Price Beliefs and Experience: Do Consumers' Beliefs Converge to Empirical Distributions with Repeated Purchases?*

Brett Matsumoto† and Forrest Spence‡

March 23, 2016

Abstract

We use data on consumers' subjective beliefs about the distribution of prices in an online marketplace to investigate two questions of interest. First, when consumers face price uncertainty, to what extent do their beliefs about the distribution of prices reflect the actual empirical price distribution? Second, do consumers learn about features of the empirical distribution through experience? Using reported expectations for online textbook prices from a survey of 1,224 college students, we find that consumers with no prior experience in purchasing textbooks online tend to expect online prices to be higher than what is observed empirically. However, consumers with more experience in the marketplace generally have more accurate beliefs about the price distribution, which is consistent with learning.

*All views and opinions expressed are our own and do not reflect those of the Bureau of Labor Statistics or the U.S. Federal Government. We would like to thank the editor, William Neilson, a very helpful referee, Meta Brown, Liran Einav, Donna Gilleskie, Robert Hammond, Brian McManus, Tiago Pires, James Roberts, Andrew Sweeting, Chad Syverson, and Andy Yates for helpful comments. All remaining errors are our own.

†Bureau of Labor Statistics; matsumoto.brett@bls.gov

‡University of Notre Dame; fspence@nd.edu

1 Introduction

Price dispersion is a feature of many markets and even occurs in markets for homogeneous goods or services (Stigler, 1961). One possible reason for the persistence of price dispersion is that consumers have limited information over prices and acquiring information may be costly. In markets with limited information and costly search, an individual may not purchase from the seller with the lowest price if she is unaware of that price. Theoretical models of consumer search incorporate the search decision into a model of consumer demand by assuming that individuals have beliefs about the empirical distribution of prices in the market and must incur a cost to reveal price information from one or more retailers before deciding whether to purchase the good or service (e.g., Reinganum, 1979; Burdett and Judd, 1983). The decision to search depends upon the magnitude of the search costs as well as the individual's subjective beliefs about the distribution of prices. When estimating models of consumer search, researchers may impose assumptions on individuals' beliefs in order to estimate search costs. In this paper, we test the validity of these assumptions using data on the observed distribution of prices for the online textbook market and data on individuals' subjective beliefs about this distribution.

There is a growing literature focusing on the development and estimation of structural models of consumer search. These models have been used to explain observed price dispersion for homogeneous goods (Hortaçsu and Syverson, 2004; Hong and Shum, 2006), test competing models of consumer search (De los Santos et al., 2012; Honka and Chintagunta, 2015), separately identify search costs from other market frictions such as switching costs or learning (Dube et al., 2010; Honka, 2014), and investigate firm pricing behavior and recover demand estimates in markets where price uncertainty and search frictions are important (Giulietti et al., 2014; Koulayev, 2014; Moraga-González et al., 2015; Pires, 2015; Seiler, 2013). A critical assumption used in these studies is that consumers have rational expectations (i.e., the price of a product is a random variable, but consumers know the parameters that govern the distribution of prices). However, if consumers have biased beliefs about the parameters of the empirical distribution of prices, this will lead to biased estimates of search costs. In particular, if consumers' beliefs about prices are biased upward, the rational expectations assumption will bias search cost estimates upwards and bias price elasticity estimates towards zero (low levels of search can be explained by either high search costs or low expected benefits from search). By comparing subjective beliefs to actual observed price

distributions, we are able to test the validity of this assumption.

In addition to testing the validity of the rational expectations assumption, we also investigate the degree to which experienced consumers have more accurate beliefs than their less experienced counterparts. Recent research has supported this idea by incorporating learning into consumer search models.[1] In these models, consumers learn about characteristics of the products offered within a single purchasing decision through a sequential search process (De los Santos et al., 2013; Koulayev, 2009, 2013). We focus instead on learning across purchasing decisions; in particular we examine the hypothesis that more experienced consumers have acquired information about the empirical price distribution through repeated participation in the market.[2]

We collect data on the empirical distribution of textbook prices from online retailers and consumers' subjective beliefs about this distribution. In order to obtain data on individuals' subjective beliefs, we provide an online questionnaire to 1,224 undergraduate students with multiple textbook purchasing scenarios in order to elicit their beliefs about prices. For each hypothetical textbook purchasing scenario, students are given the price of a textbook from the campus bookstore and are asked about their expectations of the lowest price available from an online retailer. Additional questions are asked to elicit consumers' beliefs about the variability of the lowest price. For example, if a consumer reports an expected online price of $100, she is then asked about the likelihood that the actual price is below $95.

Our results show that inexperienced consumers have price expectations that are significantly greater than the mean of the empirical price distribution. Therefore, we can reject the hypothesis that inexperienced consumers know the parameters of the price distribution for the online textbook market. Individuals with higher levels of experience, measured by the number of prior online textbook purchases, typically have price expectations that are closer to the empirical mean. Beliefs about the variation of the price distribution improve

[1] Earlier studies examined learning and search through experimental designs (e.g., Sonnemans, 1998; Einav, 2005)

[2] The research questions we address in this paper are further supported by research in the labor literature, which uses subjective beliefs about future earnings to explain college major choice (Arcidiacono et al., 2012; Stinebrickner and Stinebrickner, 2011; Wiswall and Zafar, 2015). These studies show that incorporating students' subjective beliefs leads to significantly different estimates than those obtained under the assumption of rational expectations. In addition to this, Wiswall and Zafar (2015) show that college students' beliefs about future earnings become more consistent with the actual earnings distribution as they proceed through school (i.e., become more familiar with their field).

modestly with experience. Overall, the evidence is consistent with learning, at least for learning about the mean of the price distribution. These results support a greater focus on the role of information and subjective beliefs when estimating models of consumer behavior.

The following section provides theoretical motivation for this project and expands on our goals. Section 3 describes the data and Section 4 presents results. Section 5 discusses the issue of selection, and Section 6 concludes.

2 Theoretical Motivation

In order to understand how consumers' beliefs about prices affect the search decision and to motivate the empirical section of this paper, consider the following simple model of consumer search. Individuals can purchase a given product from two locations. Assume for simplicity that the search cost is zero for one of the locations, so the individual knows the price of the product at this location. The price of the product at the other location is unknown by the individual, and there is a cost associated with determining this price.[3] Denote the price at the zero search cost location as p^* and the price at the location with a positive search cost as p, which is a random variable with cumulative density function, $F(p)$. The individual can either purchase the product from the first location or pay some cost, c, to search and discover the price at the other location. If the individual decides to search, he does not incur an additional search cost should he choose to purchase the product from the first location (i.e., search with recall).

The decision rule for the search problem is given by Equation (1). An individual chooses to search if,

$$\int_0^{p^*} \big(p^* - p\big) d\tilde{F}_i(p) > c_i \tag{1}$$

where $\tilde{F}_i$ denotes an individual's beliefs about the empirical price distribution. The LHS of Equation (1) is the expected benefit of search. A individual integrates over the difference between the known (p^*) and unknown price (p), given his beliefs about the distribution of the unknown price. The domain of integration is bounded above by p^* because an individual can costlessly revisit the first location (i.e., the benefit from search is weakly positive).

[3]This model abstracts from the decision to engage in sequential search or fixed-sample search as there is only one firm to search over. For both types of search behavior, beliefs about the mean that are biased upward or beliefs about the variance that are biased downward lead to suboptimal levels of search.

The RHS of Equation (1) is an individual specific search cost c_i. The majority of the structural consumer search literature attempts to recover the distribution of individuals' search costs. In order to do so, the econometrician must make assumptions regarding individuals' beliefs, $\tilde{F}$. A common assumption is that individuals know the parameters of the distribution of p (i.e., rational expectations).[4]

In this paper, we focus on the first two moments of individuals' beliefs. Determining if these moments match the corresponding moments of the empirical price distribution is important for the estimation of search costs. If consumers overestimate the mean of the empirical price distribution, then the model will generate an upwardly biased distribution of search costs under the rational expectations assumption. Similarly, if consumers underestimate the variance of the empirical price distribution, search cost estimates will also be biased upward.[5]

An alternative to rational expectations is to allow uncertainty and learning over the parameters of the price distribution. When individuals search and observe a price draw, they can use this information to update their beliefs according to a learning process (e.g., Bayesian). If individuals are forward looking, the value of the information gained from search is added to the expected contemporaneous benefit from search.[6] The acquisition of information provides an incentive for search, so individuals could choose to search even if the expected contemporaneous benefit was less than the search cost. Even in the learning framework, however, some variant of the rational expectations assumption is commonly used to restrict individuals' initial prior beliefs as the initial priors are typically not separately

[4]An alternative to making a parametric assumption on the empirical price distribution and consumers' beliefs is to instead assume that consumers form beliefs non-parametrically based on the empirical CDF of observed prices:

$$F(p) = \tilde{F}_i(p) = \frac{1}{N}\sum_{k=1}^{N} I[p_k < p]$$

where N is the number of observed prices. If consumers' beliefs are biased relative to the empirical distribution, this leads to similar biases in search costs that are discussed under the assumption of a parametric distribution for prices and beliefs.

[5]Misspecification of beliefs also leads to biases in price elasticity estimates. If individuals' beliefs are biased such that they underestimate the benefit of search (relative to the assumed, true benefit), then the model will recover price elasticities that are lower relative to the true elasticities.

[6]The addition of learning has implications for the tests of fixed sample versus sequential search as a model of sequential search with learning can generate similar empirical predictions as a model with fixed sample search and no learning (De los Santos et al., 2013).

identified. In the empirical section of the paper we test whether inexperienced individuals have biased beliefs about the parameters of the price distribution. We also examine whether individuals' beliefs are consistent with learning by testing whether more experienced individuals have beliefs that are closer to the parameters of the empirical price distribution.

3 Data

We collected data on subjective beliefs about the distribution of prices in the online textbook market through online questionnaires sent to students at the University of North Carolina at Chapel Hill (UNC).[7] The questionnaires asked individuals about their previous textbook purchasing behavior and presented them with hypothetical textbook purchasing scenarios. We supplement the responses to these textbook purchasing scenarios with price data scraped from an online marketplace for a large number of textbooks. Before providing a summary of both datasets, we will provide more detailed information about the textbook purchasing scenarios.

