Estimating Firm-Level Demand at a Price Comparison Site: Accounting for Shoppers and the Number of Competitors

Michael R. Baye   J. Rupert J. Gatti   Paul Kattuman

John Morgan

September 2005

Abstract

Clearinghouse models of online pricing—such as Varian (1980), Rosenthal (1980), Narasimhan (1988), and Baye-Morgan (2001)—view a price comparison site as an "information clearinghouse" where price-sensitive shoppers obtain price and product information to make online purchases. These models predict a discontinuous jump in a firm's demand when it succeeds in charging the lowest price in the market. These models also predict that the responsiveness of a firm's demand to a change in its price depends on the number of rivals. Using a unique firm-level dataset from Kelkoo.com (Yahoo!'s European price comparison site), we examine these predictions by providing estimates of the demand for PDAs. Our results indicate that both the number of competing sellers and the firm's rank in the list of prices are important determinants of an online retailer's demand. We find that an online monopolist faces an elasticity of demand of about -2, while sellers competing against 10 other sellers face an elasticity of about -6. We also find empirical evidence of a 60% jump in a firm's demand as its price declines from the second-lowest to the lowest price. Our estimates suggest that about 13% of the consumers at Kelkoo are "shoppers" who purchase from the seller offering the lowest price.
JEL Classification Numbers: D4, D8, M3, L14

1 Introduction

Estimating a firm's demand in the online marketplace—particularly at price comparison sites such as Shopper.com and Kelkoo.com—fundamentally differs from demand at most physical marketplaces. One of the key differences—first noted by Baye and Morgan (2001) but certainly anticipated by the early works of Varian (1980), Rosenthal (1980), Shilony (1977), and Narasimhan (1988)—stems from the fact that consumers in online markets typically obtain a complete list of prices charged by different sellers before making their purchase decisions. As a consequence, these models predict that a firm enjoys a discontinuous jump in demand when it succeeds in charging the lowest price because it instantly attracts the price-sensitive "shopper" segment of the market. Moreover, unlike traditional retail markets where firms like Wal Mart compete and there is little turnover in the identity of the firm charging the lowest price, the identity of the low-price seller frequently changes in online markets (Baye, Morgan, and Scholten, 2004b; Ellison and Ellison, 2004).

The online marketplace also differs from its physical cousin in its rapid turnover in the number of competing sellers. In conventional retail markets, the firms competing for customers in (say) Walnut Creek, California change infrequently owing to the barriers to entry and exit associated with setting up a physical retail location. In the online world, change comes faster. This is particularly true in the marketplace defined by a price comparison site. Here, the number of firms listing prices for a given product changes almost daily (see Baye, Morgan, and Scholten 2004a). Indeed, as pointed out by Baye and Morgan (2001), this variation in the degree of rivalry of a given online market is essential for firms to avoid pure Bertrand competition and for the information "gatekeeper"—the entity running the price comparison site—to profitably operate. Thus, in online markets, it is especially important to account for changes in the degree of rivalry; standard oligopoly models predict that the greater the number of rivals, the more elastic is a firm's demand.

A third way online markets differ from conventional markets is in the changing "locations" of firms. In conventional retail markets, the physical real estate a retailer occupies changes infrequently, and the "value-added" by its physical location is difficult to disentangle from its other characteristics. In contrast, "virtual" real estate in the online world changes rapidly.

For instance, in purchasing "adwords" (advertising space at the side of search queries on Google's site), retailers realize the advantage conferred to being the first listing on the page—and bid aggressively to obtain such a position. At any moment, a retailer can find itself displaced from this "prime" real estate to a less favorable screen location.

The rapidly changing nature of the online marketplace—the numbers of competing firms, identities of the low-price firms, and firms' screen locations—presents both a challenge and an opportunity in estimating firm-level demand in online markets. The challenge is that if one fails to properly account for these factors, one obtains biased parameter estimates. The opportunities stem from the dynamic nature of the data. Variation in the identity of the low priced firm enables us to disentangle the demand jump from other determinants of demand. Variation in the number of competitors permits us to identify the marginal impact of the number of rivals on a firm's demand elasticity (and hence its markup). Finally, variation in screen locations allows us to identify of the value of the virtual real estate separately from other firm characteristics.

Before summarizing how we seek to overcome these challenges and take advantage of the opportunities described above, it is important to point out a final challenge typically faced by researchers estimating demand in online markets—the absence of actual data on sales. The ideal dataset would consist of matched observations on the set of alternatives presented to consumers using a price comparison site, consumer clickthroughs, and final purchase decisions. Unfortunately, the principal-agent problem between the owners of price-comparison sites and e-retailers means that matched data generally do not exist. While each firm has private information about its own sales, it lacks the information enjoyed by the price comparison site. Price comparison sites, on the other hand, lack information on firms' sales, but have a wealth of other information. This includes detailed information about the prices, clickthroughs and characteristics of all sellers "inside" its market, as well as information about consumer pageviews and search patterns. While in principle the parties could "solve" this agency problem by integrating their information technologies across the supply chain, the norm is for price comparison sites to charge clickthrough-based fees. As a consequence, researchers are left with the choice between attempting to gain proprietary *sales* data from an e-retailer or attempting to obtain proprietary *clicks* data from a price comparison site.

Most researchers have opted for clickthrough data; we know of only one study (Ellison and Ellison, 2004) that is based on actual sales data—from a single firm in the online market for computer memory.

This paper addresses these challenges and opportunities. It demonstrates that theoretical "clearinghouse" models of online competition can be used, in conjunction with existing pseudo-maximum likelihood techniques specifically designed for count data, to obtain consistent estimates of consumer clickthrough behavior. Further, Proposition 1 shows that it is possible to recover underlying demand parameters (including price elasticities) from these estimates. We apply these techniques to a unique UK dataset for 18 personal digital assistants (PDAs) obtained from the Yahoo! European price comparison site, Kelkoo.com. Consistent with what one might conjecture based on the challenges we identified above, we find evidence that it is indeed important to account for the influence of shoppers, the number of rivals, and other determinants of demand (such as screen location) in estimating demand in online markets.

We find that a firm offering the best price enjoys a 60 percent increase in demand compared to what it would have enjoyed had it not charged the lowest price. Failing to account for this jump by modeling demand as continuous leads to elasticity estimates that are about *twice* as large. Our results also reveal that a firm's elasticity of demand (apart from this jump) is more elastic in online markets where competition is keener. We find that a monopoly seller faces an elasticity of demand of about -2.5, while in the most competitive markets we analyze (15 sellers), the elasticity of demand for a representative firm's product is about -6.0. We are also able to identify the effect of other determinants of demand—such as screen location—on firm demand. Our results imply that, other things equal, a firm loses about 15% of its business for every competitor listed above it on the screen.

Our results are related to a variety of papers in the literature. As noted earlier, Ellison & Ellison (2004) use sales data for computer memory chips obtained from a single firm listing on Pricewatch.com. The emphasis of their paper is on obfuscation and cross price elasticities between low and higher qualities of the same product offered by the firm; however they also indirectly obtain price elasticity estimates ranging from -25 to -40. In contrast, our analysis uses data from the complete set of firms listing prices across a broader selection of products, and where obfuscation is not prevalent. Ghose, Smith and Telang (2004) impute the sales of used books from the website of Amazon.com, which lists price offers for used books from many alternative and independent retailers. Using a multinomial logit model they estimate a price elasticity of -4.7, and note that the lowest priced firms receive disproportionately higher sales. Chevalier and Goolsbee (2003) rather ingeniously impute price elasticities for new books at two bookstores (Amazon and Barnes & Noble) using prices and relative sales rankings obtained directly from the retailers' websites rather than through a price comparison site (where price sensitivity may be expected to be greater). They estimate price elasticities of -0.6 for Amazon and -4 for Barnes & Noble, but these estimates do appear to be sensitive to the particular estimating technique adopted. Using a similar methodology, Ghose, Smith and Telang (2004) estimate the price elasticity for new books at Amazon to be -1.2.

The remainder of the paper proceeds as follows: The next section describes our data and provides an overview of the shopping environment at Kelkoo. In Section 3 we present the theory underlying our estimation methodology. Section 4 provides demand estimates based on individual as well as pooled products under the assumption that demand is continuous. These latter estimates are nested as a special case of the discontinuous demand specification, which is detailed in Section 5. Finally, Section 6 concludes.

2 Data

The proprietary data used in this paper were provided by the UK price comparison site, Kelkoo.com, which is owned by Yahoo! Within the UK, Kelkoo is the third largest retail website and attracts over 10 million individual users per month—more than twice that of its closest rival. Over 1,800 individual retailers—including 18 of the largest 20 online retailers in the UK—list prices on Kelkoo. According to Yahoo!, Kelkoo is the largest price listing service in the world, operating in seven other European countries besides the UK. It is recognized as one of the six most accessed websites in all of Europe.[1]

Consumers interested in purchasing a broad range of products can access the Kelkoo site to obtain information about the product and/or to obtain a list of retailers selling the

[1] Data taken from Hitwise Statistics and company information provided by Kelkoo.

product, together with the prices charged and other relevant information such as shipping charges. Consumers interested in making a purchase must do so from the website of the specific retailer and may easily transfer from the Kelkoo site to a retailer's site by clicking on one of the links provided.

Kelkoo's revenue is generated by charging retailers a fee for each referral made—that is, each time a consumer transfers from the Kelkoo site to a retailer's site. The fees charged vary across products and retailers, but typically range from £0.20 to £1.00 per lead. Kelkoo does not charge consumers any fees for using its site.

The products in our dataset consist of 18 of the most popular models of PDAs sold by 19 different retailers. These include models by Palm, HP, Sony, and Toshiba and span a wide range in prices. The lowest priced item is the Palm Handspring Treo, which has a median price of about £130, while the highest priced item is the Sony Clie nz90, with a median price of about £537.

Figure 1 shows a typical return from a price search on Kelkoo, which lists twelve retailers selling the HP iPAQ H5550 PDA. The information displayed includes a brief description of the product, the names of retailers selling the product, and price information detailed into item price, shipping charges ("P&P" in Kelkoo's terminology) and the total price inclusive of VAT. A consumer interested in purchasing the item may click on the "More" button, or the retailer's name or logo, to be transferred directly to the retailer's website. Figure 1 illustrates the heterogeneity in the types of retailers using the site and the wide range of prices charged for an identical product. Some retailers, such as Comet and PC World, are "bricks-and-clicks" retailers who have physical stores in addition to an online presence. Others, such as Amazon and Dell, are well-known pure e-retailers, while firms such as Big Gray Cat are less well-known specialty e-retailers.

Unlike many other price comparison sites, the order in which retailers' prices are displayed is unilaterally determined by Kelkoo; screen locations are neither auctioned nor sold directly to retailers, and are independent of the price charged.[2] Consequently, as far as both consumers and retailers are concerned, the order of price quotations for any specific screen

[2] Pricewatch.com, the price comparison site used by the e-retailer from whom Ellison and Ellison (2004) obtained sales data, lists retailers in order of price. Thus, the exercise of identifying screen location effects is not possible with their data.

appears random. Also in contrast to other studies that use price comparison site data, the complete list of prices is always displayed on a single page in our data. Thus, a change in a firm's screen location is never associated with a consumer having to click to "page two" in order to view the listing. Finally, we note that Kelkoo verifies and updates the information it displays daily.

Kelkoo maintains an information log for each "referral" generated at its site.[3] The log registers the retailer name, product name, price information, time of referral, location of the retailer on the screen and a cookie-specific reference. Kelkoo provided us with information extracted from their log files for the 18 PDA models for the period from 18 September 2003 to 6 January 2004, a period which generated over 40% of Kelkoo's annual traffic.[4]

This traffic amounted to 39,568 leads generated by 20,509 separate cookies. The majority (60.1%) of cookies generated only one lead, while a small number of cookies (0.56%) generated more than ten leads. Following Brynjolfsson, Dick, and Smith (2002), in instances where a consumer's cookie generated multiple leads, we use the consumer's last click as an indicator of her final choice.[5] Over the period of our study, there were 6,151 individual product, retailer and day specific price listings across the 18 PDA models. Our analysis is based on these observations, along with the number of last clicks generated for each PDA during the day of each listing.

