Identification and Estimation of Online Price Competition with an Unknown Number of Firms*

Yonghong An† Michael R. Baye‡ Yingyao Hu§ John Morgan¶
Matt Shum‖

This version: August 2015

Summary

This paper considers identification and estimation of a general model for online price competition. We show that when the number of competing firms is unknown, the underlying parameters of the model can still be identified and estimated employing recently developed results on measurement errors. We illustrate our methodology using UK data for personal digital assistants and employ the estimates to simulate competitive effects. Our results reveal that heightened competition has differential effects on the prices paid by different consumer segments.

Keywords: E-Retail Markets, Nonparametric Identification, Structural Estimation

*We are grateful to the co-editor Thierry Magnac and three anonymous referees for their exceptionally helpful comments. This research began while Baye was serving as the Director of the Bureau of Economics at the Federal Trade Commission. We thank his former colleagues there, especially Dan O'Brien and Dan Hosken, for helpful discussions. We also thank seminar participants at Northwestern University and University of Connecticut (Department of Agricultural and Resource Economics) for comments on a preliminary draft. Morgan thanks the National Science Foundation for financial support.

†Corresponding author. Department of Economics, Texas A&M University, College Station, TX 77845; email: y.an@tamu.edu.

‡Department of Business Economics & Public Policy, Kelley School of Business, Indiana University, Bloomington, IN 47405.

§Department of Economics, Johns Hopkins University, Baltimore, MD 21218.

¶Haas School of Business and Department of Economics, University of California, Berkeley, CA 94720.

‖Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA 91125.

1 Introduction

A fundamental query in industrial organization and antitrust concerns the impact of market structure on consumer surplus. Such analysis can be controversial and challenging in even fairly simple brick-and-mortar retail environments. One major complication is that the relevant number of competitors in such markets is often unobserved by the econometrician, especially in online markets. To the best of our knowledge, there are no ready tools available to empirically analyze and assess potential competitive effects in such situations. The absence of such tools or analyses stems, in part, from the fact that (1) online prices display considerable price dispersion, which substantially complicates predicting the price effects for a change of market structure; and (2) the number of (potential) competitors in the online channel is typically an unknown.

This paper represents a first attempt to quantify the competitive effects of changes in the number of firms in an online market when the number of sellers is unobserved. The model that we structurally estimate assumes (1) online firms are symmetric, pure-play e-retailers; (2) the number of (potential) online competitors at any point in time is known to firms but not to the econometrician; and (3) online buyers may be segmented into two types: price sensitive "shoppers," who rely on a price comparison site to find the best deal, and price insensitive "loyals," who simply visit their preferred online firm's website. This benchmark environment is the standard framework for modeling e-retail competition; see Baye et al. (2006) for a survey of this literature.

We first present a general model of online price competition that nests standard models ranging from Varian (1980) to Iyer et al. (2005) as special cases. The model enriches existing models of online price competition, including Baye and Morgan (2001), by adding two realistic features: (1) firms pay platforms for clicks; and (2) not all clicks result in sales. In such a model, the observed price distribution represents a combination of the realized number of firms choosing to list on the site together with their realized prices, both of which are stochastic and depend on the unobserved number of actual competitors. Thus, it is essential to recover the price distribution conditional on the true number of competitors in order to estimate the model parameters and conduct competitive analysis. We show that, using the results from the recent econometric literature on misclassfication (e.g., Hu (2008)), this price distribution can be nonparametrically identified from the observed number of competitors and the listed prices. Based on the results of this identification, we present a two-step procedure to estimate model parameters. A Monte Carlo experiment (see the online Appendix) demonstrates that our procedure performs well, and that failing to account for the unobserv-

ability of the potential number of firms can lead to biased estimates of model parameters. As an application of our methodology, we structurally estimate the model based on UK data for personal digital assistants (PDAs), and then use these estimates to simulate the competitive effects for changes in the number of competitors.

Our empirical results indicate that, at least in some instances, competitive effects in online markets are more similar to those predicted by the simple homogeneous product Bertrand model than might be expected given the price dispersion observed in (and predicted by theoretical models of) e-retail markets. However, unlike in Bertrand models, there are also distributional effects of equilibrium pricing, e.g., if the number of firms decreases from three to two, the average transaction price paid by price sensitive "shoppers" increases by 6.89 percent, while the average transaction price paid by consumers "loyal" to a particular firm *decreases* by 3.19 percent.[1]

This paper makes two main contributions: (1) to present a methodology for identification and estimation of online retail markets (or indeed any market characterized by a first-price all-pay auction) in which the number of competitors/bidders is observed by insiders but unknown to the econometrician; and (2) to provide some empirical evidence on how consumer surplus is affected by the potential number of competitors in an e-retail market.

Most closely related to our work is An et al. (2010), which identifies and estimates a model of first-price auctions where the number of bidders is known to the auction participants but unobserved by the researcher. The present paper differs from An et al. (2010) in a number of respects. First, we focus on a completely different pricing environment; they analyze a standard first-price winner-pay auction, while our environment mirrors that in a first-price all-pay auction. Second, our application highlights potential distributional effects of changes in competition in an online rather than auction market. Furthermore, the present paper enriches the econometric methodology employed in An et al. (2010) by showing that the method works well for a modest-sized sample even though the estimation involves large dimensional matrices.[2]

We organize the rest of the paper as follows. In Section 2, we present a general model of online price competition. In Section 3, we show the nonparametric identification results. In Section 4, we describe a two-step estimation procedure for the proposed general model. In Section 5, we present an empirical application of our methodology using UK data for PDAs. Section 6 concludes. The appendix contains miscellaneous proofs. Monte Carlo evidence

[1] Armstrong (2008) points out that similar distributional effects are theoretically possible in the context of consumer protection policy, while Baye (2008) notes that this is a theoretical possibility in antitrust.

[2] In our setting, the matrix dimension is 10×10, whereas in An et al. (2010) it is 3×3.

and some technical details omitted in the paper are provided as supporting information in an online Appendix.

2 Model of Online Price Competition

It is by now well established that price dispersion is considerable and ubiquitous in online markets. Much of the prior empirical literature presumes such dispersion stems from costly consumer search. This assumption is appropriate in environments where the information obtained from an initial search query is insufficient to make a buying decision and hence subsequent investigation is required. For example, a consumer might jump from website to website, perhaps guided by the results of the search query, paying the associating (implicit) cost to search, and stopping as in standard models of offline search. Hong and Shum (2006), Moraga-González and Wildenbeest (2008), and Moraga-González et al. (2013) use such a model to derive a family of dispersed price distributions, which they then structurally estimate to recover the implied distribution of consumer search costs.

While these models are appropriate in environments where it is costly to compare the prices that different firms charge for their desired product, they seem less appropriate when the initial search is conducted on a price comparison site (such as Kelkoo, which we examine in our empirical application). These sites return the all-in prices (including shipping and VAT) that each seller charges for an identical product. Consumers may readily sort these results to identify the firm charging the lowest price (or a variety of other criteria). Similar shopping experiences are available on platforms such as Amazon, where a search query also provides detailed information of products and a seller's reputation. In such settings, the scope of information available just from the initial search query significantly diminishes the need to visit seller websites (other than to purchase) and hence drastically reduces the implied search costs.

Search cost models typically assume that consumers are identical except for costs to search, yet e-retailers spend vast amounts of money on brand marketing. Presumably they do this because they believe that consumers can be made "sticky", loyal in the language of our model, through brand attachment. The model we describe, and which is the basis for our empirical application, takes these loyalty distinctions seriously, but abstracts from search costs. In particular, we assume that search costs are zero within the platform (e.g., it is costless to compare the prices returned at the comparison site), that N firms are in the market for a given good (though not necessarily listing at the platform), and that there are $N + 1$

distinct consumer segments. N of these segments represent customers loyal to a particular seller while the last segment consists of consumers holding no loyalties and choosing purely on the basis of price. Of course, further segmentation is possible. For instance, a richer model might allow loyalty to multiple firms, perhaps with slight preferences across each. Taken to extremes, one could have up to 2^N unique loyal segments, as well as a segment that engages in sequential or fixed-sample search across other platforms. While theoretically possible, such a model is of limited empirical use owing to the data demands required to separately identify each of the many segments, let alone the other parameters of the model. Thus, for practical as well as parsimony reasons, we stick to the simpler specification.

To fix ideas, suppose the market consists of a commonly known number of firms ($N > 1$) that produce at a constant marginal cost of $m \geq 0$. Firms offer identical products for sale through their individual websites, which may have different characteristics or provide different types of service. Some consumers, who we call "loyals," value these services and purchase by directly visiting the website of their preferred firm. Other consumers, who we call "shoppers," care only about price. They first access a price comparison site to obtain a listing of the prices charged by sellers advertising at the site and click through to the firm offering the lowest price. If no prices are listed, they visit the website of a randomly selected firm.[3] All consumers have unit demand and a maximal willingness to pay r.

It is widely recognized that conversion rates in online markets are low—only a fraction of consumers that click on a price at a comparison site follow through by making a purchase. Loyal consumers may browse at their preferred company website, perhaps in search of new offerings, but not necessarily with the intention of buying. Shoppers, even after locating the lowest price offer, may opt not to buy, perhaps because of a bad user experience or because they were simply looking rather than buying on this particular occasion. To account for this, we assume that, with probability $\gamma \in (0, 1]$ a consumer actually makes a purchase. At the price comparison site, γ may be interpreted as the conversion rate, the fraction of clicks that turn into sales. Finally, we assume a fixed total number of loyals, M, divided evenly among symmetric competing firms, so each firm attracts M/N loyals. Moreover, there are a total of $S > 0$ shoppers.