3.1 Textbook Purchasing Scenarios

Each questionnaire contained three randomly assigned hypothetical textbook purchasing scenarios from a total of twelve potential scenarios. Each of the twelve total scenarios was based on a different textbook.[8] Figure 1 is a screenshot of the information provided in one particular scenario.[9]

For each scenario, participants were presented with the title and author of the textbook and a description of the course for which the textbook is assigned. On the same page, they were asked whether they had ever taken the course and whether they had ever been assigned this particular textbook. On the following page, respondents were provided with

[7] Appendix Section A.1.1 contains the text from the online questionnaire. Individuals who agreed to participate in the survey were sent a link to the questionnaire.

[8] These twelve textbooks include four textbooks each from physical sciences, social sciences, and humanities. Of the four textbooks within these general fields, two are from introductory level courses. More information on the characteristics of the textbooks used in the hypothetical purchasing scenarios can be found in Appendix A.2.

[9] For each scenario, we randomly assigned respondents to a full information case (title, author, publisher, picture, etc.) or a limited information case. As opposed to the full information case, as seen in Figure 1, the limited information case only provided information on the title, author, and course.

the (actual) price of a new copy of the textbook from the campus bookstore, and were asked what they thought the price of the book would be if they searched only one online retailer.[10] We interpret the response to this question as the expected price in a statistical sense (i.e., the mean of the subjective price distribution). In online marketplaces, the same book can be listed at multiple different prices by different sellers. Since the textbooks are a homogeneous product, we assume that the response corresponds to beliefs about the lowest price. In the results section, we consider the possibility that individuals interpret these questions as asking about the median rather than the mean of their subjective beliefs about the price distribution.

In order to elicit information about individuals' beliefs about the higher order moments of the price distribution, we then asked respondents for the probability that the price realized after search would be less than X% or greater than Y% of their reported expected price for a new copy of the textbook. X was randomly drawn from {85, 90, 95} and Y was randomly drawn from {105, 110, 115}. For example, in Figure 1, the new price of the textbook at the campus bookstore for the Fall 2012 semester was $87.00. If the respondent reported that her expectation of the lowest price for a new copy of the textbook from one online retailer was $50.00, then the next questions would ask her the probability that the lowest price would be less than $45.00 and the probability that the lowest price would be greater than $55.00.

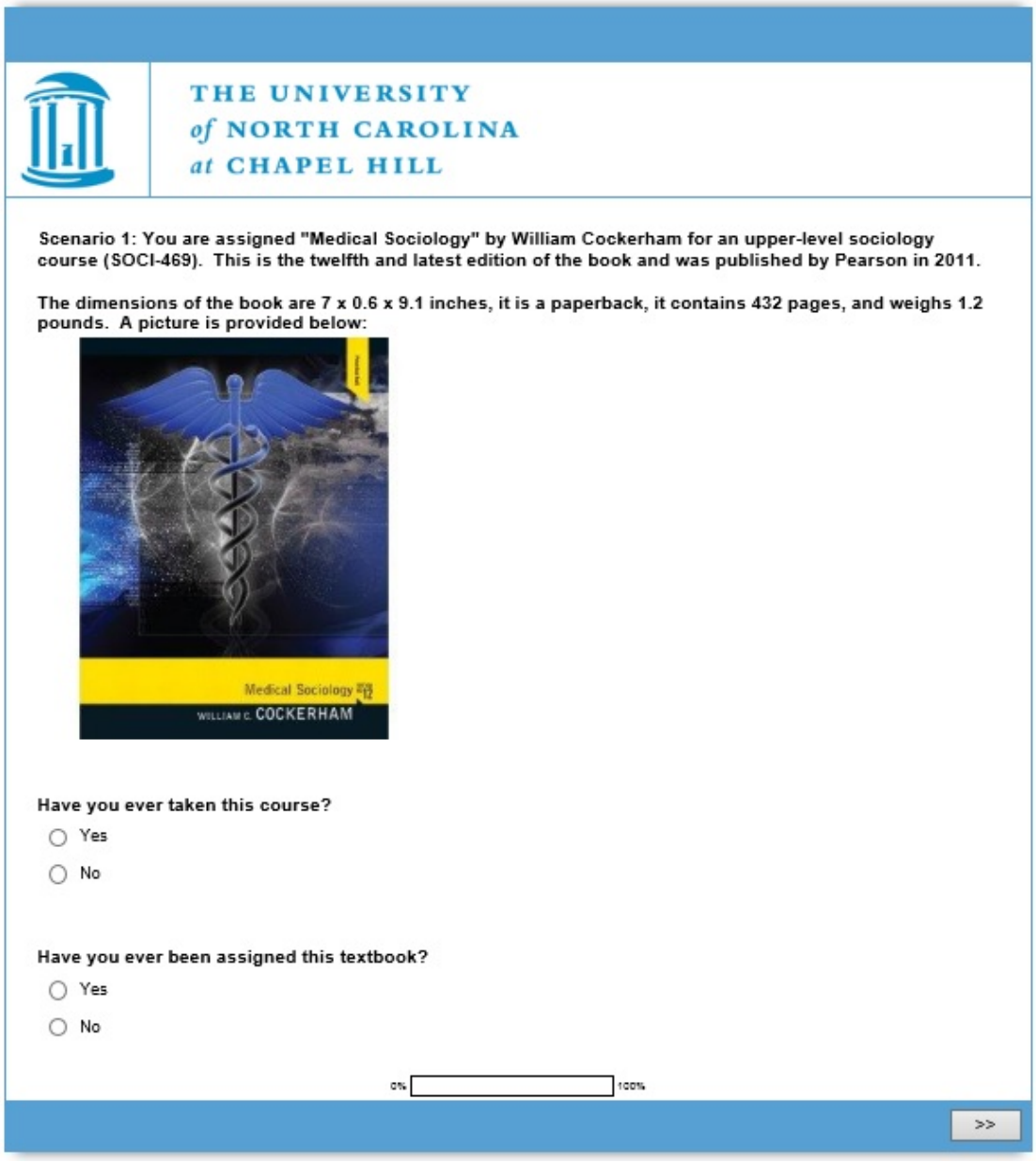
Given that individuals may not be accustomed to thinking about prices in a probabilistic manner, we first presented individuals with an example in order to help clarify the questions. In the example, we asked individuals to consider the *lowest* price they might find for a pair of jeans if they searched one retailer at the mall. This example contained information about probabilities (e.g., that their response should be between 0 and 100 percent) and clarification about the nature of price uncertainty (i.e., that although their best guess might be $20, there is some chance that the price is actually lower or higher than $20).

3.2 Online Questionnaire Data

We conducted two waves of the survey. The first was during the Fall semester of 2012, and the second was during the Spring semester of 2013. For the Fall 2012 and Spring 2013

[10]The bookstore price provided to students explicitly included sales tax. Respondents were asked to include shipping fees when providing their expectation of the lowest price available. Respondents were also reminded to not actually search for the lowest price of the textbook.

Figure 1: : Textbook Purchasing Scenario



semesters, 820 and 798 respondents completed the background questions about their previous textbook purchasing experience, respectively. The sample used in analysis is composed of 739 respondents from the Fall 2012 semester and 726 respondents from the Spring 2013 semester.[11] 104 respondents (52 from both semesters) were dropped because they had been enrolled in college for more than 10 semesters and an additional 49 respondents (29 from Fall 2012 and 20 from Spring 2013) were dropped for reporting nonsensical answers (e.g., reporting an expected price of $100,000).[12] Appendix A.3 provides a more detailed description of within survey attrition.

Table 1 displays the number of semesters enrolled for the questionnaire respondents. This count includes both traditional fall and spring semesters and any summer sessions the students had previously been enrolled in. Individuals in later semesters are slightly over sampled due to the nature of how we recruited individuals for the study. We obtained the email addresses of individuals who participated in a separate, longer running data collection project and agreed to receive follow-up emails. This other project began in the Fall of 2011 and recruited new individuals each semester. Individuals who participated at the start of this other project would be at least in their third semester at the time of data collection (assuming continual enrollment). Appendix A.1 provides more detail on how individuals were recruited.

Respondents' previous textbook purchasing behavior and major choice are also reported in Table 1. A majority of respondents have purchased textbooks at the campus bookstore and from an online retailer. There is significant variation in how many textbooks respondents have purchased online; 33.6% of the individuals in the sample have purchased five textbooks or fewer from online retailers.

[11] There were 240 individuals who participated in both surveys.

[12] In practice, this was done by removing respondents who reported expectations less than 10% or greater than 150% of the bookstore price. In Appendix B.1 we report out main results for a more relaxed omission criterion. The results are substantially the same.

Table 1: Respondent Characteristics

	Proportion
1 - 2 Semesters	0.143
3 - 4 Semesters	0.248
5 - 6 Semesters	0.242
7 or More Semesters	0.369
Ever Purchased at Campus Bookstore	0.960
Never Purchased Online	0.106
Purchased 1 - 5 Online	0.230
Purchased 6 - 10 Online	0.241
Purchased 11 or More Online	0.423
N = 1465	

3.3 Online Retailer Data

In order to construct an empirical distribution of prices for textbooks, we used a script in Perl to scrape .html files from Amazon.com. We collected daily price data for approximately 3,500 books that were assigned at UNC during the 2012-2013 school year. Using these .html files, we used a separate script in Perl to parse the lowest prices available for both new and used copies of the books on each day. For used books, the additional dimension of product quality makes mapping individuals' price beliefs to an empirical distribution more complicated. If individuals have a particular quality threshold in mind, the corresponding empirical distribution would be based on the lowest price for a book of a sufficiently high quality. We restrict the analysis to new books to make interpreting the results more straightforward.[13] In order to aggregate across textbooks, we normalize the lowest online price by dividing by the bookstore price.[14] We use the daily price data for two intervals corresponding to the timing of the surveys as well as for the start of the semester. The Fall survey period is from November 30, 2012 to December 10, 2012, and the Spring survey period is from April 11,

[13]The results for used books are consistent with the results presented in this paper.