The price used in our analysis is total purchase price—the actual cost to a consumer (including shipping and taxes) of purchasing a specific PDA model. To ensure an "apples to apples" comparison, we cleaned the data such that our analysis is based on listings of products that are identical in every respect (including condition).

Table 1 provides descriptive statistics for the data. The PDAs in our sample are somewhat pricey, and shipping accounts for only a small fraction of the total purchase price. The average total price of a PDA in our sample is £309.04 ($549.49 at exchange rates as of 8/05), of which an average of £4.16 is accounted for by shipping charges. The number

[3] Throughout the paper we use the terms "referral", "lead", and "click" interchangeably.

[4] Kelkoo is bound to protect the anonymity of retailers in disclosing information about the referrals they obtain. So in providing the information from their log files, the retailers were identified in the dataset by codes, and by some key characteristics, such as whether they had a brick and mortar presence.

[5] We also performed the analysis reported below using all clicks as well as only first-clicks data, and the results are qualitatively similar.

of sellers for a given product on any given day ranges from 1 to 15, with a mean of 4. The median number of clicks per day for a firm selling a specific PDA in our sample is 2. Consistent with the pattern observed in many traditional retail environments, referrals occur disproportionately in the fourth quarter of the year. However, as shown in Figure 2, online shopping disproportionately occurs on weekdays rather than weekends—opposite the pattern observed in many traditional retail environments.

Figure 3 suggests that price and screen location play a potentially important role in determining the business enjoyed by particular online retailers. Consumers appear to be very sensitive to price, as is evidenced by the dramatic decline in leads enjoyed by firms offering less favorable prices. Likewise, consumers tend to frequent firms that are listed above others on the screen. While screen location is not determined by price, it is possible that the results displayed in Figure 3 are the result of correlation between screen location and price. We deal with this issue formally in Section 4 of the paper.

3 Estimation Methodology

We now describe our methodology for estimating the impact of various explanatory variables on firm demand at the Kelkoo site. Given that we observe clicks and not sales, we offer conditions on expected demand which, if satisfied, allow us to interpret "click elasticities" as demand elasticities.

3.1 Data Generating Process for Leads

Recall that, to purchase a product, a consumer visiting the Kelkoo site must first process the information contained on the site and decide whether, and on which firm, to click. Following this, a consumer clicking through to the merchant's site obtains additional information about the desirability of purchasing the product and ultimately decides whether to buy it. Thus, demand can be decomposed into two parts: the click generating process and the process of converting leads into sales.

The process of generating leads depends on a number of factors, the most important of which are highlighted in Figure 4. As the figure shows, leads depend on the price a firm charges, the number of rival firms offering the same product and their prices, the identity of the firm and its rivals, the date and location of the firm's listing on the "page." Formally, let X denote this and other information obtained by the consumer directly from the Kelkoo site. Note that X may include dummy variables to control for product-specific characteristics (some products are more popular and receive more clicks, on average, than others), firm characteristics (some firms may have a brick-and-mortar presence while others do not), and time effects (firms may receive fewer clicks on weekends or products may exhibit life-cycle effects that cause clicks to vary systematically over time). Let the quantity of leads that firm i receives, Q_i, be drawn from some distribution $F_i(\cdot|X)$. Thus,

$$E[Q_i|X] = \int q dF_i(q|X) \tag{1}$$

where we use a Lebesque integral to account for the fact that Q_i is discrete. Based on the information in X—and this information alone—a representative consumer can decide to close his or her window or to click through to a particular merchant. To estimate the parameters associated with the data generating process for leads, we use a pseudo-maximum likelihood approach that does not require us to make specific assumptions about the underlying distribution generating Q_i; instead, we initially assume the underlying stochastic process has finite mean, given by

$$E[Q_i|X] = \exp[X\beta] \tag{2}$$

In order to estimate the vector of unknown parameters, β, one must account for the fact our clicks data consist of integer numbers of clicks. In fact, as shown in Table 1, over 50 percent of the data consist of days in which a firm selling a particular PDA received two or fewer clicks. For this reason, analysis of these data requires regression techniques suitable for count data. Thanks to recent advances in the econometrics of count data, a variety of estimation techniques are available.

One approach is to make a specific distributional assumption regarding the underlying stochastic process (Poisson or negative binomial, for instance), and use standard maximum likelihood estimation (MLE) methods to obtain estimates of the underlying parameters, β. Conditional on the underlying distributional assumption being correct, one obtains consistent estimates and standard errors and may perform standard hypothesis tests on β. Unfortunately, even if the mean specification in equation (2) is correct, it is known (see Gourieroux, *et al.* (1984a,b); Cameron and Trivedi, 1986) that the resulting maximum likelihood estimates of β and/or the standard errors will be inconsistent if the true stochastic process is different from that used to obtain maximum likelihood estimates.

For this reason, we adopt the pseudo-maximum likelihood (PML) approach due to Gourieroux, *et al.* (1984a,b) that has received renewed interest due to Cameron and Trivedi (1998) and Hall and Ziedonis (2001). Roughly, Gourieroux, *et al.* (1984a) show that so long as the mean specification in equation (2) is correct, any estimator for β obtained by maximizing the likelihood function based on the linear exponential class will be consistent for β even if the underlying distribution is misspecified. Since the Poisson distribution is in the linear exponential class but the negative binomial and other common specifications used for count data are not (when the parameters of the assumed distribution are unknown), we use the Poisson-based PML approach to obtain consistent estimates of β.

While the Gourieroux, *et al.* results imply that the MLE of β based on a Poisson distribution are consistent even when the underlying data generating process is not Poisson, the resulting estimates of the variance-covariance matrix are not consistent if the distribution is not Poisson. Following Hall and Ziedonis (2001), we use robust standard errors to obtain consistent estimates of the variance-covariance matrix.

To summarize, by using pseudo-maximum likelihood estimates based on a Poisson distributional assumption, we obtain a consistent estimate of β even if the underlying distribution is not Poisson. By using robust standard errors, we obtain consistent variance estimates. In contrast, maximum likelihood methods based on a specific distributional assumption (such as the negative binomial) would lead to more efficient estimates if the specification of the data generating process is correct, but inconsistent estimates if the distribution is not correct. As discussed below, we also provide MLE estimates based on specific distributional assumptions, including the negative binomial (see Cameron and Trivedi, 1998), as well as specifications that allow for unobserved firm heterogeneity (using both random and firm specific effects, as in Hausman, Hall, and Griliches 1984). Our results are robust to these alternative specifications.

3.2 Relating Leads to Demand

Up to this point, we have described a consistent methodology for estimating parameters associated with the leads generating process. In this subsection, we offer conditions under which such parameter estimates may be used to recover the elasticity of demand facing individual firms operating in the Kelkoo marketplace.

Toward this end, recall the process by which a click leads to a sale shown in Figure 4. After having observed information X at the Kelkoo site, a consumer clicking through to a firm's site receives additional information (denoted Z_i) that influences the consumer's decision to purchase. This information might include the firm's attempt at obfuscation along the lines described by Ellison and Ellison (2004), the visual attractiveness and usability of the firm's site, whether the firm is offering any guarantees on the product over and above those provided by the manufacturer, the exact restocking and return policies of the firm, and so on. Of course, a consumer's perceptions of these factors may be colored by the previous information, X, obtained on the Kelkoo site. To account for the possibility that a consumer observes Z_i's for all firms before making a purchase decision, let $Z = (Z_1, Z_2...Z_n)$ denote the vector of all such information. In this case, the probability that a click on firm i is converted into a sale, given (Z, X), is

$$\Pr\left(sale_i | Z, X\right) = G_i\left(Z, X\right)$$

Using equation (2), we may write the expected demand for a given product sold by firm i, conditional on (X, Z), as

$$\begin{aligned} E\left[D_i | X, Z\right] &= G_i\left(Z, X\right) \times E\left[Q_i | X\right] \\ &= G_i\left(Z, X\right) \times \exp\left[X\beta\right] \end{aligned}$$

The multiplicative separability of the probability of conversion, G_i, and the leads generating process follows naturally from the Kelkoo search and buying environment. Of central interest is the effect of information obtained from the Kelkoo site (various components of X) on final demand. Suppose one wished to measure the effect on firm i's demand of a change in x_i (some component of i's information posted at Kelkoo). It is useful to rewrite $X = (x_i, X_1)$ where X_1 represents all components of X other than x_i. If x_i influences firm i's leads but does not impact its conversion rate for that product, one may recover the firm's elasticity of demand with respect to x_i from its elasticity of leads. Formally,

Proposition 1 *Suppose that* $G_i(Z,(x_i,X_1)) = G_i(Z,(x_i',X_1))$ *for all* x_i, x_i'. *Then*

$$\frac{E[D_i|x_i,X_1,Z] - E[D_i|x_i',X_1,Z]}{E[D_i|x_i,X_1,Z]} = \frac{E[Q_i|x_i,X_1] - E[Q_i|x_i',X_1]}{E[Q_i|x_i,X_1]},$$

and furthermore, if demand is differentiable,

$$\frac{\partial \ln E[D_i|X,Z]}{\partial \ln x_i} = \frac{\partial \ln E[Q_i|X]}{\partial \ln x_i}.$$

Proof. We first prove the result for the differentiable case. Recall that log expected demand is given by

$$\ln E[D_i|X,Z] = \ln G_i(Z,X) + \ln E[Q_i|X]$$

Differentiating with respect to x_i yields

$$\frac{\partial \ln E[D_i|X,Z]}{\partial \ln x_i} = \frac{\partial \ln G_i(Z,X)}{\partial \ln x_i} + \frac{\partial \ln E[Q_i|X]}{\partial \ln x_i}$$

and since $G_i(Z,(x_i,X_1)) = G_i(Z,(x_i',X_1))$ for all x_i, x_i', $\frac{\partial \ln G_i(Z,X)}{\partial \ln x_i} = 0$. Hence

$$\frac{\partial \ln E[D_i|X,Z]}{\partial \ln x_i} = \frac{\partial \ln E[Q_i|X]}{\partial \ln x_i}.$$

Next, we prove the result for the non-differentiable case.

$$\begin{aligned}
\%\Delta E[D_i|(x_i,X_1),Z] &= \frac{G_i(Z,(x_i,X_1))E[Q_i|x_i,X_1] - G_i(Z,(x_i',X_1))E[Q_i|x_i',X_1]}{G_i(Z,(x_i,X_1))E[Q_i|x_i,X_1]} \\
&= \frac{E[Q_i|x_i,X_1] - E[Q_i|x_i',X_1]}{E[Q_i|x_i,X_1]}
\end{aligned}$$

where we have again used the fact that $G_i(Z,(x_i,X_1)) = G_i(Z,(x_i',X_1))$ for all x_i, x_i'. ■

Notice that the conditions of the Proposition allows firm i's conversion rate for product j to depend on additional information obtained from *all* firms' individual websites (Z), as well as all the other information posted at the Kelkoo site (X_1)—including information other than x_i posted by firm i. Provided that firm i's conversion rate is insensitive to x_i—that is, conditional on (x_i, X_1) being sufficiently "favorable" to induce the consumer to click through to firm i's site first place, the level of x_i does not influence the likelihood that the clickthrough is converted into a sale—the Proposition shows how to recover the associated demand parameter from an estimate of the leads generating process.

Two special cases of the proposition are noteworthy. First, when x_i is firm i's price listed on the Kelkoo site, one may interpret a firm's elasticity of "clicks" with respect to price as its own price elasticity of demand. Second, when x_i is a discrete variable (such as a dummy variable), one may use estimates based on clicks data to infer the percentage impact of a discrete change in x_i on a firm's demand.

Some support for this condition needed to apply Proposition 1 may be found in Brynjolfsson and Smith (2000) who show that, despite substantial variation in firms' prices, screen locations, and characteristics, conversion rates are insensitive to these differences. Their results suggest that conversion rates are insensitive to X, and *a fortiori*, insensitive to any particular information, x_i.

