have a net arrival rate of nearly 40 consumers per year.[36] This is dramatically higher than the rate at which nonshoppers are estimated to arrive. As noted earlier, however, it is only a few titles that have such shopper net arrival rates. The median title is estimated to have a shopper net arrival rate that is roughly ten times each firm's nonshopper arrival rate, and other titles are estimated to have very few shoppers at all. These estimates reflect a basic fact about demand noted earlier: we observe many sales of some titles, but zero

[35] Note, however, that prices and demand are scaled so that a potential consumer buys with probability one if he sees a price equal to the lowest 2009 online price. Offline firms that set prices below the lowest online price will sell at a somewhat higher probability.

[36] In this paragraph we use "net arrival rate" to refer to the arrival rate of shoppers willing to purchase from a store offering the lowest price observed in the data for that title.

sales over a two-month period for about one-third of the titles in our sample. Note also that the standard error on the estimate of $\ln(\gamma_0)$ is very large. This reflects a collinearity between the arrival rate and the outside good utility – it is hard to tell whether there is a large arrival rate of consumers who are each unlikely to purchase or a lower arrival rate of consumers who are each likely to purchase.

The $\hat{\gamma_0}^N = 1.23$ estimate indicates that the number of shoppers increases more than proportionately to a title's popularity. Note that the fraction of consumers who are shoppers is roughly independent of popularity because the coefficient estimate for nonshoppers is for a single store, whereas the coefficient estimate for shoppers is for the whole population. The estimates do imply, though, that the potential benefit of attracting shoppers is a more important consideration for stores selling popular titles.

Unfortunately, we are unable to say much about how Amazon's incorporation of AbeBooks's listings affected the shopper arrival rate. The imprecise estimate of $\hat{\Delta\gamma_0}^{12}$ indicates an increase of arrivals of $6\% \pm 49\%$. Hence, the 95% confidence interval includes both a doubling of the shopper arrival rate and substantial decreases. We can reject larger increases in the shopper arrival rate.

The price coefficient $\hat{\alpha} = 13.34$ indicates that online shoppers are very price sensitive. In most cases (excluding, for example, when the "outside option" is unusually bad), this estimate implies that a firm with a price close to the lowest listed price will see its demand go down by about 13% if it raises its price by 1%. A consequence of this price sensitivity coupled with our logit functional form assumption is that a standard welfare calculation will indicate that shoppers who purchase the book do not get a great deal of consumer surplus. In principle, one could avoid this implication by putting heterogeneity directly into the the shopper arrival rate and estimating a more flexible demand function that allowed consumers to have different price sensitivities among inside goods versus the outside good. But in practice, our data do not make it possible to estimate such a flexible specification.

6.4 Estimating welfare gains

In this section we examine both welfare gains from a shift from offline to online used book sales and the division of these gains between consumers and stores. Welfare gains occur when books are sold to consumers with higher valuations and/or books are sold more quickly. Online retailers must be better off if the number of nonshoppers (and the distribution of their valuations) increases, but the magnitude of the gain will be affected by the size of these increases and the relative sizes of the shopper and nonshopper populations.

Given an estimated parameter vector $\hat{\Lambda}$, the average per-listing welfare generated by the listings for title k can be calculated by integrating over the posterior distribution of the unobserved random coefficient β_{0k}:

$$E(W_k|\hat{\Lambda}) = \frac{1}{I_k}\int_\beta \left(E(CS_k|\hat{\Lambda},\beta) + \sum_i E(\pi(p_{ik},p_{-ik})|\hat{\Lambda},\beta) \right) f(\beta|p_{1k},\ldots,q_{nk}\}d\beta,$$

where I_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 is the discounted expected profit that the firm listing copy i will earn.

The profit term can be computed using the same profit functions we use in estimating the model. Given the price p_{ik}, a value β_{0k} for the outside good utility, and the other estimated parameters, we back out a value for γ_{ik}. Expected profits are then simply

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

Consumer surplus is a little more complicated. It is a sum of consumer surplus from sales to nonshoppers and sales to shoppers. Expected total consumer surplus from sales to nonshoppers can be calculated similarly to how we calculated profits:

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

Expected total consumer surplus from sales to shoppers can be calculated similarly although the final term, expected consumer surplus conditional on purchase, will be different:

$$\begin{aligned} E(CS_k^{ns}) &= \sum_i E(e^{-rt_i})\text{Prob}\{i \text{ sells to a shopper}\}E((-\alpha p_{ik}+\epsilon_{ik}) - (\beta_{0k}+\epsilon_{0kj})|\text{j purchases from i}) \\ &= \sum_i \frac{\gamma_{0k}D^o(p_{ik},p_{-ik})}{r+\gamma_{ik}D^m(p_{ik}) + \gamma_{0k}D^o(p_{ik},p_{-ik})} E((-\alpha p_{ik}+\epsilon_{ik}) - (\beta_{0k}+\epsilon_{0kj})|\text{j purchases from i}) \end{aligned}$$

We compute this via a monte carlo approach.

Recall that in the case of isoelastic demand and no shoppers, expected welfare will simply be the price, but these formulae tell us how this welfare is divided. They are also more general, applying the cases with arbitrary demand and shoppers.

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. An increase in the number of listings between 2009 and 2012 may have lead to an additional welfare increase, although per-listing profits are estimated to have declined somewhat from the 2009 level.