[14]By forming the empirical distribution in this way, we are assuming that different textbooks have the same relative price distribution (online relative to bookstore). Otherwise a separate empirical distribution would need to be formed for each individual textbook.

2013 to April 26, 2013.[15] The start of semester time period is defined as the week prior to the start of classes as well as the first week of classes. To construct the empirical price distribution, we use the average price of the textbook over the survey period. The price sample used in the analysis trims the top and bottom 0.5% of the prices for each survey period.[16] Since outliers in the left tail of the distribution are bounded by zero, removing the outliers decreases the mean of the empirical distribution. Also, removing the outliers decreases the variance of the distribution. If the outliers that are removed do not actually correspond to erroneous listings, then the mean and variance of the empirical distribution will be lower than the mean and variance of the true distribution. The empirical distribution for the start of semester prices is constructed similarly.

The total cost of purchasing books online includes shipping fees, which vary depending on the speed of delivery. For items purchased on the Amazon Marketplace from third party sellers, we added the fee for standard shipping. Items purchased directly from Amazon qualify for free standard shipping as long as the item is purchased as part of an order that exceeds a certain amount.[17] Most new textbooks will qualify for free shipping if purchased directly from Amazon, so we do not add any shipping fees to the price of these books.[18] We include sales tax in the campus bookstore prices. Sales taxes are not included in the online prices.[19]

Table 2 provides the ratio of prices of textbooks from Amazon.com relative to the price from the campus bookstore. The first row reports the prices of new books for the full sample

[15]The online questionnaire was initially distributed on November 30, 2012 for the fall survey and April 11, 2013 for the spring survey. Nearly all of the surveys were completed during these intervals. We take these periods as the time frames that individuals are forming their expectations over. This is potentially problematic as online textbook prices vary systematically across the year (e.g., they are generally higher in August than May).

[16]The trimmed sample excludes books that have an online price listing that is either a very small fraction or a large multiple of the bookstore price. In some cases, particularly for books with low sales volumes, the automated pricing algorithms used by larger book resellers can generate these extreme prices.

[17]Orders that exceeded $25 qualified for free shipping at the time of the surveys.

[18]We do not include shipping for books that do not individually qualify for free shipping because they could be purchased as part of a larger order that does qualify for free shipping. The empirical analysis focuses on higher priced books that would qualify for free shipping. All of the books in the hypothetical textbook purchasing scenarios qualify for free shipping if purchased new from Amazon.

[19]At the time of this analysis, Amazon did not collect sales taxes. Individuals were responsible for paying the sales taxes for online purchases, however compliance was low. Sellers on the Amazon Marketplace are responsible for paying any applicable sales taxes, so sales taxes are implicitly included in the listed prices.

Table 2: Ratio of Amazon Prices to Bookstore Prices

		Mean Ratio	S.D.	Min	Median	Max	Mean Diff. ($)	N
Prices During Survey Periods								
All Books	New	0.846	0.143	0.380	0.845	1.398	10.27	2139
	New_{mkt}	0.768	0.279	0.187	0.744	3.970	16.71	2312
Bookstore	New	0.811	0.138	0.388	0.814	1.327	31.51	417
Price > $100	New_{mkt}	0.662	0.179	0.205	0.652	1.508	54.33	440
Scenario Books	New	0.760	0.087	0.578	0.787	0.893	38.77	11
	New_{mkt}	0.635	0.142	0.449	0.601	0.857	60.63	11
Start of Semester Prices								
All Books	New	0.849	0.150	0.477	0.836	1.5	9.53	2162
	New_{mkt}	0.805	0.244	0.253	0.779	3.272	13.03	2319
Bookstore	New	0.829	0.107	0.477	0.819	1.213	27.83	418
Price > $100	New_{mkt}	0.750	0.169	0.301	0.753	1.471	40.16	441
Scenario Books	New	0.775	0.107	0.550	0.787	0.941	34.97	11
	New_{mkt}	0.713	0.113	0.533	0.740	0.874	47.55	11

Notes: The ratio reported is the lowest price on Amazon.com divided by the price of the same title from the campus bookstore. New_{mkt} refers to the minimum of the lowest price listed by marketplace sellers and Amazon.com for a new copy of the title. "Scenario Books" are the titles used in the hypothetical textbook purchasing scenarios in the online questionnaire.

of books for which we have data. On average, new prices on Amazon.com are approximately 85% of the bookstore price. The second to last column reports the average difference between the price of the textbook from the bookstore and an online retailer. For all textbooks in our sample, the savings in absolute terms is approximately $10. The second row reports the new prices that includes new books listed on the Amazon marketplace by third party sellers. The new price using this specification is defined as the minimum of the Amazon and Amazon marketplace listings for a new copy of the book and is denoted New_{mkt}. Including the marketplace listings increases the average savings relative to the bookstore price and increases the variation in the distribution. The sample size also increases since some titles are not available from Amazon. The price specification which includes the marketplace price is our preferred measure of the online price distribution for the subsequent analysis. Future references to the marketplace are implicitly assumed to include Amazon itself if the price

from Amazon is lower than the lowest marketplace price. The median price is lower than the mean, as the distributions are slightly skewed to the right.

The next two rows of Table 2 provide summary statistics for textbooks which are priced greater than $100 for a new copy from the campus bookstore. Books with a list price below $100 include popular press titles that have a large market outside of being assigned for a college course. Relative to the full sample, the potential savings from shopping online becomes greater (i.e. in both percentage and magnitude terms, more expensive textbooks have greater savings in the online market). The variability of prices is less relative to the full sample. The final two rows provide summary statistics for the textbooks used in the hypothetical textbook purchasing scenarios.[20] On average, these prices are slightly lower than the sample of textbooks with a price of $100 or more at the campus bookstore, but the difference in average price is not significant at the 0.05 level.[21] The bottom half of Table 2 reports the prices at the start of the semester. Prices at the start of the semester are higher than during the survey periods, which occur towards the end of each semester. The change in price is greater when the marketplace is included. The variability of prices also tends to be lower at the start of the semester.

Ideally, how we define the empirical price distribution should match the price distribution of the individuals' beliefs, but there are a few reasons why this may not be the case. First, textbook prices vary over time, and the time frame used to define the empirical distribution may not match the time frame of the individuals' subjective beliefs. To address this issue, we conduct the analysis using both the survey period and start of semester time frame to define the empirical distribution. Although the survey period prices reflect the empirical distribution at the time the individuals were asked to consider the hypothetical textbook purchasing scenario, most relevant experience in the market occurs at the start of the semester when individuals actually make the search and purchasing decision. Second, we only use price data from a single online retailer/marketplace. We believe the prices from Amazon.com provide a reasonable approximation to the empirical distribution of prices that consumers face if they only search one online retailer. Of the individuals in our sample, 75% reported Amazon.com as the first website they would visit to search for a textbook. These issues

[20] Note that the total number of textbooks in the purchasing scenarios is actually 12. However, online retailer data for one textbook is missing.

[21] Only the difference for new books purchased directly from Amazon for the start of semester is significant at the 0.10 level.

of timing and alternative retailers affect the comparison between the individuals' subjective beliefs and the empirical price distribution. The comparison of individuals' subjective beliefs across different levels of experience is not affected as long as individuals with different levels of experience do not systematically differ in the time frame considered or in the choice of the online retailer.

4 Results

The first subsection presents results using the data on reported expectations. The following subsection incorporates additional data on beliefs to examine not only individuals' expectations but also individuals' beliefs about the variance of the empirical price distribution in the context of a parametric learning model.

4.1 Expectations Results

In this section we present descriptive statistics of individuals' price expectations. Then, we test for differences in price expectations relative to the empirical prices across levels of experience in order to determine if consumers' expectations converge to the mean of the empirical price distribution. The results in this section do not require the specification of a particular distribution for the individuals' price beliefs.

The first columns of Table 3 provide the summary statistics of the reported expectations of the lowest online price as a proportion of the bookstore price for individuals with different levels of online textbook purchasing experience. In the survey, individuals were asked about the number of textbooks they had ever purchased online, and they responded by selecting one of four possible categories. Individuals with no prior online textbook purchases expect the price of a new book online to be approximately 83% of the price of a new book at the college bookstore. This corresponds to an expected savings of $27.23 on average across the hypothetical textbook purchasing scenarios. Individuals with prior online textbook purchases expect the online price to be lower, with higher levels of experience corresponding with a greater expected savings. On average, individuals with more than ten previous online purchases expect the price of a new book online to be approximately 77% of the price of a new book at the college bookstore. This corresponds to an average savings of $37.91.

Table 3: Mean Ratio Comparisons by Online Purchasing Experience, New Books

	Expectation ÷ Bookstore Price				Mean Expectation Bias			
					Scenario Books		Books > $100	
Experience	N	Mean Ratio	S.D.	Median	Survey Period	Start of Semester	Survey Period	Start of Semester
No online purchases	440	0.833	0.172	0.864	0.198***	0.120***	0.171***	0.083***
1-5 online purchases	923	0.819	0.153	0.856	0.183***	0.105**	0.156***	0.068***
6-10 online purchases	966	0.783	0.161	0.822	0.147***	0.069*	0.121***	0.032***
11+ online purchases	1706	0.770	0.157	0.795	0.135**	0.057	0.108***	0.020**

Notes: The ratio reported is an individual's expectation of the lowest price from an online retailer divided by the price of the same title from the campus bookstore. Expectation Bias refers to the difference between this ratio and the ratio of the observed Amazon Marketplace price to the bookstore price. Each observation corresponds to one hypothetical textbook purchasing scenario.

* refers to t–test p–value < .1; ** < .05; *** < .01; H_0 = No difference between ratios.