We now turn to the details of firm behavior. To advertise at the comparison site, a firm

[3]We assume that shoppers too are loyal in the following sense: They single home to the modeled price comparison site and do not jump to another site in the event of unsuccessful search. This ensures that the model also applies to environments with multiple comparison sites. Under this assumption, Proposition 1 in Baye and Morgan (2001) offers mild conditions that ensure that the posited search strategy by shoppers is indeed optimal when endogenized.

must pay an (explicit or implicit) amount $\phi > 0$ to list its price, plus a cost per click (CPC) of $c \geq 0$ each time a consumer clicks on its price advertisement (listing). Thus, firm i's strategy consists of a continuous pricing decision (p_i) and a zero-one decision to advertise its price at the comparison site. Let α_i denote the probability that firm i chooses to advertise on the comparison site. A firm that does not advertise its price on the comparison site avoids paying listing and clickthrough fees, but at the potential cost of failing to attract the shoppers visiting the comparison site. The parameters of the model are commonly known by firms.

When platform fees are not too high, one can readily show that there is an active market for listings at the comparison site. A closed form characterization of both advertising and pricing is available for the case where firms are symmetric, which we report below. For this case, we characterize the symmetric equilibrium pricing and advertising strategies of firms competing in this online environment. Please see the online Appendix (supporting information) for a proof.

Proposition 1 *Suppose that firms are symmetric, and listing and clickthrough fees are not too high,* $0 < \phi < S\left((r-m)\,\gamma\frac{N-1}{N} - c\right)$ *and* $0 \leq c < (r-m)\,\gamma\frac{N-1}{N}$. *Then in a symmetric Nash equilibrium:*

(a) Each firm lists its price on the comparison site with probability

$$\alpha^* = 1 - \left(\frac{\phi}{S\left((r-m)\,\gamma\frac{N-1}{N} - c\right)}\right)^{\frac{1}{N-1}} \in (0,1)$$

(b) Conditional on listing a price at the comparison site, a firm's advertised price may be viewed as a random draw from

$$F^*(p) = \frac{1}{\alpha^*}\left(1 - \left(\frac{(r-p)\,\gamma\frac{M}{N} + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1) - Nc}\phi}{S\left((p-m)\,\gamma - c\right)}\right)^{\frac{1}{N-1}}\right) \tag{1}$$

on $[p_0, r]$, *where* $p_0 = m + \frac{1}{(S\gamma + M\gamma)}\left(\gamma\frac{M}{N}(r-m) + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1) - Nc}\phi + Sc\right) \in (m, r)$.

(c) A firm that does not advertise on the comparison site charges a price of $p_i = r$ *on its own website.*

(d) Each firm earns an expected profit of $E\pi = (r-m)\,\gamma M/N + \frac{\phi}{N\left(1 - \frac{c}{(r-m)\gamma}\right) - 1}$.

Note that in the case of monopoly ($N = 1$), the firm's strategy is to charge r on its own website but not advertise on the comparison site. The monopolist attracts both loyals and shoppers to its site and the expected profit is $(r-m)\gamma(M+S)$.

Our model above extends the original Baye and Morgan (2001) model to an environment in which all transactions take place online, and accounts for clickthrough fees as well as conversion rates that are potentially less than unity. Consistent with the empirical literature, the model implies that prices listed at the comparison site are necessarily dispersed in equilibrium, and that the number of firms actually listing prices at the comparison site on any given date is generally less than the total number of firms in the market, e.g., see Baye et al. (2006) for a survey of about twenty studies documenting price dispersion of 10 to 50 percent in online markets. This model nests a variety of other models as special cases, including Rosenthal (1980), Varian (1980), Narasimhan (1988), Iyer and Pazgal (2003), Baye et al. (2004) , and Iyer et al. (2005). Unlike some of these special cases, our general model is flexible enough to allow differing competitive effects on consumer surplus. For instance, the Rosenthal model implies that when there are two or more competitors, average prices paid by all consumers *rise* with the number of competing firms.

Under the maintained hypothesis that firms' listed prices are distributed according to equation (1), it is, in principle, possible to estimate the underlying parameters of the model. Unfortunately, data from price comparison sites reveal A, the realized number of firms choosing to list prices at the site at a given time, but not N, the total number of firms in the market. The model implies that A is a binomially distributed random variable with parameters (α, N). The extant literature mostly finesses this problem. For example, Baye et al. (2004), Moraga-González and Wildenbeest (2008), and Moraga-González et al. (2013) use the number of observed prices as a proxy for N, in effect assuming that $N = A$. Hong and Shum (2006) assume that $N = +\infty$ in their identification of price dispersion models. The problem of the unobservability of N presents econometric challenges, especially when it varies over time. The next section offers an identification procedure that explicitly accommodates the unobservability of N.

3 Identification

The model of online price competition in Section 2 is modified from a low-bid auction in which the firm offering the lowest price secures the price sensitive shoppers when it lists on the comparison site. Unlike a standard auction, where the payoff to a losing firm is independent of its bid; in our setting the payoff (i.e. profits from sales to loyals) varies directly with one's bid—a more aggressive (i.e. lower) losing bid produces smaller rewards than a less aggressive losing bid. Thus, the setting may be thought of as a type of reverse first-price

all-pay auction. Viewed from this perspective, it is, *ex ante*, far from clear that estimation techniques designed for winner-pay auctions will also work for their all-pay cousins. In this section we show how one can adapt the techniques of econometric analysis built on a winner-pay auction model to our (all-pay) setting. Specifically, we show that the equilibrium distribution of prices in Proposition 1 (along with one additional but rather mild condition) implies the identification conditions for standard auctions pioneered by Hu (2008) and An et al. (2010).

Following these authors, suppose the maximum number of (potential) firms is K, and is known to the econometrician. The actual number of firms $N \in \{2, 3, \cdots, K\}$, which may vary, is common knowledge to the firms but unknown to the econometrician. We preclude the case of monopoly ($N = 1$) since the firm only uses its own website and there will be no data on the comparison site. Nonetheless, once the model parameters (r, m, γ, M, S) are identified, we can recover the price charged by the monopolist r as well as its expected profit $(r - m)\gamma(M + S)$. Let A denote the number of price listings on a given date and p_j be the j-th listed price, $j = 1, \cdots, A$. For reasons that will become clear, consider only dates in which two or more firms listed prices, i.e, $A \geq 2$. As will be shown, identification of price distribution in equation (1) will not be affected by dropping the observations with $A = 1$, whereas recovering probability $\Pr(N)$ requires these observations.

We randomly choose a listed price $p_l, l \in \{1, 2, \cdots, A\}$ from the A prices for each date for estimation; from the remaining $A - 1$ prices we choose another one p_m, where $m \in \{1, 2, \cdots, A\}$ and $m \neq l$, as our instrumental variable. To accommodate the discreteness of the potential number of firms, we discretize p_m to construct a discretized instrumental variable Z:

$$Z = \begin{cases} 2 & \text{if } p_m \in [\underline{p}, p_{(1)}], \\ 3 & \text{if } p_m \in (p_{(1)}, p_{(2)}], \\ \vdots & \\ K & \text{if } p_m \in (p_{(K-2)}, \overline{p}], \end{cases}$$

where the support of prices, $[\underline{p}, \overline{p}]$, is divided into $K - 1$ intervals by the $K - 2$ cutoff points: $\underline{p} < p_{(1)} < p_{(2)} < \cdots < p_{(K-2)} < \bar{p}$. The choice of the $K - 2$ cutoff points does not affect our identification and estimation given Z satisfies a mild condition (Condition 1 below) which is empirically testable.

From the econometrician's point of view: (a) N, A, and Z share the same support $\{2, ..., K\}$; (b) r, m, ϕ, γ, M, and S are unknown parameters; and (c) N is unobservable or in dispute. In the application that follows, the cost-per-click (c) is data and hence is not

included in the set of parameters to be estimated. Let $\theta \equiv (r, m, \phi, \gamma, M, S)$. Under the hypothesis that the prices at the comparison site are generated according to F^* in equation (1), we may write the underlying (undiscretized) distribution of prices as $F(p|N;\theta)$ with its density being $f(p|N;\theta)$. To ease the notational burden, we suppress θ whenever there is no confusion. The lemma below shows that the equilibrium density of listed prices is independent of A and Z:

Lemma 1 $f(p|N) = f(p|A \geq 2, Z, N)$.

This result follows directly from the fact that firms' prices are determined prior to their knowing realizations of A and Z. Next, notice that, given the data and the model, conditional on the fact that at least two firms list prices the probability that exactly A firms list at the comparison site is

$$g(A|N, A \geq 2) = \frac{\binom{N}{A}(\alpha)^A (1-\alpha)^{N-A}}{1-(1-\alpha)^N - N\alpha(1-\alpha)^{N-1}} \text{ for all } 2 \leq A \leq N. \tag{2}$$

It immediately follows that

Lemma 2 $g(A|N, A \geq 2) = g(A|Z, N, A \geq 2)$.