Nonetheless, one might still imagine circumstances where conversion rates depend both on specific information gleaned from the Kelkoo site (x_i) and information obtained by clicking through to firms' sites (Z). For instance, suppose that a consumer observes an extremely low price on the Kelkoo site (x_i), and clicks through to that firm's site to gather additional information (Z_i). In this case, the low price might induce the consumer to scrutinize the firm's site more carefully than had the firm been charging a higher price, and thus alter his or her conversion decision.

It is important to note, however, that even if the condition in Proposition 1 is not satisfied, one still obtains consistent estimates of parameters associated with the leads generating process. Furthermore, it is possible to use these estimates to obtain bounds on demand parameters. For instance, if one believes that firm i's conversion rate is increasing in x_i, then the estimated clickthrough elasticity is a lower bound for the associated demand elasticity (since in this case, $\frac{\partial \ln G_i(Z,X)}{\partial \ln x_i} + \frac{\partial \ln E[Q_i|X]}{\partial \ln x_i} > \frac{\partial \ln E[Q_i|X]}{\partial \ln x_i}$).

In the sequel, we assume that the condition in Proposition 1 is satisfied, so that we may recover the demand parameters from the estimated clicks parameters. Readers skeptical about the plausibility of this assumption are advised to interpret our results as estimates of clicks parameters or as bounds for underlying demand parameters. Even when viewed as purely estimates of clicks parameters, the results below are of considerable economic interest.

The business models of many of the most successful online companies, including Google and Yahoo!, are built on revenues derived from *clicks*— not from the conversion of clicks into sales. Understanding the determinants of click behavior in the online marketplace is arguably as important as understanding demand behavior.

4 Continuous Demand Models

In this section we provide estimates of a representative firm's demand under the assumption that each firm's demand is a continuous function of its price. As noted in the introduction, this is the usual way of estimating demand, and has led to firm elasticity estimates ranging from –0.6 (in the relatively concentrated online market for books) to about -40 (in a highly competitive online market for computer memory). The main message of this section is that differences in concentration across different markets for PDAs may help explain variation in price elasticities.

4.1 Estimates by Product

As a starting point, we pool across firms (i) and dates (t), but estimate separate elasticities for each of the 18 different models of PDAs in our data using the pseudo-maximum likelihood procedure described above. Specifically, we assume

$$E\left[Q_{ijt}|X_{ijt}\right] = \exp\left[\beta_j \ln p_{ijt} + \gamma_j X_{1,ijt}\right], \tag{3}$$

where Q_{ijt} is the number clicks firm i received on product j at time t, p_{ijt} is the total price (including VAT and shipping) firm i charged for product j at time t, and $X_{1,ijt}$ is a vector of controls. Notice that, under our maintained hypothesis that the conditions in Proposition 1 hold, β_j may also be interpreted as the own price elasticity of demand for a representative seller of a model j PDA. The vector $X_{1,ijt}$ consists of the following controls:

Position on Screen. As we showed earlier in Figure 3, when a firm's price is listed above its rivals, it tends to receive more clicks. Clicks tend to decrease as the position on the screen gets lower. Hence, we include a linear position on screen variable to capture this effect.[6]

Weekend. As displayed in Figure 2, firms systematically receive fewer clicks on weekends. Hence, we include a weekend dummy variable to control for this effect on demand.

Month. We include month dummies to control for seasonal effects on demand.

Table 2 reports the results of the individual product regressions. Notice that 13 of the estimated own price elasticities in Table 2 are statistically significantly different from zero at the 1 percent level, with values ranging from -1.75 (for the Toshiba E770) to -14.691 (for the HP Compaq iPAQ 1940). These estimates vary widely across PDAs. In interpreting these results, and to better understand the widely different estimates obtained for different models of PDAs, it is important to recognize that these estimates are firm elasticities — not market elasticities. One of the key theoretical determinants of a firm's price elasticity is the availability of substitutes—the more sellers offering the same product, the more elastic is the demand facing a firm selling that product. For instance, it is well-known that in a symmetric n-firm capacity-constrained price-setting environment, the elasticity of demand facing an individual firm (E_F) is n times the market elasticity (E_M): $E_F = nE_M$. If this is the case and different numbers of firms sold different types of PDAs, the firm elasticities of demand would vary widely across PDA models even if the market elasticity of demand were the same for each model of PDA.

Thus, it seems useful to investigate the relationship between the elasticity estimates reported in Table 2 and the average number of firms selling each PDA. This relationship is plotted in Figure 5. The estimates are divided into those that are not statistically significant at conventional levels (shown as open circles) and those that are significant at the 1% level (shown as filled-in diamonds). As Figure 5 shows, there is a strong negative relationship between the elasticity estimates for each of the products and the average number of firms offering price quotes for the product. This suggests the need to control for the number of sellers were one to pool across all products.

[6] We investigated nonlinear specifications of screen location as well and obtained similar results to those reported here. We chose the linear specification for parsimony.

4.2 Pooled Estimates

We now report estimates obtained by pooling across firms (i), dates (t), and different models of PDAs using our pseudo-maximum likelihood procedure. Here we consider two models: a baseline model that does not allow elasticities to vary with the number of sellers, and a more general model that takes into account our preliminary findings in the individual product specifications. The baseline model assumes

$$E\left[Q_{ijt}|X_{ijt}\right] = \exp\left[\beta \ln p_{ijt} + \gamma X_{1,ijt}\right]. \tag{4}$$

The controls for this specification include all of those in equation (3) as well as following:

Product. As the previous specification revealed, there are differences in clicks for each of the different PDA models. For instance, PDAs differ from one another in terms of their popularity, their operating system, various performance characteristics, add-on software, and so on. Thus we include product dummies for each of the 18 PDA models.

Product-Month Interactions. In addition, the popularity of a PDA varies depending on new entrants in the PDA product space. As technology and performance improve with the introduction of new models, the popularity of an existing PDA can decline—sometimes dramatically. To control for these effects, we include dummies interacting each of the product dummies with the month dummies mentioned above. This, in principle, allows PDA "life cycles" to differ during the time horizon of our study.

Bricks and Clicks Retailer. Some of the firms in our dataset have an established physical presence in addition to their online presence. Clearly, the reputation as well as the ease of returns and accumulated brand equity of these retailers may differ from pure online sellers. Thus, we include a dummy variable for whether a particular firm is a bricks and clicks retailer.

With these controls in place, we report pseudo-maximum likelihood estimates (Table 3, Model 1) based on the mean specification in equation (4). The bottom of Table 3 also reports the results of a likelihood ratio test for overdispersion of the negative binomial (2) type (cf. Cameron and Trivedi, 1990). This is a test of the null hypothesis that the mean and variance of the click generating process are equal, as would be the case were the data generating process truly coming from a Poisson distribution. As the table shows, we overwhelmingly reject this hypothesis, indicating that the underlying distribution is not Poisson. As discussed above, the parameter estimates are nonetheless consistent (provided the mean specification in equation (4) is correct), but the overdispersion test indicates that Poisson-based maximum likelihood estimates of their standard errors are *not* consistent. To obtain consistent variance-covariance estimates, we employ the standard error correction techniques of Rogers (1993), Huber (1967) and White (1980,1982).[7] The corresponding z-statistics are reported in Table 3.[8]

The results show a price elasticity of -4.61, which is fairly close to the average over the individual product elasticities reported in Table 2. More favorable screen positions lead to increased clicks: All else equal, a firm that moves up one screen position enjoys an 18.6% increase in demand. These results confirm what we saw earlier in Figure 3: There is a strong tendency for consumers to click on firms listed at the top of the display screen, all else equal. This may also explain why search engines that auction screen positions, such as Google, receive significant premia for positions located near the top of the screen. Interestingly, while the coefficient associated with being a bricks and clicks retailer has the expected positive sign (0.262), it is not significant at conventional levels.

To account for a potential relationship between a firm's elasticity of demand and the number of competing sellers in the pooled model, we generalize equation (4) to allow individual firm elasticities to depend on the number of sellers as follows:

$$E\left[Q_{ijt}|X_{ijt}\right] = \exp\left[\left(\beta_0 + \left(n_{jt} - 1\right)\beta_1\right)\ln p_{ijt} + \beta_2 n_{jt} + \gamma X_{1,ijt}\right], \tag{5}$$

where n_{jt} denotes the number of sellers of type j PDA on date t. Notice that, in this specification, the elasticity of demand facing a representative firm is given by

$$\beta_0 + \left(n_{jt} - 1\right)\beta_1.$$

Thus, the coefficient of total price (β_0) represents the elasticity of demand facing a monopoly seller, $\beta_0 + \beta_1$ represents the elasticity of demand in duopoly PDA markets, and more

[7] Specifically, we use the grouping technique of Rogers (1993) to relax the independence of observations for a given firm i across products and time. This allows potential autocorrelation and heteroskedasticity in the errors.

[8] Some researchers have taken the view that the rejection of the null hypothesis of no overdispersion warrants the use of a negative binomial specification. For this reason, we report ML estimates based on the negative binomial (2) specification in Table A1. As that table shows, the parameter estimates are very similar.

generally, β_1 represents the impact on a firm's elasticity of demand of facing an additional competitor. In addition to our earlier controls, we include the following:

The Number of Sellers. Besides the theoretical rationale for permitting a representative firm's elasticity of demand to depend on the number of sellers, one might expect the number of clicks received by a particular firm to directly depend on the number of sellers. For a given consumer base, adding additional sellers would tend to reduce the expected number of clicks enjoyed by any particular firm. In addition, one might speculate that consumers are more likely to click and purchase PDAs that are sold by more firms, as additional firms might stimulate online sales by making the market appear more credible in the eyes of consumers. As we will see below, our framework permits one to disentangle these two competing effects. We include a linear term for the number of sellers.[9]

The resulting estimates are displayed in the Model 2 column of Table 3. As the table shows, the number of sellers has a significant effect—both in terms of levels as well as on price elasticities. Controlling for the number of firms listing prices, we find that the price elasticity of a monopoly seller is -3.761, which implies a gross margin of 26.6%. Adding a second firm to the market raises the price elasticity to around -4.049 and cuts the gross margin to 24.7%. When ten firms list prices, the estimated elasticity becomes -6.641 or about 15.1% gross margins. The sign of the coefficient capturing the impact of the number of sellers on elasticity is also consistent with the simple capacity constrained price setting model described above.

What is the impact of a change in the number of sellers on a firm's overall demand? As we mentioned above, there is a direct effect as well as an indirect effect from increased competitiveness. Taking the derivative of equation (5) and evaluating it at the mean of our data yields

$$\begin{aligned} \frac{\partial \ln E\left[Q_{ijt}|X\right]}{\partial n_{jt}}\Big|_{\bar{p}_{ijt}} &= \hat{\beta}_1 \ln \bar{p}_{ijt} + \hat{\beta}_2 \\ &= -.288 \times 5.67 + 1.593, \end{aligned}$$

or about -0.04 ($p = .0155$). It is useful to contrast the magnitude of this "rivalry" effect with that of a change in a firm's screen position. As Table 3 shows, a reduction of one screen

[9] We also ran dummy specifications for the number of sellers and obtained similar results.

position decreases the firm's demand by 17.5%. Thus, our estimates suggest that the impact of screen position is more than four times larger than the impact of an additional competitor appearing on the price comparison site.

4.3 Potential Misspecification

One may have a number of concerns regarding the estimates based on the continuous demand specification in equation (5). As we have emphasized, the pseudo-maximum likelihood approach is robust against alternative distributional assumptions but not to the misspecification of the underlying mean of the stochastic process.

First and foremost, price comparison sites are often used by consumers looking to obtain a given product at the best price. For instance, Brynjolfsson and Smith (2000) have provided evidence that 49 percent of consumers using price comparison sites in the U.S. make purchase decisions based purely on price. The results of Ghose, *et al.* (2004) seem to indicate a jump in a firm's demand when it sets the lowest price. Moreover, recall that in our data (see Figure 3), 45 percent of the clicks are at the lowest price. These observations, coupled with the recent literature that rationalizes the observed levels of price dispersion in online markets (see Baye and Morgan, 2001; Baye, Morgan, and Scholten, 2004a) suggests that a firm lowering its price from the second-lowest to the lowest price enjoys a discontinuous jump in demand.