The relationship between the price expectation and experience would be consistent with learning if the individuals with higher levels of experience report expectations that are closer to the true mean of the price distributions. The mean expectation bias is defined as the difference between the expected price ratio and the mean of the empirical distribution, where the empirical distribution includes the Amazon Marketplace (New_{mkt} in Table 2). The final four columns of Table 3 report the mean expectation bias for scenario textbooks and textbooks with a list price greater than $100 both during the survey period and at the start of the semester. The mean expectation bias for individuals with no prior online textbook purchases is positive and statistically significant for all specifications of the empirical distribution. Individuals with no prior experience on average underestimate the savings from purchasing online. This result contradicts the standard rational expectations assumption that individuals know the online price distribution. The mean expectation bias decreases for higher levels of experience. For the survey period price distribution, the mean expectation bias remains significant at the highest experience level. For the start of semester price distribution for books with a list price greater than $100, the mean expectation bias is significant at the highest level of experience but only at the 5% level. For scenario books, the mean expectation bias is not statistically significant at the highest level of experience. For books with a list price greater than $100, the median online price is almost identical to the mean

online price (see Table 2). Therefore, these results hold if individuals interpret the question as asking about the median price. Whether individuals respond with the mean or the median price becomes a larger issue when fitting a particular distribution to individuals' beliefs.

4.2 Distribution Results

In this subsection, we examine whether the patterns observed in the data are consistent with learning over additional parameters/moments of the empirical price distribution. Unlike the results in the previous section, the results in this section require the specification of a particular distribution for the beliefs. In the hypothetical textbook purchasing scenarios, individuals report their expectations for prices as well as the probability that a draw from the price distribution is below a given threshold ($\mathbf{E}[p]$ and $F_p(p_L;\mu,\sigma)$). We use these two moments to calculate the expected parameters of each individual's beliefs (i.e., $\mathbf{E}[\mu]$ and $\mathbf{E}[\sigma]$), under the assumption that individuals believe that prices follow a log-normal distribution.[22] The log normal distribution has two properties that make it an appropriate distribution in the current context. First, the support of the distribution is non-negative real numbers and prices are bounded below by zero. The second feature is that the log normal distribution is skewed to the right, which is a feature of both the reported beliefs in the sample and the empirical distribution. The most important criteria is that the beliefs (i.e., prior and posterior distributions) of the distribution parameters are conjugate distributions, which is necessary for tractably modeling a Bayesian learning process. The results are similar under alternative distributional assumptions.[23]

Assuming that individuals believe that the distribution of prices is log-normal, then individuals' prior distribution on μ and $\frac{1}{\sigma^2}$ is Normal-Gamma. If the individual searches, she observes a price which she uses to update her beliefs. As the number of price observations increases, the individual's mean prior on μ and σ converge to the true parameters, and the variance of the priors converge to zero. In terms of the search problem, evidence of individual learning requires that individuals with more experience in the market (i.e., more observations

[22]The parameters of the log-normal distribution are calculated using the following equations for the mean and CDF of a log-normal random variable: $\mathbf{E}[p] = exp(\mu + \frac{\sigma^2}{2})$, and $F_p(p_L;\mu,\sigma) = \Phi(\frac{log(p_L)-\mu}{\sigma})$, where Φ is the standard normal CDF.

[23]The results for the gamma and normal distributions are presented in the Appendix. These are the most commonly used continuous distributions in Bayesian learning problems and can easily handle problems where individuals learn over multiple parameters of the distribution.

Table 4: Reported Probability that Lowest Price $< b *$ Expected Lowest Price

b	N	Mean	S.D.	Min	Median	Max
0.85	1359	0.271	0.176	0	0.250	1.00
0.90	1376	0.312	0.190	0	0.300	1.00
0.95	1293	0.339	0.195	0	0.300	1.00

of prices) have more accurate beliefs about the true parameters of the price distribution and more certainty in their beliefs.

Denote the individual's expected parameters as μ_i and σ_i. In the analysis, we consider the distribution of the individual's expected parameters in the population. Define $\bar{\mu}^e$ and $\bar{\sigma}^e$ as the mean of individuals' beliefs with the same level of experience e (i.e. $\bar{\mu}^e = \frac{1}{N^e}\sum_i \mu_i * \mathbf{1}[e_i = e]$ and $\bar{\sigma}^e = \frac{1}{N^e}\sum_i \sigma_i * \mathbf{1}[e_i = e]$). Similarly, define $Var[\mu]^e$ and $Var[\sigma]^e$ as the variance among individuals' beliefs with experience level e. As the number of signals the individual receives increases, the expected parameters should converge to the true price distribution parameters. Since each individual's beliefs converge to the true parameters, $\bar{\mu}^e$ and $\bar{\sigma}^e$ should converge to the true parameters as e increases. The convergence of each individual's beliefs to the true parameters as experience increases implies that the variance among individuals' beliefs goes to zero. However, at low levels of experience, $Var[\mu]^e$ and $Var[\sigma]^e$ may increase depending on the variance among individuals' initial prior beliefs. If individuals have similar initial mean priors, then the signal noise could generate greater dispersion of individuals' beliefs for low levels of experience.

Table 4 reports the summary statistics for the reported probability that a draw from the price distribution is below some threshold for different levels of the threshold. The threshold is defined as a fraction of the individual's reported expectation. On average, individuals report that the likelihood of the lowest price being less than 85% of their expected lowest price is 0.271. For higher levels of the threshold, individuals assign a larger probability that the price is below the threshold.

Some individuals report a probability of zero or 100 which cannot be justified given the distributional assumption.[24] Similarly, reported probabilities close to zero or 100 will only fit the distribution for extreme values of the parameters. Once the parameter values are

[24]There are 64 scenarios where a probability of zero is reported and 11 where a probability of 100 is reported (out of 4028 total scenarios).

calculated, individuals with parameter values in the top or bottom 2.5% of parameter values for either parameter are dropped from the sample to reduce the impact of outliers.[25]

In order to make the interpretation of the results more straightforward, we use the individual's distribution parameters to calculate the mean and standard deviation of the individual's expected price distribution, which is defined as the distribution with the individual's expected parameter values.[26] Table 5 reports the sample mean and standard deviation of these moments of the individual's expected price distribution by level of experience. Differences in the mean values from the analysis in the previous section is due to the different samples that result from the different rejection criteria.[27] The mean and standard deviation of the preferred specifications of the empirical distribution are presented for comparison as well as the associated 95% confidence intervals.

The results for mean beliefs about the mean of the price distribution across different levels of experience is qualitatively the same as the analysis in the previous section; on average, individuals with more experience are more likely to believe that the mean of the price distribution is lower. The variation in beliefs about the mean of the price distribution within groups of individuals with similar experience changes little across different levels of experience. Individuals with more experience may have more accurate beliefs about the mean of the price distribution on average, but the results do not support the conclusion that the beliefs of all individuals are converging to a single point.

The mean of the beliefs about the standard deviation of the expected price distribution initially increases with experience (from 0.233 for individuals with no online purchases to 0.249 for individuals with 1 to 5 online purchases) and then decreases with experience for higher levels of experience. Comparing the beliefs about the standard deviation of the price distribution to the empirical standard deviation suggests that individuals may overestimate the variability of prices for new books (the lowest belief for any experience level is 0.211, which lies outside of any of the reported 95% confidence intervals for the empirical distribution except for the scenario books during the survey period). The variability of the standard

[25]Probabilities of zero and 100 are replaced with .1 and 99.9 respectively. The individuals who report probabilities of zero or 100 are included in the 2.5%. All individuals who report a probability of 100 end up being dropped. Most of the individuals who report a probability of zero are dropped.

[26]Note that the individuals' two responses for each scenario exactly identify their expectations of the mean and variance of the normalized price distribution.

[27]Since individuals are dropped if their beliefs are an outlier for either of the two parameters, the resulting sample is smaller than the sample used when only considering one parameter.

Table 5: Mean and Variance Comparisons (Log-Normal Assumption)

Empirical Distribution, Amazon and Amazon Marketplace

	N	Mean	Mean 95% CI	S.D.	S.D. 95% CI
Survey Period					
Bkstr. Price > \$100	440	0.662	0.645 - 0.679	0.179	0.164 - 0.196
Scenario Books	11	0.635	0.540 - 0.731	0.142	0.111 - 0.221
Start of Semester					
Bkstr. Price > \$100	441	0.750	0.735 - 0.766	0.169	0.154 - 0.186
Scenario Books	11	0.713	0.638 - 0.789	0.113	0.081 - 0.192

Beliefs

Interpreting Response as Mean

Experience	N	Mean $\mathbf{E}_i(p)$	S.D. $\mathbf{E}_i(p)$	Mean $\sqrt{\mathbf{Var}_i(p)}$	S.D. $\sqrt{\mathbf{Var}_i(p)}$
No online purchases	380	0.811***	0.153	0.233	0.319
1-5 online purchases	842	0.809***	0.146**	0.249**	0.376**
6-10 online purchases	899	0.783	0.151	0.220	0.315
11+ online purchases	1603	0.769***	0.151	0.211**	0.297*

Interpreting Response as Median

Experience	N	Mean $\mathbf{E}_i(p)$	S.D. $\mathbf{E}_i(p)$	Mean $\sqrt{\mathbf{Var}_i(p)}$	S.D. $\sqrt{\mathbf{Var}_i(p)}$
No online purchases	298	0.857***	0.142	0.164	0.141
1-5 online purchases	657	0.844***	0.137*	0.166*	0.140
6-10 online purchases	691	0.814	0.134	0.148**	0.120**
11+ online purchases	1258	0.804***	0.142**	0.156	0.139

Notes: The significance levels reported for the mean values are from a two-sample equality of means test, where the comparison sample is all other levels of experience.

The significance levels for the standard deviations are from Brown and Forsythe's alternative formulation of Levene's robust two-sample equality of variances test.

The confidence interval for the standard deviation of the empirical distribution uses Bonett's approximation for non-normal sampling distributions.

* refers to p–value < .1; ** < .05; *** < .01

deviation of the expected price distribution across individuals with the same level of experience increases initially, but decreases for higher levels of experience, which is consistent with learning. The significance levels reported for the average of the beliefs about the mean and the variance of the price distribution are from a two-sample equality of means test that compares individuals within a particular experience group to everyone not in that group. The test for the equality of variances defines the comparison samples in the same way. These tests are for evaluating whether beliefs vary significantly across different levels of experience. The results from the statistical tests suggest that individuals with low levels of experience and the most experience tend to have beliefs that are significantly different from both each other and the other groups (no experience and moderate levels).