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. We use the 313 titles that have valid listing in both 2009 and 2012. (These are "gross" profits in that they do not account for the acquisition cost of the books being sold.) The first column indicates that average per-listing profits are estimated to have been fairly low in the offline world, just \$1.32 per listing. This is the product of the mean price for the titles, \$10.21, and a discount factor of $E(e^{-0.05t})$, reflecting the fact that sales occur probabilistically in the future. In particular, the estimated discount factor of 0.13 reflects our estimates that many books would take decades to sell: offline arrival rates are about 0.02 customers per year.

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 \$3.11. 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 three-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.

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 and price-weighted waiting times to sell are intermediate. The waiting time effect reflects that we are looking at a price-weighted average. The 2012 data contain more listings offered at very low prices and it is primarily these titles that get purchased by the shoppers. As a result, the average-priced books take a long time to sell.

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 similar upper tails suggest that nonshopper arrival rates were similar in 2009,

Average value per listing	2009 offline listings	2009 online listings	2012 online listings
Gross profit	**$1.32** (0.07)	**$3.11** (0.10)	**$2.29** (0.10)
Price × Discounting	$10.21 × 0.13	$15.04 × 0.21	$13.80 × 0.17
Consumer surplus	**$8.90** (0.09)	**$16.61** (0.85)	**$13.63** (0.50)
Nonshoppers + Shoppers	$8.85	$16.01 + $0.60	$13.36 + $0.27
Welfare	**$10.21** (——)	**$19.73** (0.87)	**$15.91** (0.49)

Table 6: Profit and welfare estimates

and the demand model infers that the growth of low-priced listings would have reduced the number of sales that high-priced firms make to shoppers. 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. The 2009 offline estimate follow mechanically from our theorem about optimal pricing with constant elasticity demand. Estimated social welfare is simply the average price of $10.21 per listing. Given the estimate that profits are only $1.32 per listing consumer surplus must make up the difference and hence is very large.

The consumer surplus figures for 2009 and 2012 online sales are not purely mechanical: the welfare result was only for the monopoly model and does not extend to the model with a shopper population.[37] The estimates here indicate that consumers benefitted substantially from the shift to online sales: 2009 online listings are estimated to generate, on average, almost twice the consumer surplus (per listing) as 2009 offline listings. As noted in the introduction, the naive intuition that higher online prices suggest that consumers are not benefitting from online sales misses the basic point that profits and consumer surplus can both be higher if Internet sales result in higher match quality and faster sales. The 2012 estimates are that consumers receive most of the social surplus in that period as well.

[37] In each case our estimate of social welfare is substantially higher than the average price: $19.72 vs. $15.04 in 2009 and $15.91 vs. $13.80 in 2012.

Intuitively, our estimate that consumer surplus is so much higher than profits reflects the distribution of prices. The model rationalizes the coexistence of low and high prices via a demand curve with a very thick upper tail, which then makes the average valuation of consumers who purchase very high. The thickness of the uppermost part of the tail, of course, relies on functional form assumptions, and one could also worry that some high prices are not actually profit-maximizing. For this reason, the precise magnitudes here should be taken with a grain of salt. Another limitation of our model, mentioned earlier, however, works in the other direction. In assuming a simple logit specification for demand, we have assumed that consumers who are very price-sensitive when comparing online listings are equally price sensitive in comparing listings to the outside option of not purchasing. This feature biases us towards a finding of relatively little consumer surplus. Although it would be preferable to estimate a model with more flexible substitution patterns, such as a nested logit model with the outside good in a separate nest if it were feasible, we are perhaps fortunate that these biases work in opposite directions.

The third row of the table presents estimates of total welfare. Our model of 2009 offline sales is a version of the monopoly-constant elasticity model discussed in section 2.1. As a result, the estimated welfare is simply the average price (and is independent of the estimated parameters). Our empirical models for the 2009 and 2012 online markets are not just monopoly models, so the welfare estimates will depend on the estimated parameters. The estimates are that per-listing welfare was almost twice as high in the 2009 online market as in the 2009 offline market. Welfare is somewhat lower in 2012 than 2009 reflecting the increased sales to nonshoppers and the somewhat lower prices.

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 to provide evidence in its favor we devoted a substantial part of our paper to developing less obvious implications that could be examined to help assess its relevance. We examined these implications using three sources of variation. First, we examined how price distributions – rather than just price levels – differ between the online and offline markets. Here, our primary supporting 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 growth of the upper tail is much less dramatic popular books (which one would expect if the valuation distribution had less of an upper tail). Our favorite characterizations 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 assumptions of our model: there is a concentration of demand among the top-ranked firms as one would expect from a price-senstive shopper segment; but firms with much higher prices also appear to have some probability of making a sale.

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. And consumer surplus and welfare are easily calculated from the estimated parameters by computing some things on a title-by-title basis and others on a consumer-by-consumer basis. 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. Amazon's subsequent incorporation of used-book listings seems to have reduced book dealer profits somewhat, but they are still much higher than they were in the offline world.

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.

References

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

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.

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

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.

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.

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.

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): "Estimating Demand in Search Markets: the Case of Online Hotel Bookings," *Rand Journal of Economics*, forthcoming.

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

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*, forthcoming.

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.

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

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