To see the potential ramifications of this on demand estimation, suppose there is a unit mass of consumers, half of which are "shoppers" who purchase at the lowest price and the other half are "loyals" who have a preference for a particular seller. Consumers within each group have identical demand functions given by $D = p^{-\theta}$. A firm that charges the lowest price in the market enjoys demand from both groups, while a firm charging a price above the minimum price in the market sells only to its loyal customers. Figure 6 illustrates the ramifications on demand estimation. The slope of the two steep lines through the data are the same, and represent the true elasticity of demand, $-\theta$, for prices above or below the minimum price. At the minimum price, there is a discontinuous jump in demand owing to the fact that the firm attracts all of the shoppers at this price.

The dashed line through the data represents the elasticity estimate that results from failing to take into account the discontinuous jump in demand that occurs when the firm charges the lowest price. Notice that, by ignoring the jump in demand at the lowest price, one obtains an estimate of the true elasticity that overstates how responsive consumers are to a change in price.

In addition to the potential problem caused by using a continuous demand specification in the presence of "shoppers," two additional econometric issues are potentially relevant. First, while there are sound theoretical reasons for elasticities (and per-firm demand) to depend on the number of firms listing prices, the estimates may be biased due to potential endogeneity. In particular, popular products are likely to (for a given number of firms) result in a firm receiving more clicks, and this may encourage additional firms to enter the market.

Second, while we have controlled for one firm characteristic—whether a firm is a bricks and clicks retailer—a variety of unobserved firm characteristics, such as the degree of accumulated brand equity or differences in consumers' perceptions of firm quality, could also potentially bias our results. Thus, it may be important to account for unobserved firm characteristics in estimating demand.

We address these and other issues in the next section.

5 Discontinuous Demand Models

We now turn to estimating demand in the presence of a mix of price-sensitive shoppers and loyals. We first sketch the theory underlying the demand estimation. We then describe the estimating equation and report results. Finally, we examine issues associated with endogeneity and unobserved firm characteristics.

5.1 Theory and Estimation Strategy

Suppose that n_{jt} firms numbered $i = 1, 2, ..., n_{jt}$ sell product j at a price comparison site on date t. Let p_{ijt} denote the total price of firm i in this market. A firm in this market sells to two types of consumers: Loyals, who purchase from their preferred firm, and shoppers, who always purchase from the firm charging the lowest price. Because of the extreme price sensitivity of shoppers, it is useful to define the set of firms offering the "best" (lowest) price for product j at time t as:

$$B_{jt} = \{i : p_{ijt} \leq p_{kjt} \text{ for all } k \neq i\}.$$

Let Q^S_{ijt} and Q^L_{ijt} denote the product j leads firm i obtains from shoppers and loyals, respectively, when charging the price p_{ijt}. Recall that firm i obtains leads from shoppers only if it is in the set B_{jt}; that is, if it offers one of the best prices. Thus, the clicks firm i obtains when it charges a price p_{ijt}, given the prices charged by other firms, is

$$Q_{ijt} = \begin{cases} Q^S_{ijt} + Q^L_{ijt} & \text{if} \quad i \in B_{jt} \\ Q^L_{ijt} & \text{if} \quad i \notin B_{jt} \end{cases}$$

Thus, firm i faces a "jump in demand" for product j when it is among those firms offering the "best" price for product j on date t.

We utilize the following functional approach that facilitates structural estimation of demand in a clearinghouse model. To account for the discontinuity in demand when the firm offers one of the best prices in the market, let $\mathbf{I}_{jt}$ be an indicator function that equals unity when $i \in B_{jt}$ and zero otherwise, and let $\#B_{jt}$ denote the cardinality of B_{jt}; that is, the number of firms offering the best price for product j. Suppose that firm i 's elasticity when it sells product j is θ_{jt}, so that we may write

$$Q_{ijt} = \alpha^L_{ijt}(X)\, p_{ijt}^{-\theta_{jt}} + \mathbf{I}_j \frac{1}{\#B_{jt}} \alpha^S_{ijt}(X)\, p_{ijt}^{-\theta_{jt}}$$

where $\alpha^L_{ijt}(X)$ and $\alpha^S_{ijt}(X)$ represent the non-price determinants of leads (such as screen location) on loyals and shoppers, respectively. To ease the notational burden, we suppress the X argument where it is clear. Hence we may rewrite Q_{ijt} as:

$$\begin{aligned} Q_{ijt} &= \left(\alpha^L_{ijt} + \frac{\mathbf{I}_{jt}}{\#B_{jt}} \alpha^S_{ijt}\right) p_{ijt}^{-\theta_{jt}} \\ &= \left(1 + \frac{\mathbf{I}_j}{\#B_j} \lambda_{ijt}\right) \alpha^L_{ijt} p_{ijt}^{-\theta_{jt}} \end{aligned}$$

where

$$\lambda_{ijt} = \frac{\alpha^S_{ijt}}{\alpha^L_{ijt}}.$$

Taking logs (and noting that $\ln[1 + \lambda_{ijt} \frac{\mathbf{I}_{jt}}{\#B_{jt}}] \approx \lambda_{ijt} \frac{\mathbf{I}_{jt}}{\#B_{jt}}$) yields

$$\ln Q_{ijt} = \lambda_{ijt} \frac{\mathbf{I}_{jt}}{\#B_{jt}} + \ln \alpha^L_{ijt} - \theta_{jt} \ln p_{ijt} \tag{6}$$

Estimation requires imposing additional structure on the parameters in equation (6). We assume

$$\theta_{jt} = \beta_0 + (n_{jt} - 1)\,\beta_1 \tag{7}$$

As in the previous section, this parsimonious specification allows a firm's elasticity for product j to depend on the number of sellers at time t. In addition, we allow different firms to have different numbers of loyals and shoppers, and also permit the number of each to vary over time and across products. However, we assume

$$\begin{aligned} \alpha^S_{ijt}(X_{ijt}) &= a^S \alpha_{ijt}(X_{ijt}) \\ \alpha^L_{ijt}(X_{ijt}) &= a^L \alpha_{ijt}(X_{ijt}) \end{aligned}$$

so that the ratio of these two expressions is constant. In particular, this assumption implies

$$\lambda_{ijt} = \lambda = \frac{a^S}{a^L}. \tag{8}$$

Under these assumptions, the mean specification is:

$$E[Q_{ijt}|X] = \exp\left[(\beta_0 + (n_{jt} - 1)\,\beta_1)\ln p_{ijt} + \beta_2 n_{jt} + \lambda \frac{\mathbf{I}_{jt}}{\#B_{jt}} + \gamma X_{1,ijt}\right], \tag{9}$$

where $X_{1,jt}$ is the matrix of controls discussed earlier (position on screen, bricks and clicks retailer, weekend, month, and product dummies as well as product-month interaction dummies). As above, we may interpret the expression $\beta_0 + (n_{jt} - 1)\,\beta_1$ as the price elasticity of a firm that has $n_{jt} - 1$ rivals. Similarly, we may interpret λ as the demand shift from shoppers; that is, the size of the discontinuous jump in demand a firm enjoys when it offers the "best" price. Notice that the continuous demand model is nested in the specification of equation (9) when $\lambda = 0$; thus, we may readily test the null hypothesis implied by the continuous demand specification.

5.2 Discontinuous Demand Estimates

Model 1 in Table 4 reports PML estimates of the parameters in equation (9). Recall that under the nested model of continuous demand, the coefficient associated with the demand shift from shoppers (λ) is equal to zero. The alternative hypothesis, predicted from the clearinghouse literature, is that this coefficient should be positive. The coefficient estimate for this effect is 0.603. Moreover, we can reject the null hypothesis of the continuous demand model in favor of the (one-sided) alternative of discontinuous demand at the 1% significance level. In short, we find considerable evidence for a discrete shift in demand when a firm offers the lowest price.

Figure 6 suggested that, in the presence of such a jump in demand, the continuous demand specification would yield more elastic estimates of a firm's demand than the discontinuous specification. The results shown in Model 1 of Table 4 compared to those in Model 2 of Table 3 are consistent with this observation. Accounting for the discontinuity in demand, the estimated elasticity for a monopoly seller becomes less elastic—going from -3.761 (in Model 2 of Table 3) to -2.459 (in Model 1 of Table 4). The difference in the elasticity estimates is greater for markets with more than one seller: The effect of an additional rival on the price elasticity is reduced by around 12.5 percent (from -0.288 in Model 2 of Table 3 to -0.252 in Model 1 of Table 4). We also note that, in contrast to the continuous demand specification, the effect of a change in the number of firms on a firm's overall demand $\left(\partial \ln E\left[Q_{ijt}|X\right]/\partial n_{jt}|_{\bar{p}_{ijt}}\right)$ is not statistically different from zero ($p = .4674$).

It is of some interest to note the economic relevance of our estimate of the demand shift from shoppers ($\lambda = 0.603$). Other things equal, a firm that sets the lowest price in the market enjoys a 60.3 percent increase in demand. In contrast, notice that the "position on screen" coefficient implies that a firm would have to move up 3 screen positions to generate the same demand increase that results from setting the lowest price in the market. Also, note that setting the lowest price in the market entails a demand shift that is about twice as large as the 32.1 percent shift associated with being a bricks and clicks retailer.

One may use our estimates of λ to obtain a very crude estimate of the fraction of consumers using the Kelkoo site who are shoppers. The total number of clicks for product j on a given date is

$$\sum_{i=1}^{n_{jt}} Q_{ijt} = \sum_{i=1}^{n_{jt}} \left(\alpha_{ijt}^{L} + \frac{\mathbf{I}_{jt}}{\#B_{jt}}\alpha_{ijt}^{S}\right) p_{ijt}^{-\theta_{jt}}$$

while the corresponding number of clicks stemming from shoppers is

$$\sum_{i=1}^{n_{jt}} Q_{ijt}^S = \frac{\mathbf{1}}{\#B_{jt}} \sum_{i\in B_{jt}} \alpha_{ijt}^S p_{ijt}^{-\theta_{jt}}$$

Hence, shoppers as a fraction of all consumers is given by

$$\begin{aligned}
\frac{S}{S+L} &= \frac{\sum_{i=1}^{n_{jt}} Q_{ijt}^S}{\sum_{i=1}^{n_{jt}} Q_{ijt}} \\
&= \frac{\mathbf{1}}{\#B_{jt}} \frac{\sum_{i\in B_{jt}} \alpha_{ijt}^S p_{ijt}^{-\theta_{jt}}}{\sum_{i=1}^{n_{jt}} \left(\alpha_{ijt}^L + \frac{\mathbf{I}_{jt}}{\#B_{jt}} \alpha_{ijt}^S\right) p_{ijt}^{-\theta_{jt}}} \\
&= \frac{\mathbf{1}}{\#B_{jt}} \frac{\sum_{i\in B_{jt}} a^S \alpha_{ijt}(X_{ijt}) p_{ijt}^{-\theta_{jt}}}{\sum_{i=1}^{n_{jt}} a^L \alpha_{ijt}(X_{ijt}) p_{ijt}^{-\theta_{jt}} + \frac{\mathbf{1}}{\#B_{jt}} \sum_{i\in B_{jt}} a^S \alpha_{ijt}(X_{ijt}) p_{ijt}^{-\theta_{jt}}}
\end{aligned}$$

Imposing symmetry across firms (so that all of the above terms are independent of i), one obtains

$$\begin{aligned}
\frac{S}{S+L} &= \frac{\mathbf{1}}{\#B_{jt}} \frac{\sum_{i\in B_{jt}} a^S \alpha_{jt}(X_{jt}) p_{jt}^{-\theta_{jt}}}{\sum_{i=1}^{n_{jt}} a^L \alpha_{jt}(X_{jt}) p_{jt}^{-\theta_{jt}} + \frac{\mathbf{1}}{\#B_{jt}} \sum_{i\in B_{jt}} a^S \alpha_{jt}(X_{jt}) p_{jt}^{-\theta_{jt}}} \\
&= \frac{a^S \alpha_{jt}(X_{jt}) p_{jt}^{-\theta_{jt}}}{n_{jt} a^L \alpha_{jt}(X_{jt}) p_{jt}^{-\theta_{jt}} + a^S \alpha_{jt}(X_{jt}) p_{jt}^{-\theta_{jt}}} \\
&= \frac{\lambda}{n_{jt} + \lambda}
\end{aligned}$$

which implies (given the estimate of $\lambda = .603$ reported in Model 1 of Table 4 and the mean number of listings (4.05) in our data) that about 13 percent of consumers at Kelkoo are shoppers. While the symmetry assumptions used to calculate this crude estimate are at odds with the data (among other things, the estimates suggest that bricks-and-clicks sellers receive 32.1 percent more clicks than pure online sellers), it nonetheless illustrates that even in online markets where nearly 90 percent of the consumers are "loyal" to a particular firm, discontinuities arising from shoppers can significantly impact elasticity estimates.