The final set of results in Table 5 calculate the distribution parameters assuming that individuals respond with the median of their expected price beliefs instead of the mean. Individuals who report a probability of 50% or higher that the price is below the threshold are dropped from the analysis.[28] The results for the beliefs about the mean are similar, but the values are larger (since the mean of the log-normal must be greater than the median). The beliefs about the variability of the distribution do not exhibit much of a pattern. These results may be difficult to interpret since individuals who expect a large variance in the price distribution report probabilities that are inconsistent with interpreting the initial response as the median. The results assuming a normal distribution are similar to the results using the log-normal and median interpretation and are presented in the appendix. Since the mean, median, and mode are the same for a normal distribution the issue of how to interpret the response to the survey question is irrelevant if prices are assumed to be normally distributed. However, a large number of individuals are dropped for reporting probabilities greater than 50%, which are not consistent with the distribution.

Overall, the evidence is consistent with learning, although the evidence suggests incomplete learning. It may be the case that individuals are only learning over a single parameter. This would explain why individuals with more experience are better able to predict the mean price, but are no better in incorporating the variance of the price distribution into their beliefs. Another possibility is that individuals in the sample do not have sufficient experience for the convergence properties of the learning process to be evident.

[28]The threshold is defined as a proportion of the initial response, so the cdf of the price distribution at the threshold must be below 0.5.

5 Learning vs. Selection

Although the evidence is consistent with learning, the differences in individuals' beliefs across levels of experience could also result from selection. If individuals have heterogeneous initial prior beliefs, then individuals who believe that the online price is similar to the bookstore price will not search and will not purchase their books online. Then, if the individuals whose initial priors are close to the true distribution are the ones who search and purchase online, the observed difference in beliefs would be the result of selection based on the initial difference in beliefs and not because of learning.

To distinguish between the effects of learning and selection, we examine the individuals who participated in the survey in both the fall and spring semesters. There were 240 individuals who participated in both surveys. Of these individuals, 89 reported an increase in their level of online textbook purchasing experience from the fall to the spring survey. If selection is generating the observed patterns in the data, then the individuals who report an increase in experience in the spring would have lower expected online prices in the fall than the individuals who do not have an increase in experience[29]. In addition to this, individuals who report an increase in experience should be more likely to report different beliefs in the spring in order for the data to be consistent with learning; individuals who do not report an increase in experience should have similar beliefs in both periods. The inherent limitation of this test is that one period of learning may not generate a significant difference in beliefs for those whose experience increased. Therefore, this test is primarily a test of the hypothesis of no selection. Another limitation is that individuals who participate in the survey multiple times may not be representative of the overall population.

Table 6 reports the mean parameter values for beliefs fitted to a log-normal distribution for the two groups for both surveys as well as the mean change in parameter values between surveys. There is not a significant difference between the mean parameter values of the two groups in the fall semester, which is consistent with the hypothesis of no selection. In the spring semester, the magnitude of the difference between the parameters of the two groups increases. The difference in the spring semester across groups is significant for the parameter σ but not for the parameter μ. The group that reports an increase in experience has a larger

[29]Individuals who do not report an increase in experience may actually be gaining experience. For example, they may be moving from 6 to 8 online purchases, or they may have engaged in an online search and not made a purchase.

Table 6: Beliefs by Change in Experience

	Fall			Spring			Difference	
Group	N	mean μ_i	mean σ_i	N	mean μ_i	mean σ_i	mean μ_i	mean σ_i
Increase Exp.	237	-0.303 (0.261)	0.266 (0.262)	240	-0.279 (0.236)	0.206 (0.207)	-0.024	0.060***
Same Exp.	395	-0.313 (0.255)	0.238 (0.236)	389	-0.304 (0.248)	0.248 (0.243)	-0.010	-0.010
Difference		-0.010	-0.025		-0.028	0.042**		

change in beliefs from the fall to the spring semester. For individuals whose experience increases, there is not a significant difference in the parameter μ across semesters, but the difference in the parameter σ is significant. There is no significant change in parameter values for the group whose experience remains the same. These results suggest that selection is not the primary cause of the differences in beliefs across experience levels and generally support the learning hypothesis. However, due to the limited sample size, no definitive conclusion can be drawn.

To further examine whether learning or selection are responsible for the observed relationship between beliefs and experience, we next focus on individuals for whom the effects of learning and selection would move in opposite directions. Individuals who report price expectations below the mean of the empirical distribution are more likely to search, but if the individuals are learning about the price distribution, then their beliefs about prices should increase on average after search. We limit the sample to individuals who have relatively low expectations for the price in the fall semester. These "optimistic" individuals are defined as those who are in the bottom 25% of the expected price responses in the fall semester after averaging over the scenarios for each individual. Table 7 reports the results for the price expectations for these individuals in the fall and spring semesters. Individuals who report a higher level of experience in the spring semester have a lower expected price in the fall semester, which is consistent with selection. However, this difference is not significant. Beliefs in the spring semester are significantly higher (closer to the empirical mean). Since the experience is measured in intervals, individuals who report the same level of experience may have more experience in the spring, just not enough to cross the upper threshold of their current experience interval. For this reason, it is not surprising that individuals who report the same level of experience in the spring also report a higher expected price. The magnitude of the increase is larger for those who report an increase in experience, but the difference is not significant.

Table 7: Change in Price Expectations for Individuals with Low Price Expectations

	Fall		Spring		Difference
Group	N	Expectation ÷ Bookstore Price	N	Expectation ÷ Bookstore Price	
Increase Exp.	46	0.603 (0.173)	46	0.699 (0.179)	0.095**
Same Exp.	90	0.645 (0.141)	85	0.735 (0.160)	0.090***
Difference		-0.041		-0.036	

Note: Sample includes individuals whose average expected price ratio was in the bottom 25% of all respondents for the fall survey.

Although the evidence presented does not find significant effects of selection and provides some support of the learning hypothesis, it is likely that both effects are responsible for the observed relationship between price beliefs and experience. The relative extent to which each is responsible is an open question. If there was learning and no selection, then individuals would have homogeneous initial beliefs that are biased upward. The individuals who search receive signals from the true price distribution and their beliefs become more accurate. If there was selection and no learning, then there would be heterogeneous initial beliefs. The individuals with lower price expectations would search while those with higher price expectations would not search. The typical assumption in consumer search models is that individuals have homogeneous initial beliefs that are unbiased (since the individuals are assumed to know the true price distribution), which does not allow for selection or learning.

6 Conclusion

The evidence is consistent with learning, but it appears that the learning process is incomplete. The beliefs of individuals with even the higher levels of experience on average do not fully converge to the empirical distribution. Also, many individuals with high levels of experience have inaccurate beliefs (i.e., the variation across individuals' beliefs does not converge to zero). There are three primary explanations for the persistence of inaccurate beliefs. The first is that the level of experience where this convergence would occur is beyond what we measure in the data. The second is that the beliefs are converging to a distribution that is more specific than what we specify. For example, books in different fields may have different online price distributions. Finally, there is likely to be some noise in the reported data as individuals may have differed in their interpretation of questions as well as the amount of consideration given to their responses.

One limitation of this analysis is the problem of external validity. Although the online market for textbooks is comparable to online markets for other goods, the individuals in our sample are not representative of consumers in other online markets. Relative to consumers in other online markets, individuals in our sample are likely to be younger with higher intellectual ability, but they may have less overall experience in online markets. If there are knowledge spillovers across online markets (i.e. if experience in one online market causes individuals to have more accurate beliefs about the prices in other online markets) then the observed bias in the initial beliefs may be more pronounced in the online textbook market, where individuals are likely to have less overall experience in online markets.

In this paper we use a novel dataset to examine subjective price beliefs and their relationship with experience in a market. We find that inexperienced consumers have biased beliefs, but that consumers appear to be learning about the empirical price distribution as they repeatedly participate in the market. This study also leaves open a wide avenue for future research. First, since we do not estimate a dynamic model of search and learning, we are not able to show how individuals incorporate their beliefs into the search decision. Thus, we are not able to determine whether individuals incorporate the benefits of the additional information obtained through search for future purchasing decisions in their decision to search. Also, if individuals have heterogeneous initial prior beliefs, one potential avenue of future research would be to determine the sources of this heterogeneity. Finally, since the price distributions are determined by the market equilibrium, future research could study the dynamic pricing decisions of firms in a market with heterogeneous search costs, heterogeneous initial prior beliefs, and learning about prices.

A Data Appendix

A.1 Online Questionnaire Data

Individuals who participated in the online questionnaire were respondents from a list of emails generated through participation in a separate online questionnaire conducted at UNC during the 2011-2012 and 2012-2013 academic calendar years. These individuals agreed to participate in follow-up surveys at the completion of the separate questionnaire. This separate questionnaire was distributed by instructors to their students, who had the option to participate. Additional details about this separate questionnaire are available in Spence (2015).

A.1.1 Online Questionnaire

[The following is a subset of the questions provided to textbook consumers using Qualtrics online survey software. Notes are provided in brackets.]

Textbook Purchasing Questionnaire

The following survey seeks to gain understanding into how consumers choose which retailers to consider when faced with purchasing decisions. Over the course of this survey you will be presented with a number of hypothetical textbook purchasing decisions. You will be asked about your price expectations from online retailers and your beliefs about the time costs involved with searching within an online market. You will not actually have to price any textbooks from online retailers or visit any website outside of this survey.

Directions: Please answer all questions to the best of your ability. Use the right arrow button at the bottom of the screen to advance to the next page. You may also use the left arrow at the bottom of the screen to move back at any time and change a previous answer. If you are uncomfortable answering a specific question you can either skip that question or exit the survey. Thank you for participating!

How many semesters in total, including this one, have you attended UNC and any other college? (Count a summer session as a semester)

Semesters: _____

What is (are) your major(s)? Please write Undecided if you do not currently have a stated major.