Lemma 1 implies that auxiliary variables A and Z only affect the equilibrium density of prices through the unobservable number of firms, N. Analogously, Lemma 2 states that the instrument Z affects the number of listed prices only through N. To recover the price distribution $f(p|N)$, we require the existence of two auxiliary variables (A, Z) that satisfy Lemma 1 and 2, where A is a mis-measured version of N and Z is an instrumental variable of N. Roughly, because A is a noisy measure of N, the two lemmas require that the noise is independent of the instrument Z, conditional on N. A good candidate for the instrument Z is a discretized second price because as shown in proposition 1, a price depends on N through all the model parameters $\theta \equiv (r, m, \phi, \gamma, M, S)$. In general, an instrumental variable Z would require $p \perp Z|N$ and $A \perp Z|N$, and Lemma 1 and 2 provide such an instrument conditional on $A \geq 2$. The use of a second price in the role of the instrument Z also requires at least two listings per date.[4]

Let $h(p, A, Z|A \geq 2)$ denote the observed joint density of p, A and Z given two or more firms list their prices. Let $\psi(N, Z|A \geq 2)$ denote the joint density of N and Z conditional on

[4]Of course, the restriction $A \geq 2$ may be relaxed if an instrumental variable other than the second price is available.

$A \geq 2$, which is unobserved because N is an unknown to the econometrician. This specification allows for the possibility that the true number of firms N might vary across products and over time without placing parametric restrictions on the data-generating process in this respect. Now, the law of total probability implies the following relationship between the observed and latent densities:

$$\begin{aligned} h\left(p, A, Z | A \geq 2\right) &= \sum_{N=2}^{K} f(p|N, A \geq 2, Z) g(A|N, Z, A \geq 2) \psi(N, Z|A \geq 2) \\ &= \sum_{N=2}^{K} f(p|N) g(A|N, A \geq 2) \psi(N, Z|A \geq 2), \end{aligned} \tag{3}$$

where the second equality follows from Lemmas 1 and 2. Define:

$$\begin{aligned} \left(H_{p,A,Z}\right)_{i,j} &= h(p, A = i, Z = j | A \geq 2) \\ \left(G_{A|N}\right)_{i,k} &= g\left(A = i | N = k, A \geq 2\right) \\ \left(\Psi_{N,Z}\right)_{k,j} &= \psi\left(N = k, Z = j | A \geq 2\right), \end{aligned}$$

and the diagonal matrix

$$F_{p|N} = diag\big(f(p|N = 2), f(p|N = 3), \cdots, f(p|N = K)\big). \tag{4}$$

All of these are $(K-1)$-dimensional square matrices. Then equation (3) may be written in matrix notation as:

$$H_{p,A,Z} = G_{A|N} F_{p|N} \Psi_{N,Z} \tag{5}$$

Next, consider the observed joint density of A and Z. Again, the law of total probability together with Lemma 2 enables us to obtain

$$b(A, Z|A \geq 2) = \sum_{N=2}^{K} g(A|N, A \geq 2) \psi(N, Z|A \geq 2)$$

or, using matrix notation analogous to that above,

$$B_{A,Z} = G_{A|N} \Psi_{N,Z} \tag{6}$$

Identification requires that the following rank condition to be satisfied:

Condition 1 $\text{Rank}\left(B_{A,Z}\right) = K - 1$.

Since both A and Z are observables, Condition 1 is empirically testable from the data. Equation (6) implies $\text{Rank}\left(B_{A,Z}\right) \leq \min\left\{\text{Rank}\left(G_{A|N}\right), \text{Rank}\left(\Psi_{N,Z}\right)\right\}$. Considering that

all the three matrices are of dimension $(K-1)\times(K-1)$, Condition 1 is equivalent to a restriction that both $G_{A|N}$ and $\Psi_{N,Z}$ are of full rank and invertible. The full rank condition for $G_{A|N}$ imposes no restrictions to our model. This is because $G_{A|N}$ is upper-triangular by construction ($A \leq N$) and a sufficient and necessary condition of full rank is that all the diagonal elements are nonzero, i.e., $\Pr(A = k|N = k) > 0$, for all $k = 2, \cdots, K$. This condition requires that for any possible number of potential firms (N), there is a positive probability that all of them advertise their products on the comparison site and this is automatically satisfied due to equation (2). The full rank condition of $\Psi_{N,Z}$ restricts its columns to be linearly independent, i.e., there is sufficient variation of the joint probability $\Pr(N, Z|A \geq 2)$. Roughly, it requires that listed prices (recall that Z is a discretized price) to vary sufficiently when there are different potential number of firms N in the market. It is worth noting that the full rankness of $B_{A,Z}$ implies N_t is varying across t. The reasoning is as follows. If there is no variation in N_t, e.g., $N_t = k \in \{2, \cdots, K\}$ for all t then $Rank(G_{A|N}) = 1$ since only one column of this matrix is nonzero. Consequently, rank of the nonzero matrix $B_{A,Z}$ will be one rather than $K-1$.

By inverting both sides of equation (6) and multiplying it to equation (5) from right, we obtain our key identifying equation:

$$H_{p,A,Z}\left(B_{A,Z}\right)^{-1} = G_{A|N}F_{p|N}\left(G_{A|N}\right)^{-1}. \tag{7}$$

The matrix on the left-hand side can be formed from the data. The right-hand side represents an eigendecomposition (also called spectral decomposition) of the left-hand side matrix since $F_{p|N}$ is diagonal (cf. equation (4)). This representation allows us to identify the unknown matrices $F_{p|N}$ and $G_{A|N}$. However, the decomposition must be unique for the purpose of identification. The uniqueness requires that any two of the eigenvalues are distinct, and the eigenvector matrix is normalized. Our theoretical model implies:

Lemma 3 *The eigendecomposition in equation (7) is unique up to a normalization and ordering of the columns of the eigenvector matrix $G_{A|N}$.*

For any price p, the matrix $H_{p,A,Z}\left(B_{A,Z}\right)^{-1}$ has $(K-1)$ eigenvalues, $f(p|N=2), \cdots, f(p|N=K)$. For a given set of parameters θ, $f(p|N) \neq f(p|N')$ holds for any p except on a set of zero Lebesgue measure, whenever $N \neq N'$.[5] It then follows that the $K-1$ eigenvalues of the

[5] An analytical proof of this statement is tedious due to the complexity of the parametric form for $f(p|N)$ in (8), and we do not provide a formal proof. Nevertheless, we do not find any counterexamples by numerically solving $f(p|N) = f(p|N')$.

decomposition in equation (7) are distinct except on a set of zero Lebesgue measure. Hence, the eigendecomposition is unique up to a normalization and ordering of the columns.

With Lemma 3 in hand, it then follows that an eigendecomposition of the observed $H_{p,A,Z}\left(B_{A,Z}\right)^{-1}$ matrix recovers the unknown $F_{p|N}$ and $G_{A|N}$ matrices up to a normalization and ordering of the columns of the eigenvector matrix $G_{A|N}$. There is a clear, appropriate choice for the normalization of the eigenvectors because each column of $G_{A|N}$ should add up to one. The model also implies a natural ordering for the columns of $G_{A|N}$, since the matrix is upper-triangular and has non-zero diagonal entries as we discussed before.

Finally, having recovered $G_{A|N}$, from equation (6) , we have

$$\Psi_{N,Z} = \left(G_{A|N}\right)^{-1} B_{A,Z}$$

and hence $\Psi_{N,Z}$ is also recovered. To summarize, we have shown:

Proposition 2 *Suppose Condition 1 holds. Then $F_{p|N}$, $G_{A|N}$ and $\Psi_{N,Z}$ are identified (with $F_{p|N}$ pointwise in p).*

The identified joint distribution $\psi(N, Z|A \geq 2)$, i.e., the matrix $\Psi_{N,Z}$ needs to be treated with caution. Consider that $\psi(N, Z|A \geq 2) = \frac{\psi(N,Z,A\geq 2)}{\Pr(A\geq 2)} = \frac{\psi(Z,A\geq 2|N)\Pr(N)}{\Pr(A\geq 2)}$, where both $\Pr(A \geq 2)$ and $\psi(Z, A \geq 2|N)$ can be expressed as functions of model parameters but not for the probability $\Pr(N)$. Thus, we consider $\psi(N, Z|A \geq 2)$ as reduced form in the identification argument above. The probabilities $\Pr(N), N = 2, 3, \cdots, K$ might be of interest for the purpose of some counterfactual analyses. A convenient procedure to identify $\Pr(N)$ is to construct a linear system $f(p_j) = \sum_{N=2}^{K} f(p_j|N)\Pr(N)$ where $f(p)$ is the density of listed prices (including prices for $A = 1$); $p_j, j = 1, \cdots, K-1$ are $(K-1)$ prices and $f(p_j|N)$ is identified in Proposition 2. $\Pr(N)$ is then identified as the solution of the linear system if the matrix constructed by $f(p_j|N)$ is full rank. We leave the technical details in Appendix A. Note that we preclude the case of $N = 1$. Thus a monopoly firm's behavior is merely characterized by the model parameters r, m, M, S, and γ, as we discussed in Section 2.

Upon identifying the price distribution conditional on the "true" number of firms $F_{p|N}$, the probability distribution $\Pr(N)$, and the probability of A firms listing their prices $g(A|N)$, the structural link between these identified objectives and the parameter θ specified in Proposition 1 allows us to identify θ. First off, the listing probability α^* for each N can be recovered from $g(A|N)$. The equilibrium condition that characterizes those firms who advertise a price r at the comparison site (please see equation (2) in the online Appendix) implies a tradeoff between the profit $(r - m)\gamma$ and the cost of listing ϕ/S per shopper. By varying N, the profit from shoppers changes, whereas the equilibrium condition holds; this allows us to

identify $(r-m)\gamma$. Consequently, ϕ/S can be recovered from its relationship with $(r-m)\gamma$ at equilibrium. Next, we investigate the dependence of the price distribution $F_{p|N}$ on θ to further pin down the parameters. The CDF $F_{p|N}$ evaluated as a given price p for different N describes the effects of number of firms on listing prices, and we can identify ϕ (recall that $(r-m)\gamma$ is identified) and $(r-p_0)\gamma M$ by varying N. Therefore, S is determined by ϕ/S. For a given N, the distribution of price $F_{p|N}$ evaluated at different prices provides a system of equations for parameters, which allows us to identify γ, r and M.