5.3 Potential Misspecification

While the PML approach used to obtain the estimates reported above does not make specific distributional assumptions about the underlying clicks generating process (apart from assuming that the conditional mean specification is correct), it is nonetheless useful to compare the shape of the empirical distribution with the distribution based on the number of clicks predicted by the model. As Figure 7 reveals, the distribution of predicted clicks (based on Model 1 in Table 4) resembles that observed in the data. While this is somewhat reassuring, it does not rule out the possibility that Model 1 in Table 4 is misspecified due to endogeneity or unobserved heterogeneity across firms.

Endogeneity

One concern that might be raised with the preceding analysis is that two of the key variables of interest—price and the number of listing firms—might be "endogenous" in the sense that the regressors may be correlated with omitted variables, thus creating the possibility of inconsistent parameter estimates. This problem might arise through unobserved demand shifters such as variations in a product's popularity. Such changes in popularity would obviously impact a given firm's demand. In addition, changes in popularity would also be correlated with the number of sellers (an increase in popularity would presumably induce more sellers to enter) as well as the prices charged for the product (firms would raise prices for "hot" products and reduce them for "cold" products).

To be concrete, suppose that the correct clicks generating process is given by:

$$Q_{ijt} = \exp\left[\left(\beta_0 + \left(n_{jt} - 1\right)\beta_1\right)\ln p_{ijt} + \beta_2 n_{jt} + \lambda \frac{\mathbf{I}_{jt}}{\# B_{jt}} + \gamma X_{1,ijt} + pop_{jt}\right] + \varepsilon_{ijt} \tag{10}$$

where pop_{jt} denotes an unobserved latent variable and ε_{ijt} is a zero mean error term. Then, the correct conditional mean specification is:

$$E\left[Q_{ijt} | X_{ijt}, pop_{jt}\right] = \exp\left[\left(\beta_0 + \left(n_{jt} - 1\right)\beta_1\right)\ln p_{ijt} + \beta_2 n_{jt} + \lambda \frac{\mathbf{I}_{jt}}{\# B_{jt}} + \gamma X_{1,ijt} + pop_{jt}\right]$$

and hence equation (9) is misspecified. However, the parameter estimates based on equation (9) remain consistent provided:

$$E\left[Q_{ijt} | X_{ijt}, pop_{jt}\right] = E\left[Q_{ijt} | X_{ijt}\right] \tag{11}$$

That is, to the extent that fluctuations in pop_{jt} only occur *inter-month*, any influence of pop_{jt} on $E\left[Q_{ijt} | X_{ijt}, pop_{jt}\right]$ will be absorbed in the product-time interaction dummies contained in $X_{1,ijt}$, and equation (11) will hold—the mean specification given in equation (9) will be correct. If there is *intra-month* variation in pop_{jt}, however, this variation will not be absorbed in the product-time interaction dummies. Furthermore, for the reasons described above, pop_{jt} is likely to be correlated with p_{ijt} and n_{jt}. In this case, equation (11) will not hold, and the parameter estimates reported above will be inconsistent.

To deal with this potential endogeneity problem, we collected additional data from the US to serve as a proxy for unobserved demand shifters in the UK (i.e. the variable pop_{jt}). In particular, for each of the PDAs in our UK sample and for each date, we obtained data on that PDA's product popularity ranking (for that same day) from Shopper.com—a US price comparison site.[10] The product popularity ranking are integers (1 represents the most popular product, 2 the second most popular product, and so on), and are constructed by Shopper.com based on consumer pageviews at its site. It seems plausible that the product popularity ranking of an identical PDA model in the US on a given date is a valid proxy for its popularity in the UK for that same date, as it is likely correlated with pop_{jt} but uncorrelated with the other UK regressors included in equation (10).

Formally, suppose that

$$pop_{jt} = \eta D_{jt}^{USPOP} + \nu_{jt}$$

where D_{jt}^{USPOP} is a matrix of dummy variables for the popularity of product j at time t from the US data, η is a vector of parameters associated with each product rank, and ν_{jt} is an error term. Substituting this relation into equation (10), one obtains

$$Q_{ijt} = \exp\left[(\beta_0 + (n_{jt} - 1)\beta_1)\ln p_{ijt} + \beta_2 n_{jt} + \lambda \frac{\mathbf{I}_{jt}}{\#B_{jt}} + \gamma X_{1,ijt} + \eta D_{jt}^{USPOP}\right] \exp[\nu_{jt}] + \varepsilon_{ijt}$$

Assuming ν_{jt} is independent of the other regressors, this implies a conditional mean specification of

$$E[Q_{ijt}|X_{ijt}, pop_{jt}] = \exp\left[(\beta_0 + (n_{jt} - 1)\beta_1)\ln p_{ijt} + \beta_2 n_{jt} + \lambda \frac{\mathbf{I}_{jt}}{\#B_{jt}} + \gamma X_{1,ijt} + \eta D_{jt}^{USPOP}\right] \tag{12}$$

where we have assumed, without loss of generality, that $E[\exp[\nu_{jt}]] = 1$.

Under the stated assumptions, pseudo-maximum likelihood estimation of equation (12) gives consistent parameter estimates. We report the results of this specification as Model 2 in Table 4. As the table shows, controlling for potential endogeneity does little to the

[10] The US data is discussed in more detail in Baye, Morgan and Scholten (2004a).

magnitude or the significance of the coefficient estimates.[11]

Unobserved Firm Heterogeneity

Another potential shortcoming of the PML approach used in Models 1 and 2 of Table 4 is that the specification presumes there is no unobserved heterogeneity across firms. While we have attempted to control for differences across firms that stem from their having different online and offline presences, as well as different screen locations, it is still possible that a particular firm's demand is also driven by unobserved factors. For this reason, we also report in Table 4 results that allow for the effects of unobserved firm heterogeneity.

Model 3 in Table 4 reports maximum likelihood estimates of the discontinuous demand model based on the random effects specification for unobserved firm heterogeneity pioneered by Hausman, Hall and Griliches (1984), while Model 4 reports conditional maximum likelihood estimates based on a fixed effects specification for unobserved firm heterogeneity. Note that these results require the specification of the actual likelihood function, which we have take to be Poisson. However, Table A2 in the appendix shows that the results reported in Table 4 and discussed below are similar if one uses the likelihood function for a negative binomial (2) specification.

Notice that, in both the random effects (Model 3) and fixed effects (Model 4) specifications, the coefficients of interest are roughly comparable to those obtained ignoring potential unobserved heterogeneity (Models 1 and 2). As the coefficient on the demand shift from shoppers reveals, we still reject the null hypothesis of the continuous demand model in favor of the discontinuous demand specification. Further, the economic value of the coefficient associated with the demand shift is largely unchanged by allowing for potential unobserved heterogeneity. Likewise, the coefficient associated with the elasticity of demand for a monopoly firm remains at about -2.5, similar to the estimate obtained in Model 1.

In contrast, both the effect of a change in the number of rivals on a firm's price elasticity as well as the effect of change in screen position on a firm's demand are dampened in Models 3 and 4 compared to Model 1. This suggests that part of the effect that was previously attributed to changes in screen position or rivalry is more properly accounted for

[11] There is some evidence that the US popularity dummies address the endogeneity issues discussed above. In all specifications that include US popularity dummies, we reject the null hypothesis that the coefficients on D_{jt}^{USPOP} are jointly equal to zero ($p < 0.001$).

by unobserved firm heterogeneities. Nonetheless, the coefficients on these variables remain economically and statistically significant.

6 Conclusions

Using a unique dataset consisting of consumer clickthroughs for 18 PDAs obtained from the Yahoo! price comparison site, Kelkoo.com, we estimated key demand parameters for firms competing in an important UK e-retail market. Taking advantage of pseudo-maximum likelihood techniques for count data and applying insights from theoretical “clearinghouse” models of online competition, we provided consistent parameter estimates without making strong distributional assumptions about the data generating process. Furthermore, we showed in Proposition 1 that one may recover key demand parameters when only clicks data are available.

What do our results suggest about the competitiveness of the UK e-retail market? The conventional wisdom is that price is paramount in online markets—firms are forced into cutthroat competition to attract the business of consumers who only care about price. Theoretical clearinghouse models take a more nuanced view, suggesting that a firm’s demand exhibits a jump when it offers the lowest price and thus succeeds in attracting price sensitive shoppers. This leads to two questions: Do these price-sensitive shoppers really exist, and, if so, are there enough of them to be economically relevant?

Our results suggest that the answer to both questions is yes. Using a novel estimation strategy that allows for the possibility of such a jump, we find that the lowest-priced seller of a PDA enjoys a 60 percent increase in demand compared to the case where its price is not the lowest in the market. We also showed that a jump of this magnitude arises in clearinghouse models even when only 13 percent of consumers are price-sensitive “shoppers.” Moreover, failing to account for such a jump when estimating a firm’s demand in online markets results in biased parameter estimates. In the data we analyzed, price elasticity estimates based on a continuous demand specification were about two times more elastic than those obtained in a more general specification that allows for a jump in demand.

In addition to price, much of the industrial organization literature has stressed the importance of market structure—the number of sellers competing in a given market—on competition. Our dataset provided a unique opportunity to examine this question owing to daily variation in the number of sellers of a given product on the Kelkoo site. We find that the number of rivals offering the same product strongly affects a firm's price elasticity. When a firm is the sole seller of a product, its price elasticity is -2.459. In contrast, when four firms sell the same product (approximately the mean in our data), a firm's demand becomes considerably more elastic at -3.215.

Firms' efforts to differentiate themselves from rivals also play an important role in the UK e-retail markets we examined. For example, a standard prescription in the e-retail strategy literature is that a firm can gain a competitive advantage in an online market by leveraging its offline presence to vertically differentiate itself from purely online competitors. The lack of success by Barnes & Noble in pursuing this strategy against Amazon, however, raises questions about the value of this perceived advantage. Our results suggest that, in the UK market for PDAs, the "bricks and clicks" advantage is substantial: Other things equal, bricks and clicks e-retailers enjoy 32.1 percent higher demand than their purely online competitors. Effects of this magnitude should clearly be an important practical consideration for conventional retailers in the UK looking to pursue a "straddle" strategy of moving into the online space.

Finally, we find that non-pecuniary considerations also play an important role in influencing a firm's demand online. All else equal, a firm listed first on the display screen at a price comparison site enjoys 17.5 percent higher demand than when it is second on the list—despite the trivial cost to the consumer (one mouse click) of reordering the list of offers by price.

References

[1] Baye, Michael R. and John Morgan, "Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets," *American Economic Review* (2001) 91(3), pp. 454-474.

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

[3] Baye, Michael R., John Morgan, and Patrick Scholten, "Temporal Price Dispersion: Evidence from an Online Consumer Electronics Market," *Journal of Interactive Marketing* (2004b) 18, pp. 101-115.

[4] Brynjolfsson, Erik, Astrid Dick, and Michael Smith, "Consumer Heterogeneity and Search Costs on the Internet," Working Paper, Carnegie Mellon University, 2002.

[5] Brynjolfsson, Erik and Michael D. Smith, "The Great Equalizer? Consumer Choice Behavior at Internet Shopbots." MIT Working Paper, 2000.

[6] Chevalier, J. and A. Goolsbee, "Measuring Prices and Price Competition Online: Amazon.com and BarnesandNoble.com," *Quantitative Marketing and Economics*, (2003) 1(2), pp. 203 - 222.