Major(s): _____

Which of the following have you ever purchased a textbook from? (Please check all that apply)

☐ UNC Student Stores (campus bookstore)

☐ Ram Book and Supply

☐ Another college bookstore

☐ Amazon.com

☐ Half.com or Ebay.com

☐ Ecampus.com

☐ Chegg.com

☐ Another online retailer

☐ Another student (directly)

Which of the following have you ever rented a textbook from? (Please check all that apply)

☐ I have never rented a textbook

☐ UNC Student Stores (campus bookstore)

☐ Ram Book and Supply

☐ Another college bookstore

☐ Amazon.com

☐ Half.com or Ebay.com

☐ Ecampus.com

☐ Chegg.com

☐ Another online retailer

☐ Another student (directly)

Please write in the other online retailers you have ever rented or purchased a textbook from:

When do you normally purchase (or order) your textbooks?

☐ More than 2 weeks before the semester starts

☐ 1 - 2 weeks before the semester starts

☐ A few days before the semester starts

☐ The day the semester starts

☐ A few days after the semester starts

☐ 1 - 2 weeks after the semester starts

☐ More than 2 weeks after the semester starts

Have you purchased or rented any textbooks for an upcoming summer session?

☐ Yes

☐ No

How many textbooks have you ever purchased or rented online?

☐ 1 - 5

☐ 6 - 10

☐ 11 or more

When purchasing or renting a textbook online, have you ever used a website that shows the lowest prices available from multiple online retailers?

☐ Yes

☐ No

On average, when you purchase a textbook online, how many different online retailers do you visit?
Number of Retailers Visited: _____

Do you have an Amazon Prime membership?

☐ Yes

☐ No

Do you have a Paypal account?

☐ Yes

☐ No

How many online purchases do you typically make in a three month period?

Number of Purchases: _____

If you were given an isbn number or textbook title and wanted to purchase or rent this textbook online, what is the first website you would visit?

Website Name: _____

Hypothetically, if you only visited one online retailer, how many minutes do you think it would take to look up **one** textbook and purchase it? (Include the time to search, find the option you want, enter your information, and complete the transaction)

Minutes: _____

Hypothetically, if you only visited one online retailer, how many minutes do you think it would take to look up **three** textbooks and purchase them? (Include the time to search, find the option you want, enter your information, and complete the transaction)

Minutes: _____

You will now be given a number of hypothetical textbook purchasing decisions. In each case, you will be given information about the textbook and asked to give your best guess about what the price of this textbook is from an online retailer. This survey is concerned about what your expectations

are about prices from online retailers, so please do not actually search for the price of the textbook online. Before presenting you with the hypothetical purchasing decisions, you will be provided with an example of what the questions will be like.

Example: If you searched only one clothing store in the mall (ex. Old Navy), what do you think is the lowest price you could find for a pair of jeans in your size? Please enter your answer as a number. Note that this question does not have a right or wrong answer, it simply asks for your best guess.

$\_\_\_\_\_$ [Denoted "Example Expectation" in following questions]

Example Continued: Given that you dont know the lowest price of a pair of jeans with certainty, there is some chance that the lowest price is lower than $[Example Expectation] and some chance that the lowest price is greater than $[Example Expectation]. In the following questions, you will be asked about your beliefs about the chance that the lowest price you could find would be below $[Example Expectation] and also the chance the lowest price you could find would be above $[Example Expectation].

What do you think is the chance that the lowest price of the pair of jeans is less than $[90% of Example Expectation]? Please enter the chance as a percentage (i.e. a number between 0 and 100). For example: I think there is a 30% chance that the lowest price of the pair of jeans is less than $[90% of Example Expectation];

Percent Chance: _____

What do you think is the chance that the lowest price of the pair of jeans is more than $[110% of Example Expectation]? Please enter the chance as a percentage (i.e. a number between 0 and 100). For example: "I think there is a 35% chance that the lowest price of the pair of jeans is greater than $[Example Expectation]."

Percent Chance: _____

You will now be given three textbook purchasing scenarios, each similar to the previous example.

[The following is an example using one of the possible twelve textbooks. Respondents were given three scenarios randomly drawn from three groups of four textbooks (one from each group).]

Scenario: You are assigned "Economics: Principles and Policy" by William Baumol and Alan Blinder for an introductory economics course (ECON-101). [The following information on textbook characteristics was randomly assigned to respondents with 50% likelihood. The likelihood of receiving this information for the subsequent scenarios did not depend on whether the information on characteristics was shown for previous textbook purchasing scenarios.] This is the twelfth and latest edition of the textbook, it was published by South-Western College Publishing, and was last

revised in 2012. The dimensions of the book are 8.4 x 1.5 x 11.1 inches, it is a hardcover, it contains 880 pages, and it weighs 4.4 pounds. A picture is provided below:

[Picture presented such as the one shown in the screenshot in Section 3]

Have you ever taken this course?

☐ Yes

☐ No

Have you ever been assigned this textbook?

☐ Yes

☐ No

You know that a new copy of this book costs $212 (including taxes) at the UNC Student Stores. If you searched one online retailer, what do you think the price of a new copy at this online retailer would be (include shipping costs)? Reminder: Please do not actually search for this price. Provide your best guess instead.

$ _____ [Denoted "New Expectation" in future questions]

What do you think is the probability that the lowest price for a new copy of this book is less than $[85%, 90%, or 95% of New Expectation] (including shipping costs) at the online retailer?

Percent Chance: _____

What do you think is the probability that the lowest price for a new copy of this book costs more than $[105%, 110%, or 115% of New Expectation] (including shipping costs) at the online retailer? Note that your answer to this question added to your answer from the previous question should not exceed 100.

Percent Chance: _____

You know that a used copy of this book costs $159 (including taxes) at the UNC Student Stores. If you searched one online retailer, what do you think the price of a used copy at this online retailer would be (include shipping costs)?

$ _____ [Denoted "Used Expectation" in future questions]

What do you think is the probability that the lowest price for a used copy of this book costs less than $[85%, 90%, or 95% of Used Expectation] (including shipping costs) at the online retailer?

Percent Chance: _____

What do you think is the probability that the lowest price for a used copy of this book costs more than $[105%, 110%, or 115% of Used Expectation] (including shipping costs) at the online retailer? Note that your answer to this question added to your answer from the previous question should

not exceed 100.

Percent Chance: _____

A.2 Textbook Purchasing Scenarios

Table 8 provides information on the textbooks used in the hypothetical textbook purchasing scenarios. Respondents that completed the survey faced three scenarios; in each scenario, one textbook from each group was randomly assigned to the respondent. The first group is composed of social science textbooks; the second group is composed of hard science textbooks; the third group is composed of humanities textbooks. In the Fall 2012 semester, individuals were presented these scenarios in the previous ordering (social sciences, hard sciences, then humanities). The Spring 2013 questionnaire assigned the order of the groups randomly.

Textbooks were chosen to provide variation in the following characteristics: the number of total editions of the textbook, whether the textbook is the latest edition, the year of publication, whether the course was designed for an introductory or upper-level course, the type of cover (hardback vs. paperback), the number of pages, and the weight. In the following tables, course number refers to the numbering at UNC for the Fall 2012 and Spring 2013 semesters. New bookstore price refers to the price from UNC's campus bookstore during these semesters.

A.3 Survey Attrition and Estimation Sample

Table 9 summarizes the number of respondents at various points within the survey. For the Fall 2012 semester, 979 respondents began the questionnaire and 734 (75%) completed the questionnaire. For the Spring 2013 semester, 1002 respondents began the questionnaire and 703 (70%) completed the questionnaire. We only exclude individuals who did not complete the background questions. This leaves 820 respondents from the Fall and 798 respondents from the Spring. Of these individuals, we exclude 104 respondents (52 from both semesters) because they had been enrolled in college for more than 10 semesters and/or summer sessions and an additional 49 respondents (29 from Fall 2012 and 20 from Spring 2013) for reporting nonsensical answers (e.g., reporting an expected price of $100,000).

Table 8: Textbook Scenarios

Book #	Group	Title	Author	Edition	Latest Edition
1	1	Economics: Principles and Policies	Baumol and Blinder	12	Yes
2	1	Introductory Econometrics: A Modern Approach	Wooldridge	4	No
3	1	Experience Sociology	Croteau and Hoynes	1	Yes
4	1	Medical Sociology	Cockerham	12	Yes
5	2	Chemistry: The Essential Science	Brown et al.	12	Yes
6	2	Animal Psysiology: Adaptation and Environment	Schmidt-Nielsen	5	Yes
7	2	Earth: Portrait of a Planet	Marshak	3	No
8	2	Data Structures and Algorithm Analysis in Java	Weiss	3	Yes
9	3	Norton Anthology of Short Fiction	Bausch and Cassill	7	Yes
10	3	Medicine and Morality in Haiti	Brodwin	1	Yes
11	3	Western Civilization, Volume 1	Perry	10	Yes
12	3	Voces de Hispanoamerica	Chang-Rodrigues and Filer	4	Yes

Book #	Year Published	Subject	Course	Cover	Pages	Weight (lbs)	ISBN	New Bookstore Price
1	2012	Economics	101	Hard	880	4.4	9780538453577	$286
2	2008	Economics	570	Hard	896	3.4	9780324581621	$262
3	2012	Sociology	101	Paper	576	2.8	978007319353	$141
4	2011	Sociology	469	Paper	432	1.2	9780205054183	$87
5	2011	Chemistry	101	Hard	1200	5.4	9780321696724	$253
6	1997	Biology	451	Hard	617	3.4	9780521570985	$104
7	2007	Geology	101	Paper	880	4.4	9780393935189	$146
8	2011	Computer Science	410	Hard	640	2.1	9780132576277	$151
9	2006	English	123	Paper	1776	2.8	9780393926118	$81
10	1996	Anthropology	470	Paper	260	1.0	9780521575430	$52
11	2012	History	151	Paper	496	1.6	9781111831707	$172
12	2012	Spanish	400	Hard	706	2.2	9781111837921	$213

Table 9: Survey Attrition

	Fall 2012 Respondents	Percent Remaining	Spring 2013 Respondents	Percent Remaining
Began the questionnaire	979	100	1002	100
Completed the background questions	820	83.8	798	79.6
Completed at least one scenario	759	77.5	761	75.9
Completed at least two scenarios	741	75.7	716	71.5
Completed the questionnaire	734	75.0	703	70.2

Table 10: Results Using Alternative Samples

	Main Sample			Extended Sample			Single Scenario		
Experience	N	Mean	SD	N	Mean	SD	N	Mean	SD
No Online Purchases	440	0.833	0.172	473	0.833	0.186	140	0.833	0.205
1-5 Online Purchases	923	0.819	0.153	1016	0.812	0.161	296	0.797	0.175
6-10 Online Purchases	966	0.783	0.161	1015	0.778	0.170	297	0.758	0.190
11+ Online Purchases	1706	0.770	0.157	1905	0.771	0.169	531	0.754	0.174

B Robustness Checks

This section investigates the robustness of the results presented in the paper by providing results from a number of other specifications. Specifically, we explore the robustness of our results by varying the following:

B.1 The criteria for being omitted from the sample and the number of scenarios used for each respondent.