Note that the parameters associated with loyals, both their number, M and their willingness to pay r, entirely relies on the functional form in equation (1) owing to the absence of sales data to estimate these parameters directly. In principle, this reliance on functional form represents an important limitation to the analysis. Fortunately, our nonparametric identification procedure offers a test of its validity. Specifically, if the nonparametrically identified $F_{p|N}$ agrees with the parametric specification, worries about the dependence of our identification on functional form is reduced.

It is worth noting that, if the analogs of Lemmas 1 and 2 as well as Condition 1 continue to hold in some alternative model of price dispersion, then, since our identification is independent of the functional form of the equilibrium price distribution, the same procedure may be used to recover the parameters of this alternative model as well.

4 Estimation

We now describe how one may use the identification argument to estimate the structural model, given data from a price comparison site. Let $t = 1, 2, \cdots, T$ index each set of price observations. For each t, we observe A_t, the number of firms choosing to list their prices at the comparison site. Let $p_{it},\ i = 1, \ldots, A_t$ denote the $A_t \geq 2$ listed prices at t. Our estimation procedure accounts for the fact that N_t is known to the competing firms at time t but is, in effect, a random variable from the perspective of the econometrician. While we cannot recover the specific value of N_t pertaining to each set of prices at each point in time, we are able to recover its marginal distribution.

To estimate the vector of parameters θ, we use the following two-step estimation procedure: In the first step, we use our key equation (7) to nonparametrically estimate $G_{A|N}$ and $\mathbf{\Gamma}_N$, where $\mathbf{\Gamma}_N$ is defined as a vector of marginal probabilities over the number of listings N conditional on $A \geq 2$. Our methodology closely parallels the approach taken in An et al. (2010) and hence we relegate the detailed derivation of these expressions to Appendix B. In

the second step, based on the parametric form of $F(p|N;\theta)$ in equation (1), we recover the parameters θ by MLE. Let $l(p, A;\theta)$ denote the joint density of prices and number of listings ($A \geq 2$). In equilibrium, A and p are independent conditional on N. Thus, this density may be written as

$$l(p, A;\theta) = \sum_{N=2}^{K} g(A|N, A \geq 2) f(p|N;\theta)\Gamma(N) = e_A G_{A|N} F_{p|N;\theta} \mathbf{\Gamma}_N,$$

where $e_A = (0, 0, ..., 1, ..., 0)$ is a row vector where the 1 appears as the A-th element. Hence the likelihood function $\mathcal{L}_t$ for the t-th set of prices is

$$\mathcal{L}_t = \prod_{i=1}^{A_t} l(p_{it}, A_t;\theta) = \prod_{i=1}^{A_t} e_{A_t} G_{A_t|N} F_{p_{it}|N;\theta} \mathbf{\Gamma}_N.$$

Using the first step estimates, we can express the likelihood function $\mathcal{L}$ as

$$\ln \mathcal{L} = \sum_{t=1}^{T} \sum_{i=1}^{A_t} \ln \left(e_{A_t} \widehat{G}_{A_t|N} F_{p_{it}|N;\theta} \widehat{\mathbf{\Gamma}}_N \right).$$

where $F_{p_{it}|N;\theta}$ is a diagonal matrix with diagonal element being $f(p_{it}|N;\theta)$. From equation (1), it may be shown that the density associated with $F^*(p|N;\theta)$ is given by

$$f^*(p|N;\theta) = \frac{1}{N-1} \left(\frac{1}{\alpha^*} - F^*(p|N;\theta) \right) \left(\frac{\gamma M/N}{(r-p)\gamma M/N + \frac{(r-m)\gamma N - Nc}{(r-m)\gamma(N-1) - Nc}\phi} + \frac{\gamma}{(p-m)\gamma - c} \right) \tag{8}$$

for $p \in [p_0, r]$ and zero otherwise. Note that $\widehat{G}_{A_t|N}$ and $\widehat{\mathbf{\Gamma}}_N$ are estimated using the data, whereas $F_{p_{it}|N;\theta}$ is based on the theory model.

As a check, we also consider the "naïve" case where the number of potential firms is simply taken to be the number of actual firms, i.e., $N = A$. In this case, the likelihood is

$$\ln \mathcal{L} = \sum_{t=1}^{T} \sum_{i=1}^{A_t} \ln f(p_{it}|A_t;\theta), \text{where } f(p|A;\theta) = f^*(p|N;\theta) \tag{9}$$

Notice that we deal with the probability $G_{A|N}$ nonparametrically while its parametric form (2) is available. The advantage of such an approach is twofold. First, a nonparametric probability matrix $G_{A|N}$ has the flexibility to incorporate general models of entry. For example, a tractable parametric form similar to (2) may be difficult to obtain in a model with endogenous entry while our nonparametric approach may still apply. Second, the approach provides a specification test of the model, i.e., to test the difference between the

parametric and nonparametric version of $G_{A|N}$, where the former matrix can be computed using the estimate of parameters and the latter is nonparametrically estimated in the first step. Similarly, the parametric specification of the equilibrium price distribution $F(p|N)$ in (1) can also be tested.

We present some Monte Carlo evidence in the online Appendix (supporting information) to demonstrate that our estimation procedure performs well in a controlled, small-sample environment. Moreover, we also show that failing to account for the unobservability of the potential number of firms can lead to biased estimates of model parameters.

5 Empirical Application

In this section, we apply our methodology to daily price data obtained from the UK price comparison site, Kelkoo, during the 18 September 2003 through 6 January 2004 period. Baye et al. (2009) use these data along with proprietary data on clicks to demonstrate that behavior at the site is consistent with a "clearinghouse" model that is nested as a special case of our formulation in Section 2. The goal of this section is to use the econometric procedure described above to structurally estimate parameters of the model, using only (publicly available) price data rather than both price and clicks data. Before proceeding, we briefly explain why the model in Section 2 more closely matches the shopping environment at Kelkoo than alternative models of online price dispersion, such as models with costly consumer search. Our discussion summarizes key points in Baye et al. (2009); the interested reader may refer to that paper for a more detailed description of the shopping environment at Kelkoo.

At the time these data were collected, Kelkoo was the largest price listing service in the world, operating in nine other European countries besides the UK. Within the UK, Kelkoo was the third largest retail website, with over 1,800 participating retailers including 18 of the largest 20 online retailers in the UK. It attracted an average of 10 million individual users per month, more than twice that of its closest rival. With a simple keyword search (e.g., "Compaq iPaq H3630"), a consumer visiting the site during 2003 obtained a complete list of the prices that different sellers listing on Kelkoo charged for the exact same product. Information about shipping and VAT were also displayed, such that it was virtually costless for a shopper to identify the seller charging the lowest "all in" price. By simply clicking the name of the firm offering the lowest price, the consumer was directed to the checkout page of the firm's website to purchase the product.

For these reasons, the Kelkoo environment of 2003 shares many of the key features of the idealized environment modeled in Section 2. Its dominant size, depth of participation, and ease of use meant there were few realistic alternative platforms for search by price conscious UK shoppers. UK branding activities at the time were no less sophisticated than those in the US; thus, it seems realistic to suppose that many consumers were loyal to one of the competing retail brands, which included such UK giants as Tesco and Dixon's, located on virtually every High Street in the country. Of course, a fringe of consumers no doubt pursued other search strategies or were loyal to no firm in particular, but such consumers were unlikely to be of a size so as to undo the main economic forces driving the model and, presumably, real world firm decisions by competitors in the consumer electronics retail market. In any event, the price to be paid for the ability to recover deep parameters via structural estimation is commitment to a model that, like all models, is necessarily an imperfect representation of the rich tapestry comprising the actual real world setting.

5.1 Data and Empirical Issues

The estimation dataset, which is fully described in Baye et al. (2009), includes 6151 daily listed prices (inclusive of taxes and shipping) charged by firms selling 18 models of PDAs over the period from 18 September 2003 through 6 January 2004. Given the theory model and shopping environment at Kelkoo, one empirical approach would be to produce separate estimates for each of these 18 different PDAs. Unfortunately, we have a limited number of observations for each product and therefore pool data across different PDAs. Fortunately, this problem is common in the literature (see Hong and Shum (2006), Moraga-González and Wildenbeest (2008), as well as Moraga-González et al. (2013)), and we may use standard techniques to overcome these data limitations. In particular, we utilize information about the characteristics of each product (e.g., brand, CPU clock, ram capacity, and whether bluetooth and built-in camera are available) to construct "homogenized prices" that control for product heterogeneity, where characteristics associated with each model were obtained by merging the original dataset with information from PDAdb.net, a mobile device database.

Table 1 provides summary statistics for our data. We can see that products with high average prices also have higher standard deviation. This implies products with high value are subject to a larger price dispersion, which may be caused by product characteristics. The table indicates that product characteristics, as well as manufacturer and month effects, contribute to the observed variation of listed prices over time and across different PDAs. This is more clearly seen in Model 1 of Table 2, which shows the results of a regression of prices on

the number of listings, characteristics of products, manufacturer and month dummies. As the results of Model 1 show, controlling for product heterogeneity and seasonal (month) effects, the number of listing firms continues to have a negative and statistically significant effect on price. On average, the listed price declines by 10.19 GBP with each additional competing firm. Second, product characteristics and the specific brand of PDA all influence price. For example, a built-in camera adds as much as 80 GBP to the price of a PDA. Third, the average listed price in December (the omitted month) is higher than in September, October and January at the 5% significance level.