[7] Cameron, A. and P. Trivedi, "Econometric Models Based on Count Data: Comparisons and Applications of Some Estimators," *Journal of Applied Econometrics* (1986) 1, pp. 29-53.

[8] Cameron, A. and P. Trivedi, "Regression-based tests for overdispersion in the Poisson model," *Journal of Econometrics* (1990) 46(3), pp. 347-364.

[9] Cameron, A. and P. Trivedi, *Regression Analysis of Count Data,* Econometric Society Monograph No. 30 (1998).

[10] Ellison, Glenn and Sara Fisher Ellison, "Search, Obfuscation, and Price Elasticities on the Internet." *mimeo,* 2004.

[11] Ghose, A., M. Smith, and R. Telang, "Internet Exchanges for Used Books: An Empirical Analysis of Welfare Implications," *mimeo*, 2004.

[12] Gourieroux, Christian, A. Monfort, and A. Trognon, "Pseudo Maximum Likelihood Methods: Theory," *Econometrica* (1984a) 52(3), pp. 681-700.

[13] Gourieroux, Christian, A. Monfort, and A. Trognon, "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica* (1984b) 52(3), pp. 701-20.

[14] J. Hausman, B. Hall, and Z. Griliches, "Econometric Models for Count Data with an Application to the Patents-R&D Relationship," *Econometrica* (1984) 52, pp. 909-938.

[15] Hall, B. and R. Ziedonis, "The Patent Paradox Revisited: An Empirical Study of Patenting in the US Semiconductor Industry, 1979-1995,"*RAND Journal of Economics* (2001) 32(1), pp. 101-128.

[16] Huber, P., "The Behavior of Maximum Likelihood Estimates under Non-standard Conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (1967) 1, pp. 221-233.

[17] Narasimhan, Chakravarthi, "Competitive Promotional Strategies." *Journal of Business*, (1988) 61, pp. 427-449.

[18] Rogers, W. "Regression Standard Errors in Clustered Samples," *Stata Technical Bulletin* (1993) 13, pp. 19-23.

[19] Rosenthal, Robert W., "A Model in Which an Increase in the Number of Sellers Leads to a Higher Price," *Econometrica* (1980) 48(6), pp. 1575-1580.

[20] Shilony, Yuval, "Mixed Pricing in Oligopoly," *Journal of Economic Theory* (1977) 14, pp. 373-388.

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

[22] White, H. "A Heteroskedasticty-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity," *Econometrica* (1980) 48, pp. 817-830.

[23] White, H., "Maximum Likelihood Estimation for Misspecified Models," *Econometrica* (1982) 50, pp. 1-25.

[24] Wooldridge, J., "Quasi-likelihood Methods for Count Data," in *Handbook of Applied Econometrics, Vol. 2* (M. Pesaran and P. Schmidt, eds.) Oxford: Blackwell (1997), pp. 352-406.

Figure 1: Kelkoo Screenshot

Product	Shop	Price	Total Price inc P & P	More
HP IPAQ H5550 COLOUR POCKET PC	PC WORLD Shop info	£429.98 + P & P: £3.95	£433.93	More
HP iPAQ H5550 Pocket PC	[illegible] Shop info	£417.95 + P & P: Free	£417.95	More
HP iPAQ Pocket PC h5550 - RAM: 128 MB - ROM: 48 MB - Windows Mobile 2003 - display 3.8" TFT	dabs.com Shop info	£464.12 + P & P: £5.88	£470.00	More
FA107A HP iPAQ H5550	BigGreyCat Shop info	£486.45 + P & P: £5.20	£491.65	More
IPAQ 5550 INTEL 400MHZ 48MB ROM 128MB 3.8 TFT 802.1X SDIO BIOME	SiMPLY computers Shop info	£504.00 + P & P: £5.29	£509.29	More
HP iPAQ h5550 Pocket PC (English)	expansys Shop info	£382.71 + P & P: £6.98	£389.69	More
iPAQ H5550 400MHz Bluetooth FA107A#ABU	Technoworld Shop info	£491.83 + P & P: £8.17	£500.00	More
iPAQ H5550/EN 400MHz 48MB128MB Bluetooth	easyetrader Shop info	£551.63 + P & P: Free	£551.63	More
HEWLETT PACKARD IPAQ H5550	COMET comet.co.uk Shop info	£429.98 + P & P: Free	£429.98	More
HP iPAQ H5550 Pocket PC (English)	amazon.co.uk Shop info	£429.99 + P & P: Free	£429.99	More
Hewlett Packard iPAQ 5550 Pocket PC	C&A Electronics Shop info	£424.99 + P & P: £6.99	£431.98	More
HP iPAQ H5550 New	totalpda Shop info	£445.33 + P & P: £17.63	£462.95	More