B.2 The distributional assumption.

B.3 Sample for the test for selection.

We also provide evidence supporting the log-normal distribution assumption for the price distribution in section B.4

B.1 Omission Criteria

Respondents are omitted from our sample in the main body of the paper for two reasons:

1. Being enrolled in more than 10 semesters of college.

2. Reporting expectations less than 10% of the bookstore price or greater than 150% of the bookstore price.

The first criteria is used to focus on traditional college students. The second criteria is used to eliminate respondents who we believe did not take the questionnaire seriously (for example, individuals who reported expectations of $0 or $100,000). To make sure that our results are not biased because of these omissions, we relax these criteria.

We proceed to report the main findings from the paper for a less stringent omission criteria. Specifically, we only omit respondents who report expectations less than 1% of the bookstore price or greater than 200% of the bookstore price, and students in more than 14 semesters of college. This results in one respondent being omitted from the Fall sample and one respondent being omitted from the Spring sample for reporting expectations below 1% of the bookstore price, and three respondents being omitted from the Fall sample and two respondents being omitted from the Spring sample for reporting expectations greater than 200% of the bookstore price. Eight respondents in the Fall sample and 7 respondents in the Spring sample are omitted for reporting being enrolled in more than 14 semesters of college.

Table 10 reports the mean ratio of expectations to bookstore prices for the main sample as well as the extended sample. Including outliers does not significantly change the estimates of mean price expectations. Also included in Table 10 are the results that only use the first hypothetical textbook purchasing scenario that an individual responded to (out of a potential of six for individuals who completed the questionnaire in the fall and spring semester). The price expectations for the first scenario are lower than for the full sample, but the relationship between experience and price expectations is the same for both groups.

B.2 Alternative Distributional Assumption

Table 11 reports the results for the normal and gamma distributions. Also included are the results for the log-normal and gamma distributions using a more restrictive sample. Since the normal distribution requires dropping individuals who report a greater than 50% probability of being below the threshold, the restricted samples are constructed using a similar rejection criteria. The samples are constructed by dropping individuals who report a 50% or greater probability of being below the threshold. Then the parameter values are

Table 11: Distribution Results

	New		
Experience	N	mean $\mathbf{E}_i(p)$	mean $\sqrt{\mathbf{Var}_i(p)}$
	Normal		
No online purchases	301	0.835*** (0.135)	0.139 (0.099)
1-5 online purchases	663	0.822*** (0.129**)	0.141* (0.097)
6-10 online purchases	696	0.796 (0.130)	0.130 (0.089*)
11+ online purchases	1257	0.785*** (0.134)	0.133 (0.096)
	Gamma		
No online purchases	412	0.836*** (0.164)	0.237 (0.288)
1-5 online purchases	870	0.819*** (0.152*)	0.243* (0.312)
6-10 online purchases	908	0.785* (0.158)	0.224 (0.276)
11+ online purchases	1604	0.774*** (0.153)	0.217* (0.269)
	Log-Normal - Restricted Sample		
No online purchases	301	0.835*** (0.135)	0.120 (0.069**)
1-5 online purchases	663	0.822*** (0.129**)	0.122** (0.066)
6-10 online purchases	696	0.796 (0.130)	0.114 (0.060**)
11+ online purchases	1257	0.785*** (0.134)	0.115* (0.064)
	Gamma - Restricted Sample		
No online purchases	325	0.847*** (0.164)	0.126** (0.072***)
1-5 online purchases	690	0.821*** (0.151*)	0.125*** (0.070**)
6-10 online purchases	708	0.788** (0.153)	0.114* (0.063**)
11+ online purchases	1263	0.780*** (0.150)	0.114*** (0.063**)

Notes: Standard deviations in parenthesis. The significance levels reported for the mean values are from a two-sample equality of means test. The significance levels for the standard deviations are from Brown and Forsythe's alternative formulation of Levene's robust two-sample equality of variances test.

* refers to p–value $< .1$; ** $< .05$; *** $< .01$

calculated for each individual, and the final sample includes individuals whose parameter values are not in the top or bottom 2.5% of values for either parameter. The results for the log-normal and gamma using the restricted samples does not change the mean of the price expectations by a large amount, but the variability of the expected price distribution falls substantially. The higher mean variability of the expected price distribution using the less restrictive sample for the log-normal and gamma distributions is driven by the individuals who report a high probability of the price being below the threshold. Ultimately, the results are similar regardless of the distribution used.

B.3 Alternative Sample for the Test for Selection

In this section, we restrict the test for selection to the individuals who report no experience in the fall semester (22 individuals with between 47 and 56 scenarios). Since the measure of experience is an interval, individuals who remain in the same interval for both fall and spring may or may not have gained experience. By restricting the sample to individuals who have no experience in the fall semester, we are able to test between a group that gained experience from fall to spring and one that does not gain experience. The results of this test are that there is no clear evidence of selection, however, the sample is too small to make any definitive conclusion.

B.4 Price Distribution

In this section we provide some evidence supporting the use of the log normal price distribution as well as discussing some limitations of the distribution in fitting certain features of the empirical price distribution. Tests for normality reject the assumption of normality for both the distribution of prices and the log of prices for most specifications of the empirical distribution. Generally, the normal distribution is able to fit the data slightly better than

Table 12: Parameter Values by Change in Experience

	Fall			Spring			Difference	
Group	N	mean μ_i	mean σ_i	N	mean μ_i	mean σ_i	mean μ_i	mean σ_i
Increase Exp.	45	-0.243 (0.163)	0.216 (0.148)	51	-0.291 (0.272)	0.236 (0.274)	0.048	-0.020
Same Exp.	58	-0.311 (0.289)	0.269 (0.278)	54	-0.310 (0.269)	0.271 (0.283)	-0.001	0.002
Difference		-0.068	0.052		-0.019	0.035		

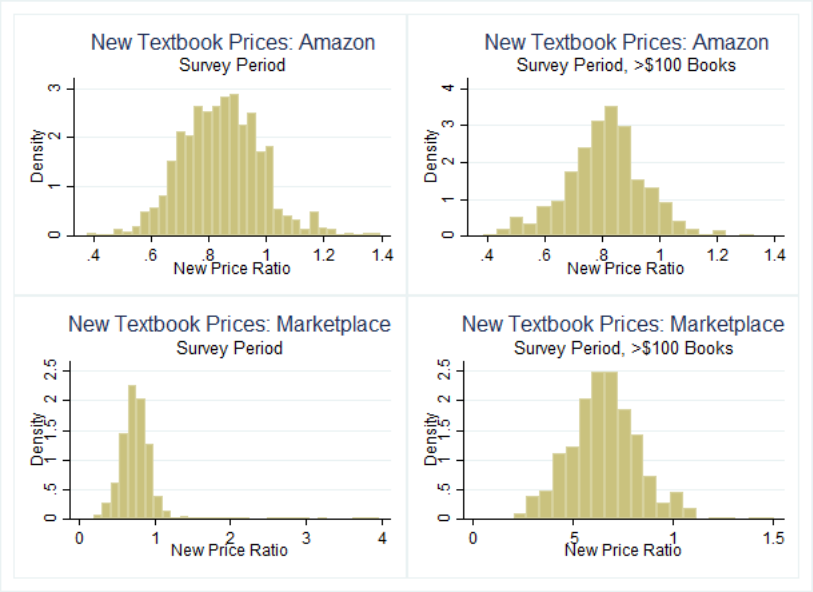


Figure 2: Histograms of Prices, Survey Period

the log-normal distribution. In order for the log normal distribution to fit the long right tail of the price distribution, the result is that it places too little weight on the left tail relative to the empirical distribution. The analysis in this paper is not dependent on a particular distributional assumption. In structural search models, however, an incorrect distributional assumption on the individual's beliefs about the price distribution or about the empirical price distribution can significantly bias estimates. Figure 2 displays histograms of the empirical prices for the Fall survey period. Figures 4 and 5 display kernel density estimates for used and new prices.