To deal with product heterogeneities, we pool the data and construct homogenized prices by regressing price on the product characteristics and brand fixed effects (Model 2 in Table 2) and then construct "homogenized" prices using the regression results. These homogenized prices are the basis for our structural estimation.[6] Ideally, we may take into account the product heterogeneity in our analysis by two alternative ways. First, we condition the estimation on the heterogeneity. This requires a large sample size and is not plausible for our application due to the limitation of data. Second, we employ a structural method that allows the model parameters r, m, S and M to depend on the characteristics. The dependence may be due to firms' asymmetry in size, reputation, etc. However, such a generalization comes at the cost of extending the theoretical model in Section 1 to an asymmetric one, which requires a different characterization of equilibria as well as a quantitively different argument of identification. It is still an open question whether such a more sophisticated model could be identified using the kind of data scenario we consider here.

Homogenized prices control for differences in product characteristics or manufacturer, but they do not (and cannot) account for the seasonal effects observed in Model 1 of Table 2. A key difficulty stems from the multiple factors impacting price (which increased) during the holiday season (December). During this time period, there is undoubtedly a demand shock, though its composition between shoppers (S) and loyals (M) is unclear. As well, there may be a supply shock (larger N) as some firms enter the market to take advantage of increased demand. This is consistent with the data, where we observe an average number of listings of $4.08, 4.68, 5.86, 7.84$ and 6.54 during September, October, November, December and January, respectively. Since we only observe the number of listings A and the average

[6]Such an approach has been used in literature to control for observed heterogeneity, e.g., see Haile et al. (2006) and Bajari et al. (2014) among others. The main idea is to express prices as $p = W\beta + u$, where W is a vector of characteristics and $u \perp W$. Thus the CDF of prices $F_{p|W} = \Pr(W\beta + u \leq p) = F_u(p - W\beta)$. Homogenized prices $p - W\beta$ are then constructed using the residuals when we regress prices on manufacturer and product characteristics as in Model 2 of Table 2.

price, we cannot disentangle the impact of the shock on various channels (N, M, S, and so on) without detailed sales data. Therefore, as a robustness check, we separately estimate the structural model at the exclusion of December, when a (positive) demand shock is most likely. As we shall see, the parameter estimates are consistent with the estimates over all periods.

5.2 Structural Estimation

We now estimate the model parameters θ by applying the two-step estimation procedure to the homogenized data. Our estimation is based on clickthrough fees at Kelkoo in 2003, which were 20 pence per click or $c = 0.20$. Out of $1,591$ product-dates available, 1229 had two or more firms listed prices.

In constructing the $G_{A|N}$ matrix, we collapse observations where there are 10 or more listings into a single bin owing to a paucity of data. Such pooling may be justified theoretically (see Baye and Morgan (2009)) as well as on the practical grounds that the estimates in our simulation were relatively invariant to the cutoff used for pooling. Correspondingly, both A and N take on ten distinct values from $\{2, 3, ..., 10, 11+\}$. We use a discretized second listed price as the instrument Z, which also has the same support. Hence, $G_{A|N}$ is a 10×10 matrix for purposes of estimation, with the first 9 columns corresponding to $N = 2, ...10$ and the last bin corresponding to $N > 10$.

In estimating the matrix of conditional probabilities, $G_{A|N}$, two potential problems must be overcome. The first concerns the structure of the matrices comprising equation (7) , used in the first step of estimation. Since $A \leq N$, it then follows that the matrix comprising the right-hand side of this equation should be upper triangular (as there is no chance of more listings than competing firms). Consequently, so too should the left-hand side matrix, $H_{Ep,A,Z}B_{A,Z}^{-1}$, but this latter matrix is formed using the data and so may not be upper triangular. Fortunately, this is of no consequence since, performing the first step of the estimation without constraining the data in any way, while constraining the estimated matrix $\widehat{G}_{A|N}$ to be upper-triangular in the second step of estimation, does no harm to the asymptotic consistency or convergence of $\widehat{G}_{A|N}$. The reason, as argued in An et al. (2010), is that $H_{Ep,N,Z}B_{A,Z}^{-1}$ must be upper-triangular asymptotically.

The second key hurdle concerns the non-negativity of the elements of $\widehat{G}_{A|N}$. Our main procedure does not constrain these estimates to be non-negative and indeed, in some cases, negative values do arise. This too may be overcome by adopting an alternative estimation procedure where we impose non-negativity constraints to all elements of $\widehat{G}_{A|N}$ and choose

values to minimize the distance between the left and right-hand sides of equation (7) rather than solving it explicitly from the decomposition. It is routine to show that this alternative procedure retains the asymptotic consistency and convergence properties of the original estimation procedure. We opted to compute $\widehat{G}_{A|N}$ both with and without imposing constraints and derived similar estimates. The reported estimates derive from the constrained version of $\widehat{G}_{A|N}$.

5.3 Estimation Results and Discussions

Table 3 reports the results of the first-stage estimates of the matrix $G_{A|N}$. The element $\widehat{G}_{A=i|N=j}(j \geq i)$ corresponds to the estimated probability (in the data used) that there are A firms listing prices on the comparison site when the population of firms is N. According to the discussion on the full rankness of $B_{A,Z}$ in Section 3, this result also implies that there are sufficient variation of N across time.

The resulting parameter estimates, along with bootstrapped standard errors, are reported in Table 4, where columns (a) and (b) are obtained under different ways of pooling the larger values of A. For comparison, we also provide in column (c) the "Naïve" estimates which assume $N = A$ and ignore the potential unobservability of the number of potential firms. The monetary parameters (r, m and ϕ) are denominated in GBP. As the table reveals, all of the parameters are precisely estimated by our two-step procedure.

The parameter estimates in column (a) of Table 4 indicate that, on an average day in the UK during 2003, a total of $M = 25.84$ loyal consumers were interested in purchasing a PDA online, while $S = 12.75$ consumers were interested in purchasing online from the firm charging the lowest price. These estimates imply that about 33 percent of consumers in this online market are price-sensitive shoppers, while 67 percent are loyals. It is interesting to contrast our estimates with those of Brynjolfsson and Smith (2000), who find that around 13% of consumers in US e-retail markets were shoppers during that time period. Given the somewhat less-developed state of e-retail in the UK compared to the US in 2003, it is not altogether surprising to find that fewer UK customers had become "attached" to a particular online retailer. The estimated conversion rate, $\gamma = .08$, implies that a firm listing on Kelkoo.com had to receive, on average, about 12 clicks in order to generate one sale. At a cost of 20 pence per click, this translates into an average cost per sale of 2.4 GBP in addition to the fixed listing fee of $\phi = 4.82$ GBP. Finally, notice that the estimated *monopoly markup* for a PDA, $(r - m)/r$, is about 80 percent.[7]

[7]This is the monopoly markup of the "homogenized price". As shown in column (h) of Table 5, the profit

Consistent with our simulation (see the online Appendix), the results in column (c) also show that the naïve approach yields parameter estimates that substantially understate the level of e-retail competition, compared to our estimates that account for the unobservability of N. This is most clearly seen from columns (g) and (h) in Table 5, which report the expected profit for each firm using the estimates in Columns (a) and (c) of Table 4 and the formula in Proposition 1. For each N, the expected profit using the naïve approach almost doubles those based on our estimates.

We also present several robustness checks to address concerns that our results may be sensitive to accounting for seasonal demand shocks, pooling data for large N, or employing different instruments and methods of discretization. These checks reveal that our findings are robust to these potential concerns. The results in column (f) of Table 4 deal with seasonality issues by using data at the exclusion of December 2003. If interaction effects from time varying demand were distorting our results, we would expect wildly varying estimates compared to those reported in column (a) based on all the data. Comparing the two sets of results we see, as expected, an increase in the number of shoppers and loyals, but little change in the other parameters, thus mitigating concerns about results being driven by a failure to account for demand shocks. Similarly, columns (d) and (e) of Table 4 permit one to compare results based on pooling observations with 11 or more listings (column (a)) with those arising by pooling observations with 8 or more listings. The estimates do not vary substantially with the method of pooling. Finally, we varied the construction of Z as well. Specifically, we chose alternative observed prices as the instrument Z and also used different methods of discretization. These variations did not materially impact our parameter estimates. Indeed, provided the rank condition $Rank(B_{A,Z}) = K - 1$ holds, the estimates are similar to those in columns (d) and (e) of Table 4, and quite robust to variation in discretization. Overall, these estimates suggest that the potential concerns we highlighted earlier do not materially impact our parameter estimates.

We conclude this section by examining the goodness-of-fit of our parametric specification of the price distribution $F(p|N)$ in Proposition 1.[8] In particular, we compare the non-parametrically estimated $\widehat{F}(p|N)$ in the first step with the one obtained by plugging $\hat{\theta}$ into equation (1), i.e, $F(p|N;\hat{\theta})$. For all $N \in \{2, 3, ..., 10, 10+\}$, the parametric distribution of

of non-monopoly retailers is much smaller.

[8] An alternative goodness-of-fit test for specification involves the probability $g(A|N)$; one may use the Chi-Square diagnostic tests proposed in Andrews (1988a,b) to test the difference between the nonparametrically estimated $\widehat{g}(A|N)$ and the parametric one $g(A|N;\hat{\theta})$. Nevertheless, we focus on $F(p|N)$ since it incorporates the parametric specification of the listing probability, α^*.

price, $F(p|N;\hat{\theta})$ is contained in the 95% point-wise confidence band (estimated by bootstrapping 200 times) of $\widehat{F}(p|N)$. Figure 1 illustrates the result for the case where $N = 8$, and is typical of the patterns for other N. In the figure, the solid and dotted lines represent $F(p|N;\hat{\theta})$ and the mean of $\widehat{F}(p|N)$, respectively, and the confidence band consists of two dash lines. An important caveat to this comparison is that, due to the small sample size, the confidence band around $\widehat{F}(p|N)$ is fairly wide, so having $F(p|N;\hat{\theta})$ lie inside this band is not a stringent test.