Figure 2: Average Number of Leads by Day of Week

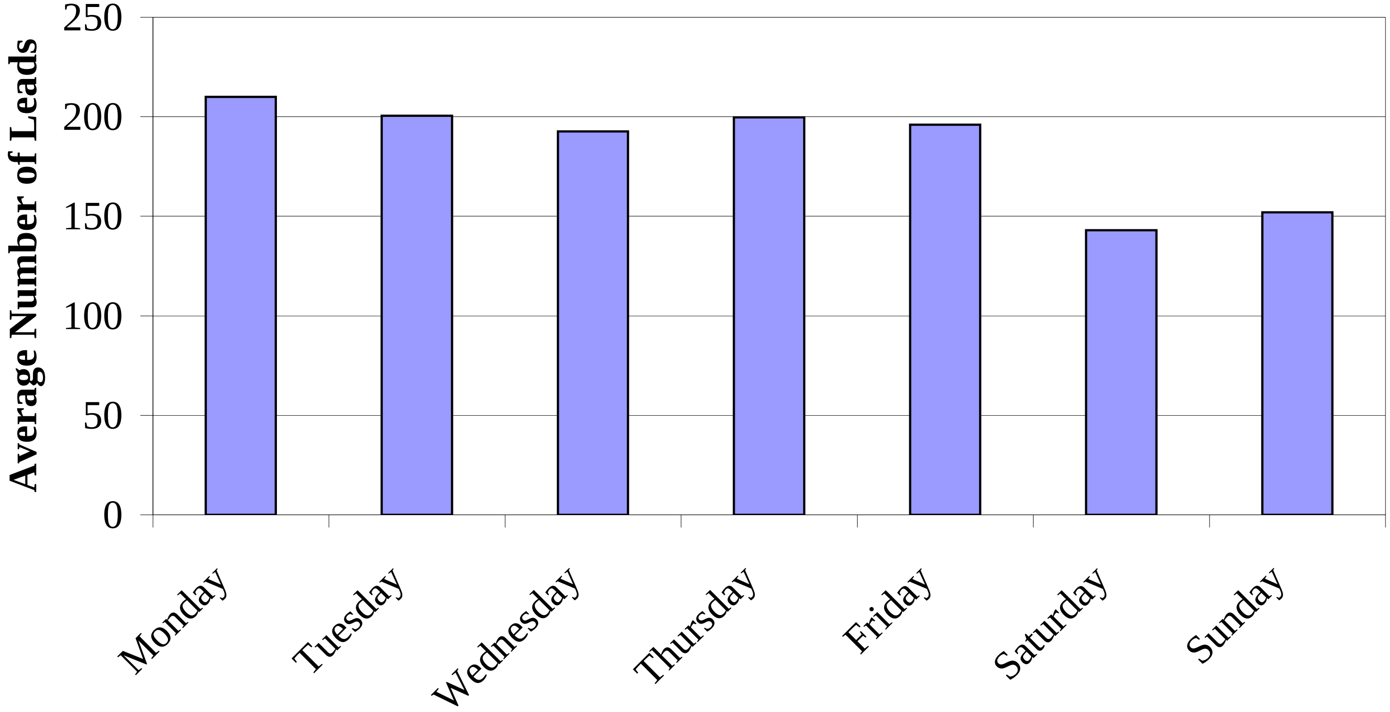

Figure 3: Histogram of Leads by Price Rank and Screen Location

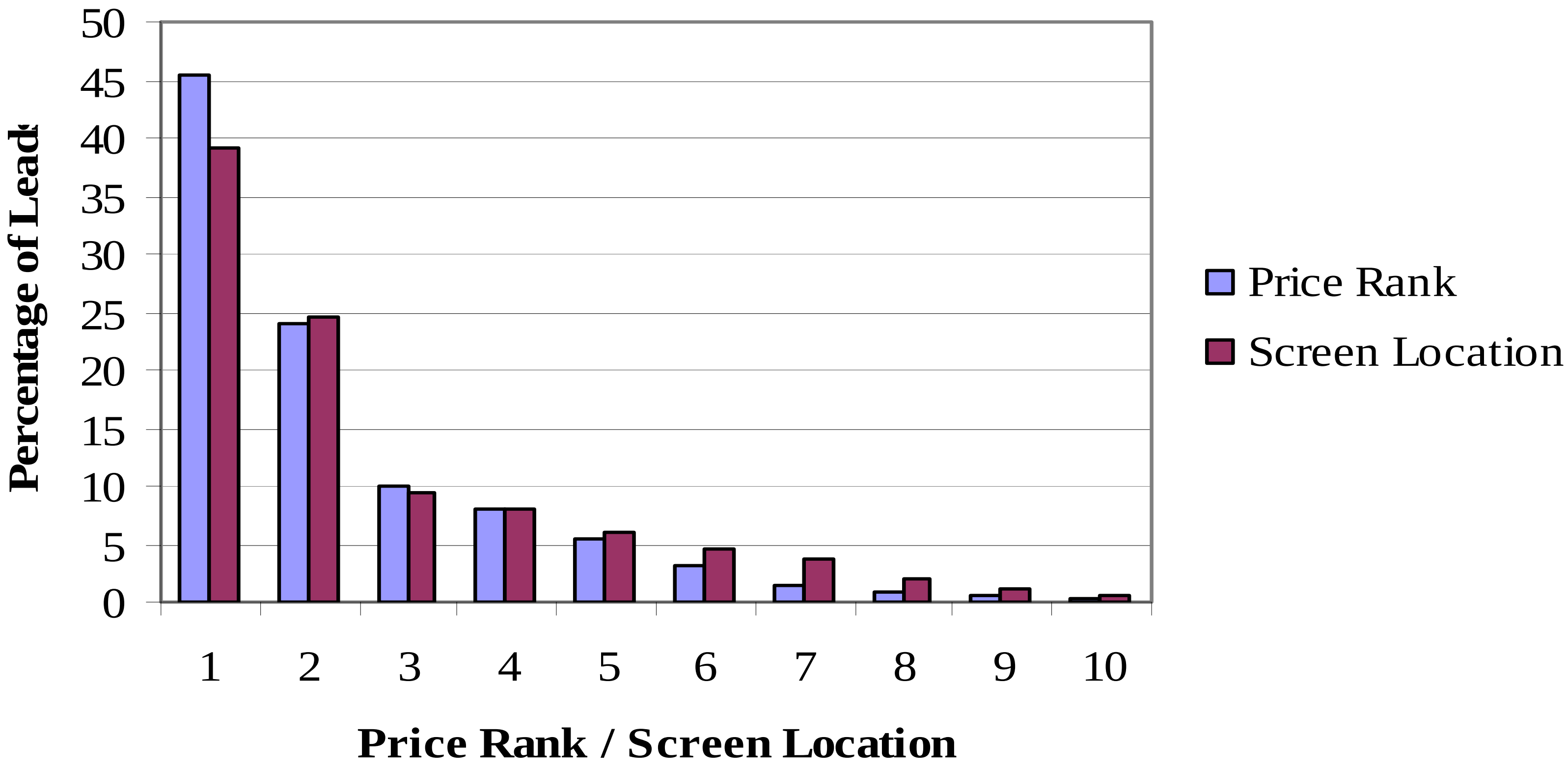

Figure 4: Factors Influencing Numbers of Leads

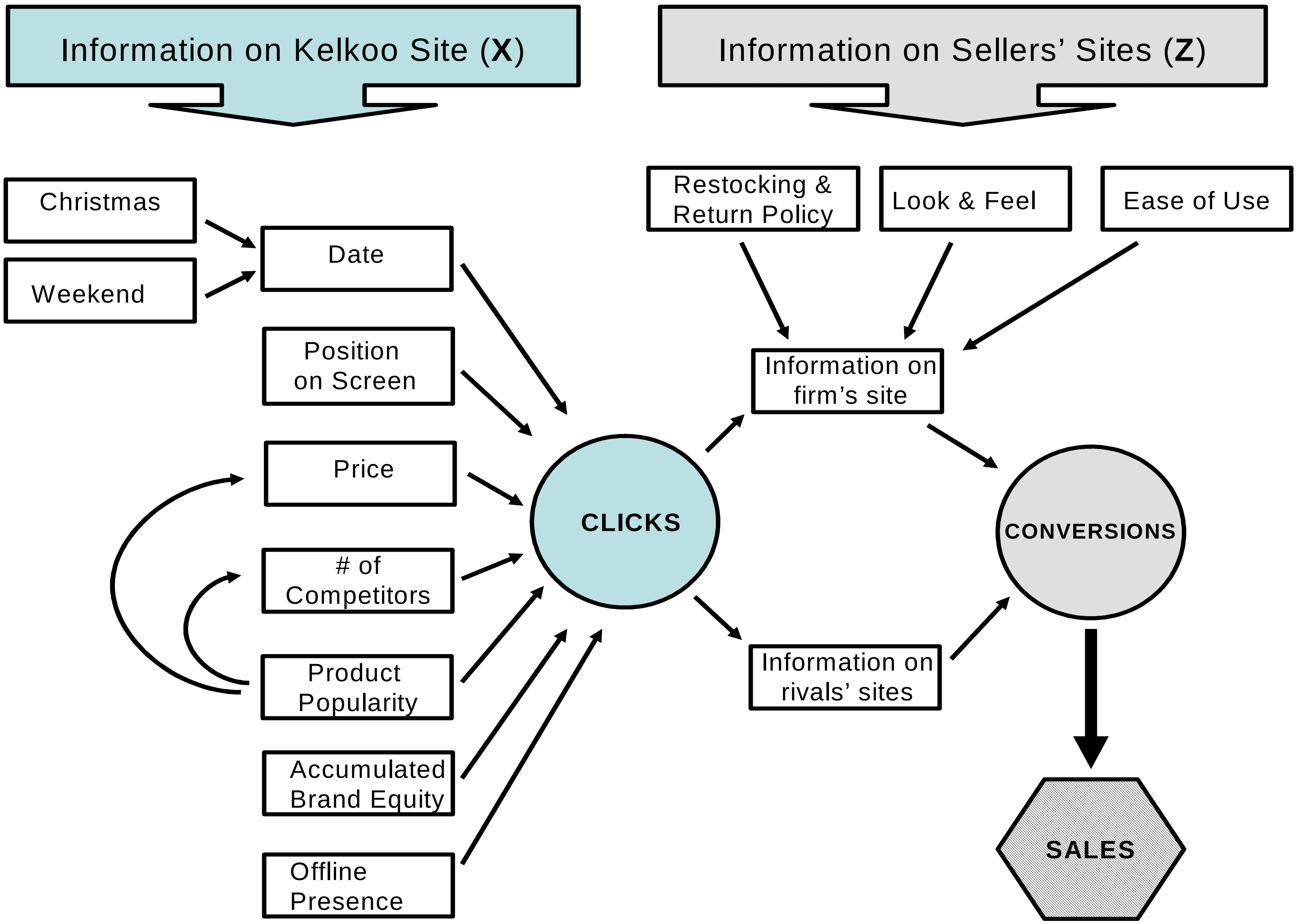

Figure 5: Estimated Demand Elasticity and Number of Sellers

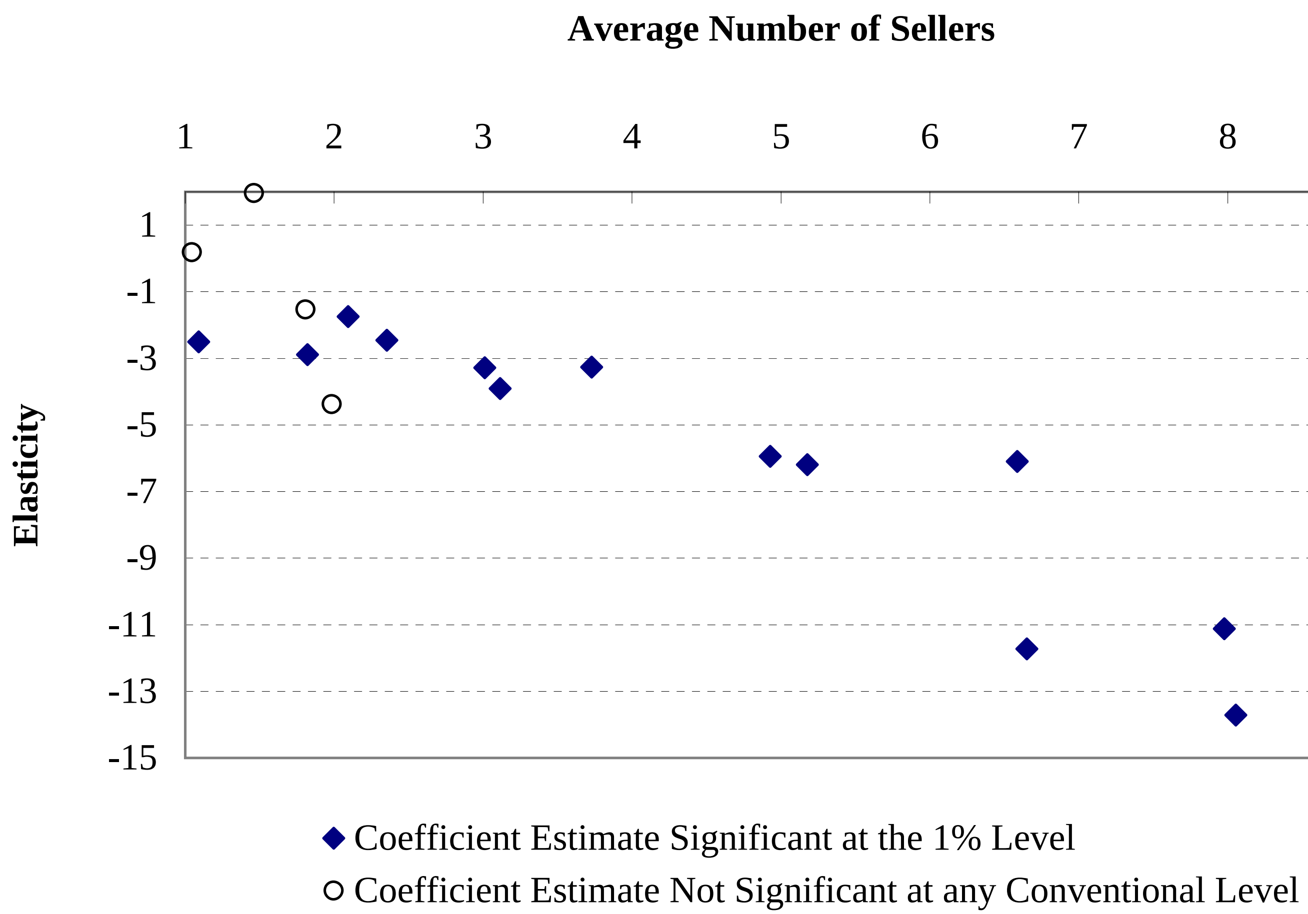

Figure 6: Misspecification from using Continuous Demand Model in Split Market Setting