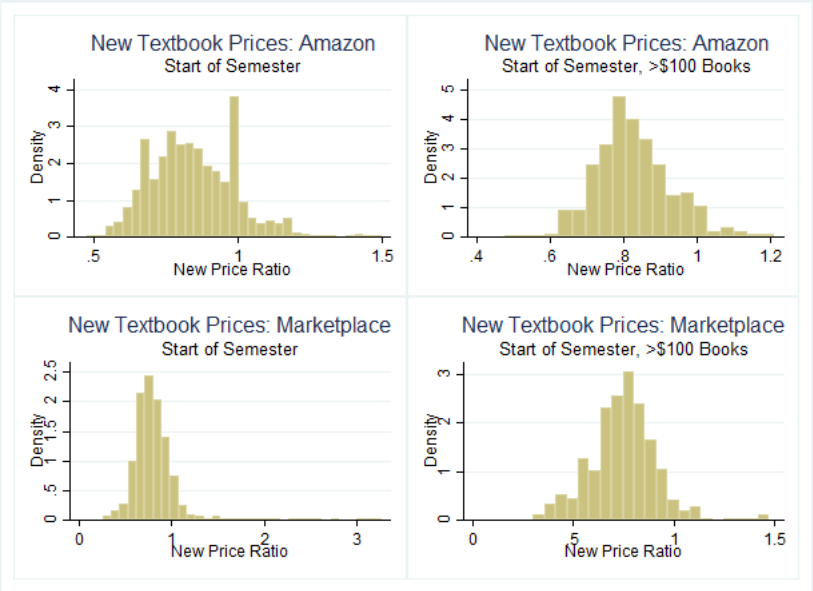


Figure 3: Histograms of Prices, Start of Semester

Figure 4: Kernel Density Estimate, Survey Period

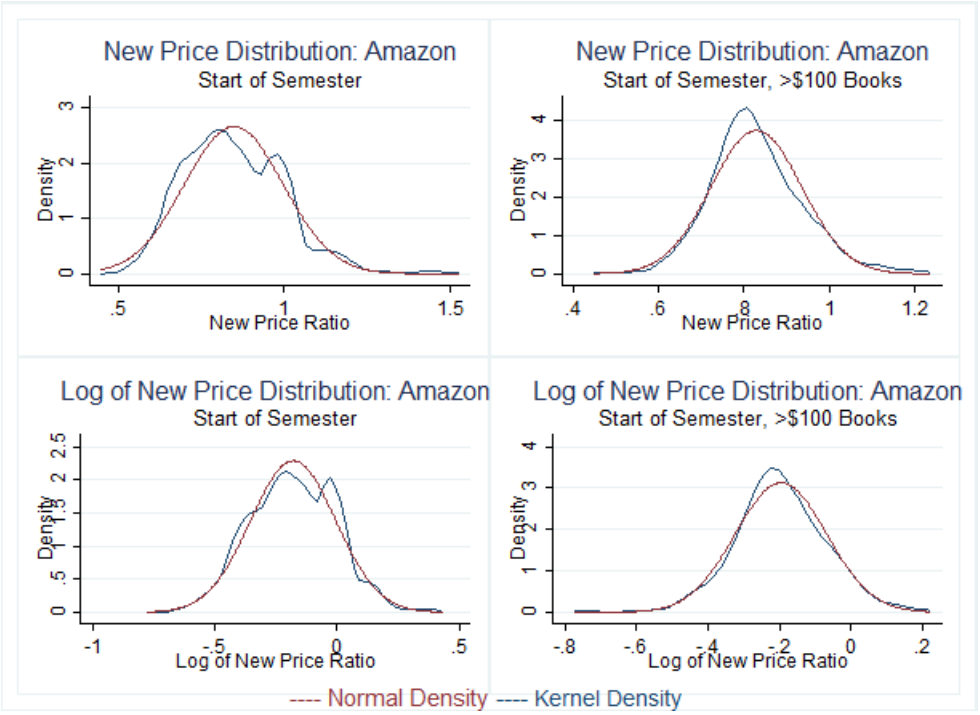


Figure 5: Kernel Density Estimate, Start of Semester

Figure 6: Kernel Density Estimate, Amazon Marketplace, Survey Period

Figure 7: Kernel Density Estimate, Amazon Marketplace, Start of Semester

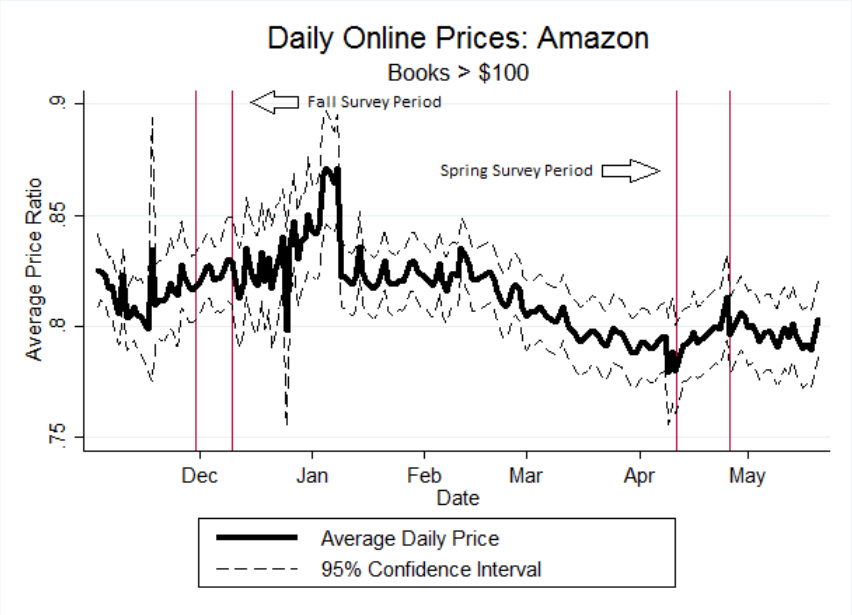


Figure 8: Daily New Prices

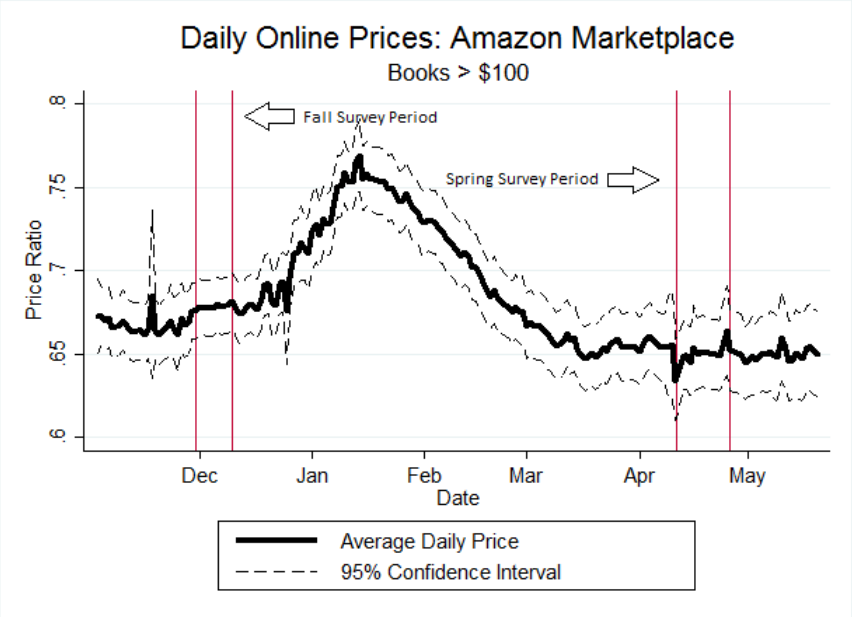


Figure 9: Daily New Prices: Amazon Marketplace

References

Arcidiacono, P., Hotz, J., and Kang, S. (2012). Modeling College Major Choice using Elicited Measures of Expectations and Counterfactuals. *Journal of Econometrics*, 166(1):3–16.

Burdett, K. and Judd, K. (1983). Equilibrium Price Dispersion. *Econometrica*, 51(4):955–969.

De los Santos, B., Hortacsu, A., and Wildenbeest, M. (2012). Testing Models of Consumer Search using Data on Web Browsing and Purchasing Behavior. *American Economic Review*, 102(6).

De los Santos, B., Hortacsu, A., and Wildenbeest, M. (2013). Search with Learning. *Working Paper*.

Dube, J., Hitsch, G., and Rossi, P. (2010). State Dependence and Alternative Explanations for Consumer Inertia. *RAND Journal of Economics*, 41(3):417–445.

Einav, L. (2005). Informational Asymmetries and Observational Learning in Search. *Journal of Risk and Uncertainty*, 30(3):241–259.

Giulietti, M., Waterson, M., and Wildenbeest, M. (2014). Estimation of Search Frictions in the British Electricity Market. *Journal of Industrial Economics*, 62(4):555–590.

Hong, G. and Shum, M. (2006). Using Price Distributions to Estimate Search Costs. *RAND Journal of Economics*, 37(2):257–275.

Honka, E. (2014). Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry. *RAND Journal of Economics*, 45(4):847–884.

Honka, E. and Chintagunta, P. K. (2015). Simultaneous or Sequential? Search Strategies in the U.S. Auto Industry. *Working Paper*.

Hortaçsu, A. and Syverson, C. (2004). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. *Quarterly Journal of Economics*, 119(2):403–456.

Koulayev, S. (2009). Estimating Search with Learning. *Net Institute Working Paper No. 08-29*.

Koulayev, S. (2013). Search with Dirichlet Priors: Estimation and Implications for Consumer Demand. *Journal of Business and Economic Statistics*.

Koulayev, S. (2014). Search with Dirichlet Priors: Estimation and Implications for Consumer Demand. *RAND Journal of Economics*, 45(3):553–575.

Moraga-González, J., Sándor, Z., and Wildenbeest, M. (2015). Consumer Search and Prices in the Automobile Market. *Working Paper*.

Pires, T. (2015). Consideration Sets in Storable Goods Markets. *Working Paper*.

Reinganum, J. (1979). A Simple Model of Equilibrium Price Dispersion. *Journal of Political Economy*, 87(4):851–858.

Seiler, S. (2013). The Impact of Search Costs on Consumer Behavior: A Dynamic Approach. *Quantitative Marketing and Economics*, 11(2):155–203.

Sonnemans, J. (1998). Strategies of Search. *Journal of Economic Behavior and Organization*, 35(3):309–332.

Spence, F. (2015). Consumer Experience and the Value of Search in the Online Textbook Market. *Working Paper*.

Stigler, G. (1961). The Economics of Information. *Journal of Political Economy*, 69(3):213–225.

Stinebrickner, T. and Stinebrickner, R. (2011). Math or Science? Using Longitudinal Expectations Data to Examine the Process of Choosing a College Major. *NBER Working Papers 16869*.

Wiswall, M. and Zafar, B. (2015). Determinants of College Major Choices: Identification from an Information Experiment. *Review of Economic Studies*, 82(2):791–824.