5.4 The Effects of Changes in Market Structure

The econometric framework described above, along with the structural estimates of the model of online price competition, permit us to address a number of issues that arise in the evaluation of the competitive effects when the number of online firms changes.

To accomplish this, we first substitute the parameter estimates reported in column (a) of Table 4 into the expressions summarizing equilibrium behavior in Proposition 1. We use carets to denote the resulting estimates. Next, we calculate the implied average prices conditional on a given number of firms and display them in Table 5. Column (a) in Table 5 lists the total number of firms in the relevant market (N), which is unknown. Column (b) provides the estimated average price listed at the comparison site conditional on different numbers of competitors, where the average listed price is $E[p] = \int_{\widehat{p}_0}^{\widehat{r}} p d\widehat{F^*}(p)$. As expected, Table 5 shows that the estimated average listed price declines in the number of firms—rather abruptly as one moves from monopoly to a duopoly, and modestly thereafter. Column (c) reports the estimated average minimum listed price, which is given by $E[p_{\min}] = \frac{1}{1-(1-\widehat{\alpha^*})^N} \sum_{A=1}^{N} \binom{N}{A} \widehat{\alpha^*}^A (1-\widehat{\alpha^*})^{N-A} \int_{\widehat{p}_0}^{\widehat{r}} pA[1-\widehat{F^*}(p)]^{A-1} d\widehat{F^*}(p)$. Notice that this calculation takes into account the effect of a change in N on the equilibrium distribution of prices, firms' propensities to advertise prices at the comparison site, and the impact of a larger number of listings on the minimum order statistic. Accounting for this, Column (c) of Table 5 shows that the estimated average minimum listed price also declines as the number of firms increases.

Note that in our setting neither the average prices nor the average minimum prices represent average transaction prices. To calculate the average transaction price paid by loyals, one needs to account for a firm's propensity to list prices on the comparison site. When a firm chooses not to list, the model implies that it charges the monopoly price, but this price is unobservable to us. Thus, the average transaction price paid by a loyal customer is $E[p^L] = \widehat{\alpha^*} E[p] + (1-\widehat{\alpha^*})\widehat{r}$. Column (d) of Table 5 reports the estimated average transaction

prices of loyal consumers. Notice that it declines abruptly as one moves from monopoly to duopoly, but then rises as the number of firms increases further.[9] Likewise, the average transaction price for shoppers must also account for listing decisions: The average transaction price paid by a price-sensitive shopper is given by $E\left[p^S\right] = [1-(1-\widehat{\alpha^*})^N]E\left[p_{\min}\right]+(1-\widehat{\alpha^*})^N\widehat{r}$. Column (e) of Table 5 reports the estimated average transaction price of shoppers, which declines as the number of firms increases.[10]

Columns (d) and (e) highlight that shoppers and loyals are impacted differently by heightened competition: So long as there are at least two firms in the market, loyal consumers are harmed by heightened competition, while shoppers are unambiguously made better off. The overall transaction price, reported in Column (f) of Table 5, is merely an average of the shoppers' and loyals' estimated transaction prices, weighted by the estimated fraction of are shoppers vs. loyals: $E\left[p^T\right] = \frac{\widehat{M}}{\widehat{S}+\widehat{M}}E\left[p^L\right] + \frac{\widehat{S}}{\widehat{S}+\widehat{M}}E\left[p^S\right]$. The first row of columns (b)-(f) presents the price charged by a monopolist, $(r-m)\gamma(M+S)$.

In summary, the estimates in Table 5 reveal that the average listed price and the average minimum listed price both decline as the number of firms increases. This is consistent with standard reasoning, which suggests that heightened competition leads to lower prices. However, this ignores the endogenous listing decisions of firms, which is, of course, relevant for the *transaction* prices paid by consumers. Here, a more subtle story emerges. Both shoppers and loyals pay lower average transaction prices as the online market moves from monopoly to duopoly. Thereafter, the effects of increased competition diverge: Loyal consumers are harmed (pay higher average transaction prices) as the number of firms further increases, while shoppers benefit from heightened competition. Intuitively, as the number of firms increases, the minimum listed price decreases due to competition, so does the profit of a firm earns from the comparison website. This results in a smaller listing probability for a

[9]That the transactions price increasing for loyals as competition increases is not purely an artifact of our two consumer type model. In fact, the result readily extends to situations where some of the loyals for each firm are only partially loyal ("quasi-loyals") in the sense that they also search at a second firm. To see this, let ε denote the fraction of "quasi-loyals" in the population. If the proportion of loyals $1-\varepsilon$ is sufficiently large, the effects of competition on loyals remains under this alternative model.

[10]There is an apparent discrepancy between the predictions as to average transaction prices on the comparison site, which are falling in N (and hence also in A because the two are positively correlated) and our reported average transaction prices offered in Table 1, which are not strictly monotone in A. The difference stems from the fact that the measure contained in the summary statistics does not account for instances where shoppers buy without using the comparison site because of the lack of listings. These events are more likely when N (and hence A) is small than when it is large; thus, the summary statistics *understate* actual transactions prices, though to a lessening degree as A increases.

larger number of firms (using the estimate of column (a) in Table 4, we can verify that α is decreasing in N). Thus for a larger N, loyal consumers pay the monopoly price r and the average listed price $E[p]$ with a larger and smaller probability, respectively, and this may lead to an increasing average transaction price. Similarly, shoppers also pay r with an increasing probability $(1-\alpha^*)^N$ and the average minimum listed price $E[p_{min}]$ with a decreasing probability for a larger N. However, since the probability $(1-\alpha^*)^N$ is much smaller than $[1-(1-\alpha^*)^N]$, the decreasing $E[p_{min}]$ dominates shoppers' payment as N increases. Thus shoppers benefit from heightened competition by paying less.

Table 6 uses the results in Table 5 to compute the price effects when the number of firms declines from N to $N-1$, where column (a) represents the post-decline number of firms. So long as there is more than one firm in the market, a change of market structure will not harm the "average" online consumer. This conclusion is based on the assumption that firms in the online channel do not compete against firms in other channels. In effect, column (f) reveals that—even though models of online competition are more complex than standard homogenous product Bertrand competition and the "law of one price" does not hold online—the conclusions based on our estimates are similar to what one would have concluded based on the simple Bertrand model, at least in this particular online market: There are no adverse effects from a consolidation in market structure in this online market so long as there are two firms in the market.

It is interesting to compare our results above with a Bertrand competition model without segmentation where all the $M+S$ consumers are shoppers (if all the consumers are loyal, each firm charge r on its own site). In such a model, a monopoly firm ($N=1$) charges r on its own website. Once $N>1$, a firm without advertising its product on the comparison website attracts no shoppers and earns zero profit. On the other hand, due to Bertrand competition, firms list on the comparison site charge a uniform price, which is the sum of marginal cost, listing and click cost. Therefore, similar to our model, if there is more than one firm in the market, a change of market structure will benefit the consumers.

We may also compute the competitive effects using the naïve estimates in Table 4. The results show that one may underestimate competitive effects by failing to account for the unobserved number of potential firms in the market. For example, if the number of firms decreases from three to two, our estimates indicate that the average transaction price paid by price sensitive "shoppers" and "loyals" increases by 6.89%, and declines by 3.19%, respectively. Using the naïve estimates results in smaller predicted price effects (12% smaller for "shoppers" and 15% smaller for "loyals").

6 Conclusions

We showed that the econometric methodology from winner-pay auctions may be used to identify and structurally estimate standard models of online competition, even when the number of competing firms is unobserved. The estimates can be employed to analyze the competitive effects induced by the change of number of firms. Our empirical results suggest that: (1) Online markets are less vulnerable to adverse competitive effects from reductions in the number of firms than one might expect given the plethora of papers documenting significant price dispersion in online markets; (2) reductions in the number of competitors in online retail markets harm price sensitive shoppers but benefit customers who are loyal to a particular firm; and (3) using the observed number of firms as a proxy for the number of potential firms may lead to estimates that significantly understate the degree of competition in online markets. We stress, however, that these findings are based on data from one e-retail market in the UK and a particular structural model that we believe fits that market environment. While the model and econometric techniques developed in this paper are useful more generally, one should tread cautiously in generalizing the results from our study to environments where search costs play an important role in determining equilibrium prices.

References

An Y, Hu Y, Shum M. 2010. Estimating First-Price Auctions with an Unknown Number of Bidders: A Misclassification Approach. *Journal of Econometrics* **157**: 328–341.

Andrews DW. 1988a. Chi-square diagnostic tests for econometric models: Introduction and applications. *Journal of Econometrics* **37**: 135–156.

Andrews DW. 1988b. Chi-square diagnostic tests for econometric models: theory. *Econometrica* : 1419–1453.

Armstrong M. 2008. Interactions between competition and consumer policy. *Competition Policy International* **4**: 97–147.

Bajari P, Houghton S, Tadelis S. 2014. Bidding for incomplete contracts: An empirical analysis of adaptation costs. *American Economic Review* **104**: 1288–1319.

Baye MR. 2008. Market definition and unilateral competitive effects in online retail markets. *Journal of Competition Law and Economics* **4**: 639–653.