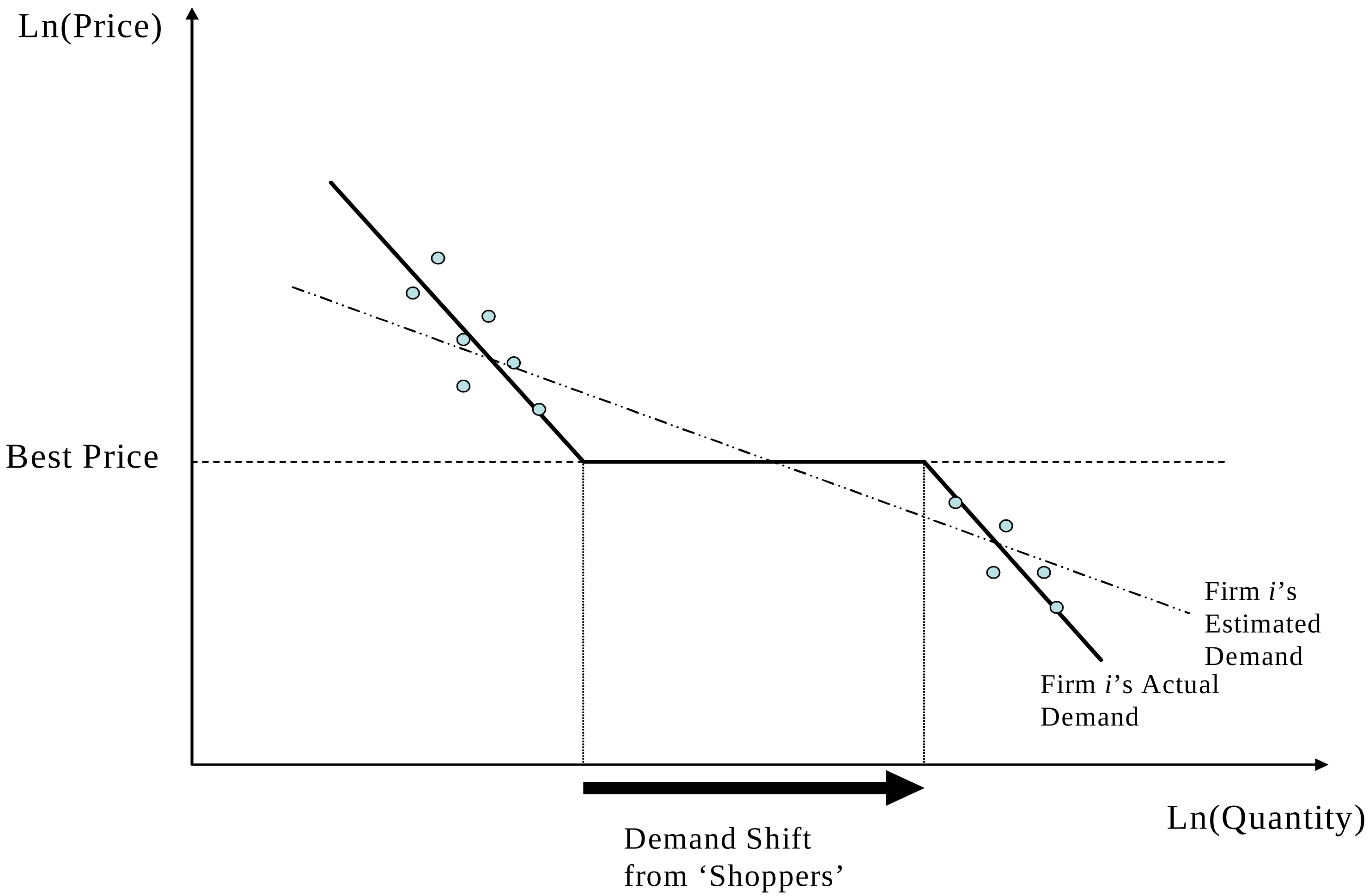

Figure 7: Actual versus Predicted Clicks Frequency

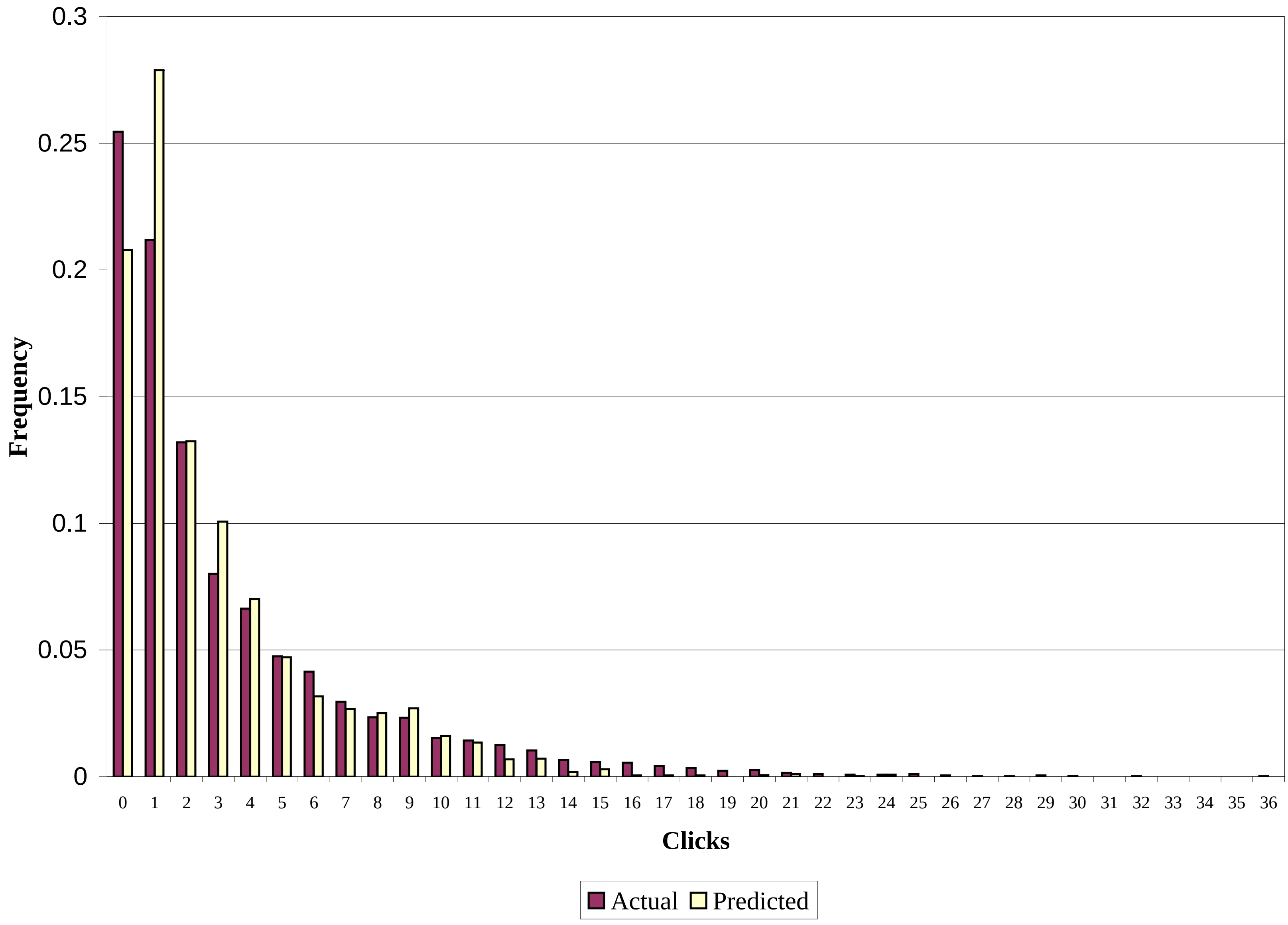

Table 1: Descriptive Statistics

Variable	Mean	Standard Deviation	First Quartile	Median	Third Quartile	Maximum	Minimum
Clicks	3.33	4.27	0	2	5	36	0
Price	304.88	106.84	229.99	279.98	396.63	601.95	104.57
Shipping	4.16	4.50	0	3.95	5.82	17.63	0
Total Price	309.04	107.01	234.42	283.94	396.63	607.77	108.10
Number of Sellers	4.05	2.93	2	3	6	15	1
Location on Screen	3.40	2.43	1	3	5	15	1
Bricks and Clicks Retailer	0.29						
Weekend	0.28						
September	0.11						
October	0.29						
November	0.29						
December	0.27						
January	0.05						
Total Number of Products	18						
Total Number of Firms	19						
Total Number of Dates	111						
Total Number of Observations	6151						

Table 2: Product Specific Demand Estimates

Product	Log Total Price	Position on Screen	Weekend	Month Dummies	# of Obs.	Average Number of Sellers
	Likelihood Specification for Clicks: Poisson PML					
Toshiba E740 WIFI	-1.75	0.272	-0.214	4	216	2.093
	(8.64)**	(3.23)**	(2.35)*			
HP Compaq IPAQ 1910	-3.281	-0.591	-0.215	2	171	3.012
	(5.68)**	(4.73)**	(2.29)*			
HP Compaq IPAQ 1940	-14.691	-0.165	-0.255	4	898	8.942
	(20.39)**	(13.98)**	(4.45)**			
HP Compaq IPAQ 2210	-11.725	-0.058	-0.251	1	184	6.652
	(10.54)**	(2.04)*	(2.43)*			
HP Compaq IPAQ 3950	1.961	-0.351	-0.152	3	91	1.462
	(1.56)	(1.02)	(0.62)			
HP Compaq IPAQ 3970	-1.53	-0.262	-0.12	4	131	1.809
	(1.91)	(3.10)**	(1.14)			
HP Compaq IPAQ 5550	-13.712	-0.153	-0.288	4	851	8.055
	(22.97)**	(13.92)**	(5.17)**			
Palm m515	-2.503	-0.458	-0.444	2	44	1.091
	(3.88)**	(0.99)	(2.82)**			
Sony Clie NX70V	-2.455	-0.227	-0.116	3	164	2.354
	(9.41)**	(2.99)**	(0.91)			
Sony Clie NX73V	-5.941	-0.258	-0.163	4	501	4.928
	(10.82)**	(7.18)**	(1.73)			
Sony Clie NZ90	-2.884	-0.144	-0.331	4	151	1.821
	(1.51)	(0.82)	(1.60)			
Sony Clie SJ22	-3.263	-0.085	-0.278	4	368	3.728
	(8.65)**	(3.04)**	(3.54)**			
Sony Clie SJ33	0.182		-0.215	2	44	1.045
	(0.08)		(1.51)			
Sony Clie TG50	-6.188	-0.049	-0.202	4	428	5.178
	(6.28)**	(1.22)	(1.87)			
Handspring Treo 90	-4.375	-0.723	-0.225	2	136	1.985
	(1.67)	(0.79)	(2.92)**			
Palm Tungsten T2	-6.096	-0.153	-0.265	4	678	6.587
	(11.90)**	(6.30)**	(3.04)**			
Palm Tungsten W	-3.902	-0.328	-0.406	4	295	3.115
	(4.37)**	(4.08)**	(2.30)*			
Palm Zire 71	-11.115	-0.157	-0.316	4	800	7.978
	(11.47)**	(7.71)**	(3.65)**			

Note: Robust z statistics in parentheses. * Significant at 5%. ** Significant at 1%.

Table 3: Continuous Demand Specifications

		Model 1	Model 2
Likelihood Specification for Clicks		Poisson PML	Poisson PML
Log Total Price		-4.61	-3.761
		(8.91)**	(7.45)**
Log Total Price x (Number of Sellers − 1)			-0.288
			(4.14)**
Number of Sellers			1.593
			(4.05)**
Position on Screen		-0.186	-0.175
		(4.54)**	(4.47)**
Bricks and Clicks Retailer		0.262	0.236
		(1.58)	(1.67)
Weekend		-0.242	-0.265
		(10.82)**	(11.46)**
Product Dummies		17	17
Month Dummies		4	4
Product x Month Dummies		55	55
Robust Standard Errors Clustered by Firm		Yes	Yes
Observations		6151	6151
Overdispersion Test	Test Statistic	2656.46	2488.77
	P-Value	0	0

Note: Robust z statistics in parentheses. * Significant at 5%; ** Significant at 1%

Table 4: Discontinuous Demand Specifications

		Model 1	Model 2	Model 3	Model 4
Likelihood Specification for Clicks		Poisson PML	Poisson PML	Poisson CML	Poisson CML
Log Total Price		-2.459	-2.386	-2.446	-2.449
		(9.11)**	(9.64)**	(23.78)**	(23.75)**
Log Total Price x (Number of Sellers − 1)		-0.252	-0.289	-0.175	-0.173
		(4.60)**	(5.44)**	(9.94)**	(9.83)**
Demand Shift from Shoppers		0.603	0.591	0.599	0.6
		(7.11)**	(6.61)**	(26.60)**	(26.55)**
Number of Sellers		1.415	1.614	0.98	0.97
		(4.52)**	(5.31)**	(9.93)**	(9.82)**
Position on Screen		-0.175	-0.174	-0.149	-0.149
		(4.37)**	(4.46)**	(21.61)**	(21.32)**
Bricks and Clicks Retailer		0.321	0.317	0.367	
		(2.41)*	(2.43)*	(-1.87)	
Weekend		-0.268	-0.272	-0.263	-0.263
		(13.79)**	(14.24)**	(15.44)**	(15.43)**
Product Dummies		17	17	17	17
Month Dummies		4	4	4	4
Product x Month Dummies		55	55	55	55
US Product Rank Dummies		No	52	52	52
Robust Standard Errors Clustered by Firm		Yes	Yes	No	No
Controls for Unobserved Firm Heterogeneity		No	No	19	19
				Random Effects	Fixed Effects
Observations		6151	6151	6151	6151
Overdispersion Test	Test Statistic	1942.27	1822.4	NA	NA
	P-Value	0	0		

Note: *z* statistics in parentheses. * Significant at 5%; ** Significant at 1%

Table A1: Continuous Demand - Alternative Specifications

	Model 1	Model 2
Likelihood Specification for Clicks	Negative Binomial ML	Negative Binomial ML
Log Total Price	-4.81	-3.696
	(10.29)**	(8.66)**
Log Total Price x (Number of Sellers − 1)		-0.343
		(5.54)**
Number of Sellers		1.897
		(5.37)**
Position on Screen	-0.178	-0.166
	(4.70)**	(4.46)**
Bricks and Clicks Retailer	0.316	0.272
	(2.26)*	(2.23)*
Weekend	-0.263	-0.288
	(11.62)**	(13.42)**
Product Dummies	17	17
Month Dummies	4	4
Product x Month Dummies	55	55
Robust Standard Errors Clustered by Firm	Yes	Yes
Observations	6151	6151

Note: Robust z statistics in parentheses. * Significant at 5%; ** Significant at 1%

Table A2: Discontinuous Demand - Alternative Specifications

	Model 1	Model 2	Model 3	Model 4
Likelihood Specification for Clicks	Negative Binomial ML	Negative Binomial ML	Negative Binomial CML	Negative Binomial CML
Log Total Price	-2.343	-2.304	-2.334	-2.333
	(8.18)**	(8.89)**	(16.72)**	(16.64)**
Log Total Price x (Number of Sellers − 1)	-0.314	-0.342	-0.168	-0.165
	(5.38)**	(5.95)**	(7.22)**	(7.09)**
Demand Shift from Shoppers	0.619	0.608	0.578	0.579
	(8.24)**	(7.78)**	(18.92)**	(18.84)**
Number of Sellers	1.77	1.912	0.932	0.918
	(5.30)**	(5.82)**	(7.15)**	(7.02)**
Position on Screen	-0.166	-0.165	-0.145	-0.144
	(4.31)**	(4.35)**	(15.93)**	(15.57)**
Bricks and Clicks Retailer	0.324	0.319	0.272	
	(2.66)**	(2.64)**	(2.81)**	
Weekend	-0.29	-0.297	-0.247	-0.247
	(14.86)**	(15.13)**	(10.96)**	(10.95)**
Product Dummies	17	17	17	17
Month Dummies	4	4	4	4
Product x Month Dummies	55	55	55	55
US Product Rank Dummies	No	52	52	52
Robust Standard Errors Clustered by Firm	Yes	Yes	No	No
Controls for Unobserved Firm Heterogeneity	No	No	19	19
			Random Effects	Fixed Effects
Observations	6151	6151	6151	6151

Note: z statistics in parentheses. * Significant at 5%; ** Significant at 1%