Baye MR, Gatti JRJ, Kattuman P, Morgan J. 2009. Clicks, discontinuities, and firm demand online. *Journal of Economics & Management Strategy* **18**: 935–975.

Baye MR, Morgan J. 2001. Information gatekeepers on the internet and the competitiveness of homogeneous product markets. *American Economic Review* : 454–474.

Baye MR, Morgan J. 2009. Brand and price advertising in online markets. *Management Science* **55**: 1139–1151.

Baye MR, Morgan J, Scholten P. 2006. Information, search, and price dispersion. *Handbook on economics and information systems* **1**.

Baye MR, Morgan J, Scholten PA. 2004. Price dispersion in the small and in the large: Evidence from an internet price comparison site. *Journal of Industrial Economics* **52**: 463–496.

Brynjolfsson E, Smith MD. 2000. The great equalizer? the role of shopbots in electronic markets. Technical report, Working Paper, MIT Sloan School of Management, Cambridge, MA.

Haile P, Hong H, Shum M. 2006. Nonparametric Tests for Common Values in First-Price Auctions. NBER working paper #10105.

Hong H, Shum M. 2006. Using Price Distributions to Estimate Search Costs. *RAND Journal of Economics* **37**: 257–275.

Hu Y. 2008. Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. *Journal of Econometrics* **144**: 27–61.

Iyer G, Pazgal A. 2003. Internet shopping agents: Virtual co-location and competition. *Marketing Science* **22**: 85–106.

Iyer G, Soberman D, Villas-Boas JM. 2005. The targeting of advertising. *Marketing Science* **24**: 461–476.

Moraga-González JL, Sándor Z, Wildenbeest MR. 2013. Semi-nonparametric estimation of consumer search costs. *Journal of Applied Econometrics* **28**: 1205–1223.

Moraga-González JL, Wildenbeest MR. 2008. Maximum likelihood estimation of search costs. *European Economic Review* **52**: 820–848.

Narasimhan C. 1988. Competitive promotional strategies. *Journal of Business* : 427–449.

Rosenthal RW. 1980. A model in which an increase in the number of sellers leads to a higher price. *Econometrica: Journal of the Econometric Society* : 1575–1579.

Varian HR. 1980. A model of sales. *The American Economic Review* : 651–659.

A Identification of $\Pr(N)$

In this section, we discuss the identification of $\Pr(N)$, the probability distribution of number of firms, N. We first construct a linear system $f(p_j) = \sum_{N=2}^{K} f(p_j|N) \Pr(N)$, where $f(p)$ is the density of listed prices including $A = 1$. $p_j, j = 1, 2, \cdots, K-1$ are $(K-1)$ prices and $f(p_j|N)$ is identified in Proposition 2. For ease of exposition, we rewrite the linear system as a matrix equation

$$\Phi_{p,N} \mathbf{\Sigma}_N = \mathbf{P}, \tag{10}$$

where the matrix $\Phi_{p,N}$ is defined as $(\Phi_{p,N})_{j,k} = f(p = p_j|N = k)$. $\mathbf{\Sigma}_N$ denote the vector of the unknown frequency distribution of N, $\mathbf{\Sigma}_N \equiv \big(\Pr(N=2), \Pr(N=3), \cdots, \Pr(N=K)\big)^T$ and $\mathbf{P}$ is the vector of $K-1$ prices, $\mathbf{P} \equiv \big(f(p_1), f(p_2), \cdots, f(p_{K-1})\big)^T$. The vector $\mathbf{\Sigma}_N$ is uniquely determined by the linear system if the following full rank condition holds:

Condition 2 *There exist $K-1$ prices, $p_1, p_2, \cdots p_{K-1}$ such that* $\text{Rank}\,(\Phi_{p,N}) = K-1$.

Note that Condition 2 is testable because $f(p|N)$ is identified in Proposition 2 and the matrix $\Phi_{p,N}$ is observable once the choice of the $K-1$ prices is given. Under Condition 2, all the elements of the vector $\mathbf{\Sigma}_N$ are uniquely solved from the linear system using Cramer's rule. We summarize the results above in the following corollary of Proposition 2.

Corollary 1 *Suppose Conditions 1-2 hold. The probability distribution of N is identified for $N = 2, 3, \cdots, K$.*

B Estimation: The First Step

In this section, we describe how to use observable data on prices (p) and the number of listing firms (A) to estimate $G_{A|N}$ and the probability distribution $\mathbf{\Gamma}_N$ using the data with $A \geq 2$. We suppress the condition $A \geq 2$ whenever there is no ambiguity. Our methodology

closely parallels the approach taken in An, Hu and Shum (2010). While the key identification equation (3) is stated in terms of the joint density $h(p, A, Z)$, faster convergence is achieved if instead we take the expectation over all prices given (A, Z). Specifically, let $E[p|A, Z] = \int p \frac{h(p,A,Z)}{b(A,Z)} dp$, i.e. the expected price conditional on some realization A, Z. It then follows from equation (7) that

$$E[p|A, Z]\, b(A, Z) = \sum_{N=2}^{K} E[p|N] \times g(A|N)\psi(N, Z)$$

where $E[p|N] = \int p f(p|N)\, dp$.

Now define the matrices:

$$H_{Ep,N,Z} \equiv \left[E(p|A = i, Z = j)\, b(A = i, Z = j)\right]_{i,j}, \tag{11}$$

and

$$F_{Ep|N} \equiv diag\left(E[p|N = 2], \cdots, E[p|N = K]\right).$$

Then, we have

$$H_{Ep,A,Z} = G_{A|N} F_{Ep|N} \Psi_{N,Z}$$

which is analogous to equation (5). Similarly, we can obtain the estimating equation by postmultiplying both sides of this equation by $B_{A,Z}^{-1}$. This yields the analogous identification equation:

$$H_{Ep,A,Z} \left(B_{A,Z}\right)^{-1} = G_{A|N} F_{Ep|N} \left(G_{A|N}\right)^{-1} \tag{12}$$

Consequently,

$$G_{A|N} = \zeta\left(H_{Ep,A,Z} \left(B_{A,Z}\right)^{-1}\right),$$

where $\zeta(\cdot)$ denotes the mapping from a square matrix to its eigenvector matrix. Note that if the distribution of listed prices is such that the average price is monotonically ordered in N, then an analog of Lemma 3 holds for expected prices as well. This guarantees that ζ is a unique mapping. Following Hu (2008), we may estimate the relevant matrices using sample averages:

$$\widehat{G}_{A|N} \equiv \zeta\left(\widehat{H}_{Ep,A,Z} \left(\widehat{B}_{A,Z}\right)^{-1}\right), \tag{13}$$

where

$$\widehat{H}_{Ep,A,Z} = \left[\frac{1}{T} \sum_{t} \frac{1}{A_j} \sum_{i=1}^{A_j} p_{it} \mathbf{1}(A_t = A_j, Z_t = Z_k)\right]_{j,k}. \tag{14}$$

In general, Z can be obtained by different choices of discretization, and this may result in different matrices $H_{p,A,Z}$, and $\Psi_{N,Z}$. However, both the main identification equation (8) and

Proposition 2 hold for those methods of discretization such that Condition 1 holds. Our procedure of identification and estimation are based on the main identification equation (8), which is pointwise in p, a continuous variable. Therefore, we are still dealing with continuous price and discretization will have no impact on our estimation results.

Finally, let $\mathbf{g}(A)$ be a vector of marginal probabilities over the number of listings and let $\mathbf{\Gamma}_N$ denote the vector of the unknown frequency distribution of N conditional on $A \geq 2$. Then

$$\mathbf{g}(A) = G_{A|N}\mathbf{\Gamma}_N$$

and we may estimate the unknown distribution $\mathbf{\Gamma}_N$ using the data as follows:

$$\widehat{\mathbf{\Gamma}}_N = \left(\widehat{G}_{A|N}\right)^{-1} \hat{\mathbf{g}}(A) \tag{15}$$

where $\hat{\mathbf{g}}(A)$ denotes the empirical frequency of the number of listings.

Finally, we propose a least square estimate for $\mathbf{\Sigma}_N$ according to the linear system equation (10) on Corollary 1. Without taking average with respect to price p, the eigen-decomposition in the preceding step can be used to estimate the conditional density $f(p|N)$. Let $\hat{f}(p|N)$ denote the estimate, then we have a linear system $\hat{f}(p_{it}) = \sum_{N=2}^{K} \hat{f}(p_{it}|N) \Pr(N)$, where $\hat{f}(\cdot)$ is a kernel density estimate. The probabilities $\Pr(N)$ then are estimated using a least square estimator. It is worth noting that we use all the observation including those $A = 1$ in this step of estimation.

Table 1: Summary Statistics

Variable		# of observations	Mean	Std. Dev.	Min	Max
# of listings	1	362	323.10	121.64	141.56	607.77
	2	714	315.21	142.51	108.1	607.77
	3	465	303.24	126.43	108.1	607.77
	4	644	300.12	94.50	133.77	607.77
	5	705	287.30	91.65	132.75	499
	6	660	320.87	92.22	133.77	509.29
	7	791	315.87	90.16	183.94	509.29
	8	608	319.98	104.19	179.99	509.29
	9	315	306.89	103.63	179.99	509.29
	10	350	327.41	115.16	179.99	509.29
	11	297	307.68	106.19	179.99	509.29
	12	132	256.49	35.26	183.94	329.71
	13	65	271.48	26.75	244.95	329.71
	14	28	272.29	26.22	244.95	329.71
	15	15	272.83	24.75	244.95	323.89
Month	Sep.2003	655	303.75	103.42	108.1	529.94
	Oct.2003	1781	305.62	108.23	108.1	599
	Nov.2003	1764	314.29	109.28	110.45	607.77
	Dec.2003	1674	309.48	104.91	134.88	596.98
	Jan.2004	277	307.43	104.62	141.56	596.98
Product*	e740wi (Toshiba)	216	434.12	90.11	251.46	491.57
	h1910 (HP)	171	223.83	30.54	169.49	309.98
	h1940 (HP)	898	275.45	23.13	240.88	331.69
	h2210 (HP)	184	332.26	26.22	295.98	374.83
	h3950 (HP)	91	291.06	21.18	273.77	377.11
	h3970 (HP)	131	328.30	25.94	298.74	415.07
	h5550 (HP)	851	464.96	27.14	427.98	509.29
	m515 (Palm)	44	198.85	19.79	166.98	219.00
	nx70v (Sony)	164	291.22	61.02	234.42	379.00
	nx73v (Sony)	501	379.66	28.58	338.94	446.45
	nz90 (Sony)	151	541.47	34.57	499.95	607.77
	sj22 (Sony)	368	151.62	16.02	132.75	183.94
	sj33 (Sony)	44	173.89	5.85	168.29	179.95
	tg50 (Sony)	428	272.55	20.09	203.94	309.54
	treo90 (Handspring)	136	132.02	22.56	108.10	156.17
	tungstent2 (Palm)	678	265.82	30.14	202.39	327.03
	tungstenw (Palm)	295	406.17	42.08	376.98	499.00
	zire71 (Palm)	800	210.73	15.76	179.99	253.94
Total		6,151	309.04	107.01	108.1	607.77
# of listings		6,151	5.90	3.07	1	15
Characteristics	CPU speed (MHz)	6,151	220.23	120.47	33	400
	RAM (MB)	6,151	46.45	38.34	16	128
	Bluetooth	6,151	.55	.49	0	1
	Built-in Camera	6,151	.31	.46	0	1

* A product is indicated by both its model and the brand (in parentheses), e.g., e740wi is produced by Toshiba.

Table 2: Reduced form analysis

	Model 1	Model 2
Number of listings	-10.19***	
	(0.40)	
Processor (MHZ)	0.17***	0.29***
	(0.023)	(0.023)
Ram (GB)	2.57***	2.30***
	(0.061)	(0.062)
Bluetooth	55.18***	28.70***
	(3.68)	(3.70)
Camera	78.86***	50.17***
	(4.08)	(4.10)
HP	-102.47***	-116.95***
	(4.41)	(4.60)
Sony	24.58***	41.76***
	(2.67)	(2.72)
Toshiba& Handspring	-61.68***	-34.74***
	(5.13)	(5.28)
September	-27.53***	
	(3.50)	
October	-15.0***	
	(2.63)	
November	-3.98	
	(2.48)	
January	-11.50**	
	(4.56)	
Constant	201.5***	142.88***
	(3.98)	(2.71)
N	6151	6151
R^2	0.575	0.528
adj. R^2	0.574	0.528

Standard errors in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: Estimated $G_{A|N}$ Matrix

	Number of firms (N)									
Number of listings(A)	2	3	4	5	6	7	8	9	10	> 10
2	1.00	0.56	0.66	0.40	0.31	0.16	0.42	0.63	0.49	0.35
3	0	0.44	0.20	0.10	0.15	0.18	0.04	0.09	0.15	0.13
4	0	0	0.14	0.22	0.15	0.03	0.17	0.14	0.14	0.08
5	0	0	0	0.28	0.28	0.53	0.18	0.03	0.02	0.09
6	0	0	0	0	0.11	0.04	0.03	0.02	0.06	0.16
7	0	0	0	0	0	0.06	0.11	0.03	0.06	0.09
8	0	0	0	0	0	0	0.03	0.02	0.05	0.01
9	0	0	0	0	0	0	0	0.04	0.02	0.05
10	0	0	0	0	0	0	0	0	0.01	0.02
> 10	0	0	0	0	0	0	0	0	0	0.02

Table 4: Parameter Estimates and Robustness Check

	Estimate			Robustness Check		
Params.	Two-step estimate (a)	Two-step estimate (b)	"Naïve" estimate (c)	Discretization # 1 (d)	Discretization # 2 (e)	Exclude December (f)
ϕ	4.82	4.77	3.02	4.91	4.92	4.01
	(0.45)	(0.30)	(0.22)	(0.68)	(0.81)	(0.78)
r	452.75	452.76	319.49	466.18	445.37	426.78
	(32.87)	(25.12)	(23.43)	(56.21)	(76.34)	(65.34)
m	86.34	82.46	94.29	83.47	80.27	82.12
	(25.08)	(14.39)	(17.09)	(26.55)	(28.19)	(34.12)
M	25.84	23.82	35.36	27.81	25.20	29.13
	(9.75)	(14.10)	(12.92)	(8.10)	(11.12)	(8.66)
S	12.75	11.88	22.59	9.56	10.09	14.32
	(1.91)	(1.82)	(2.02)	(1.23)	(2.65)	(2.42)
γ	0.076	0.069	0.14	0.065	0.069	0.077
	(0.013)	(0.0084)	(0.01)	(0.03)	(0.02)	(0.03)

Note: Column (d) contains the estimates when we pool observations with 8 or more listings, and column (e) reports the results when we choose different price to be discretized to get Z. Specifically, the cutoff points are the 9th, 19th, 30th, 38th, 46th, 60th, 70th, 80th, and 95th percentile of the chosen price.

Figure 1: Price Distributions

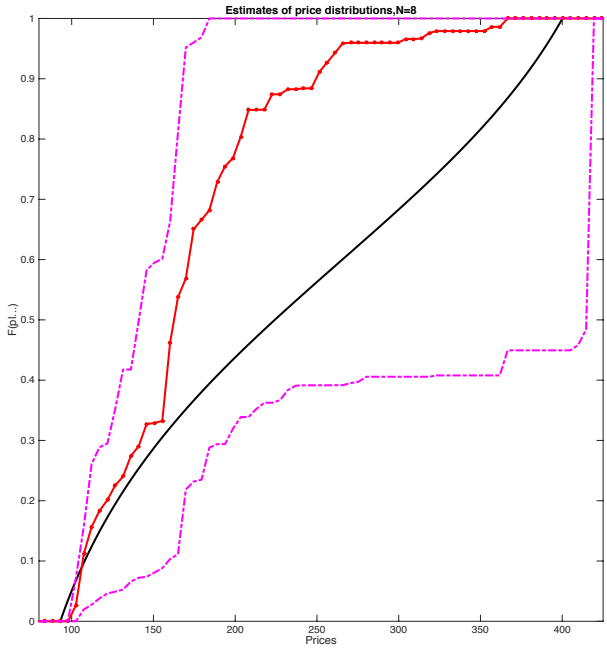

Table 5: Estimated Transaction Prices and Profit

Number of Firms	Prices					Profit	
	Avg. Listed Price	Avg. Minimum Listed Price	Avg. Trans. Price Loyals	Avg. Trans. Price Shoppers	Avg. Trans. Price	"Naïve" Profit	Two-step Profit
(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
1	452.75	452.75	452.75	452.75	452.75	1827.05	1074.62
2	345.85	319.31	348.80	319.41	339.09	560.47	364.68
3	344.38	296.95	359.93	297.41	339.27	373.13	242.29
4	339.16	279.04	369.08	279.87	339.61	279.72	181.51
5	333.22	264.55	376.46	265.72	339.87	223.73	145.13
6	327.39	252.61	382.51	254.06	340.07	186.41	120.90
7	321.89	242.58	387.56	244.27	340.22	159.77	103.61
8	316.79	234.02	391.86	235.91	340.33	139.79	90.64
9	312.06	226.62	395.56	228.69	340.43	124.25	80.56
10	307.69	220.15	398.79	222.37	340.50	111.82	72.50
11	303.63	214.43	401.64	216.79	340.57	101.65	65.90
12	299.85	209.33	404.18	211.81	340.62	93.18	60.41
13	296.33	204.76	406.45	207.35	340.67	86.01	55.76
14	293.04	200.63	408.50	203.32	340.71	79.86	51.77
15	289.95	196.88	410.36	199.66	340.74	74.54	48.32

Table 6: Percentage Change of Transaction Prices

Number of Firms	Estimated Change in Average Listed Price	Estimated Change in Avg. Minimum Listed Price	Estimated Change in Average Transaction Price Loyals	Estimated Change in Average Transaction Price Shoppers	Estimated Change in Average Transaction Price
(a)	(b)	(c)	(d)	(e)	(f)
1	23.61%	29.47%	22.96%	29.45%	25.10%
2	0.43%	7.00%	-3.19%	6.89%	-0.05%
3	1.51%	6.03%	-2.54%	5.90%	-0.10%
4	1.75%	5.19%	-2.00%	5.06%	-0.08%
5	1.75%	4.51%	-1.61%	4.39%	-0.06%
6	1.68%	3.97%	-1.32%	3.85%	-0.04%
7	1.59%	3.53%	-1.11%	3.42%	-0.03%
8	1.49%	3.16%	-0.95%	3.06%	-0.03%
9	1.40%	2.86%	-0.82%	2.76%	-0.02%
10	1.32%	2.60%	-0.71%	2.51%	-0.02%
11	1.24%	2.38%	-0.63%	2.29%	-0.02%
12	1.17%	2.18%	-0.56%	2.11%	-0.01%
13	1.11%	2.02%	-0.50%	1.94%	-0.01%
14	1.05%	1.87%	-0.46%	1.80%	-0.01%