View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Research Papers in Economics

An Empirical Model of Search with Vertically Differentiated Products*

Matthijs R. Wildenbeest†

First draft: November 2006

Current version: April 2009

Abstract

This paper presents a non-sequential search model that allows for vertical product differentiation. In the unique symmetric equilibrium firms with different characteristics draw utilities from a common utility distribution, resulting in asymmetric price distributions. The model therefore provides a theoretical rationale for explaining price dispersion as a result of quality differences and search frictions together. More specifically, the model can explain the frequent and asymmetric price changes reported in several empirical papers, but also why some firms have persistently higher prices than others. Using the equilibrium conditions derived from the model, we show how to estimate search costs by maximum likelihood using only prices. The method is applied to a data set of prices for grocery items from supermarkets in the UK. Estimates reveal that most of the observed price variation can be explained by supermarket heterogeneity and that the estimated amount of search is low in this market. We show that ignoring vertical product differentiation results in an overestimation of search costs. Moreover, estimated search costs using a basket of organic items are on average higher than that of a similar non-organic basket. We also simulate how changes in search costs will affect behavior of stores and consumers.

Keywords: consumer search, product differentiation, price dispersion, structural estimation

JEL Classification: C14, D83, L13

*This paper is based on Chapter 6 of my PhD thesis. I would like to thank mySupermarket.co.uk for providing a list of items used to construct their staple basket. In addition, I wish to thank my advisor José Luis Moraga-González for his valuable comments and suggestions. Ambarish Chandra, Eric Rasmusen, Michael Rauh, Mike Waterson, Chris Wilson, and the seminar participants at Indiana University, London School of Economics, Universidad Carlos III de Madrid, Erasmus University, the University of Groningen, and the University of Warwick also provided me with useful remarks. The paper has benefited from presentations at IIOC 2008 (Arlington, VA), the EARIE 2007 Meetings (Valencia), the European Meeting of the Econometric Society 2007 (Budapest), the Royal Economic Society Annual Conference 2007 (Coventry), the European Winter Meeting of the Econometric Society 2006 (Turin), and the Nake Research Day 2006 (Amsterdam).

†Indiana University, Kelley School of Business, E-mail: mwildenb@indiana.edu.

1 Introduction

Price dispersion is an observed feature in many markets. Recent empirical studies have documented substantial price dispersion in markets for grocery products (Lach, 2002), mutual funds (Hortaçsu and Syverson, 2004), online electronics (Baye, Morgan, and Scholten, 2004), and online books (Hong and Shum, 2006).

Figure 1 provides another example of price dispersion by looking at the prices charged for Golden Delicious apples at four major supermarkets in the United Kingdom between September and October 2008. Several factors may have contributed to the relatively large differences in prices across sellers. First, a large theoretical literature on consumer search behavior (e.g., Reinganum, 1979; Burdett and Judd, 1983; Stahl, 1989; Janssen and Moraga-González, 2004) has shown that imperfect information about sellers' prices may lead to equilibrium price dispersion.[1] Second, Golden Delicious apples are particularly prone to bruising and shriveling and supermarkets might differ in service levels offered, so product differentiation offers another explanation for the observed differences in prices.

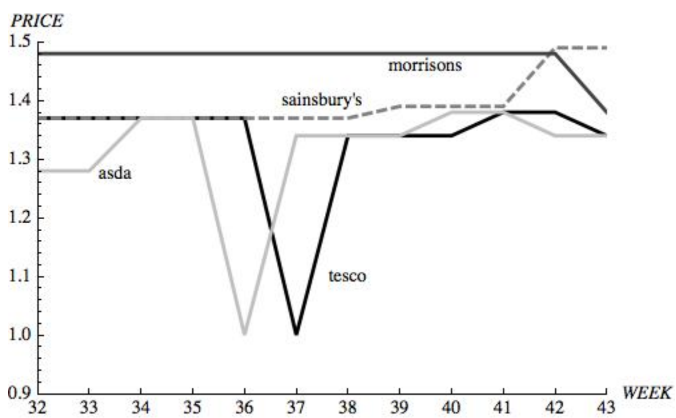


Figure 1: Prices Golden Delicious apples

The substantial variation in prices at a given point in time shown in Figure 1 as well as the persistence of the price dispersion over time are consistent with a product differentiation as well as a search friction explanation. However, Figure 1 also suggests that some supermarkets set on average higher prices for Golden Delicious apples than others, while at the same time three out of four supermarkets change their prices of apples frequently, and – more importantly – not necessarily in similar ways. Although the first observation is consistent with a product differentiation explanation,

[1]For an overview of studies on price dispersion, see Baye, Morgan, and Scholten (2006).

unless the seemingly random changes in prices have gone together with changes in the quality of the apples offered these observed pricing patterns are difficult to explain using the product differentiation explanation alone. On the other hand, while search models can explain random pricing patterns (e.g., Burdett and Judd, 1983; Stahl, 1989), at the same time they cannot explain why some firms have higher average prices than others.

Although several other papers have found evidence for both randomness and persistent differences in average prices (e.g., Lach, 2002; Baye, Morgan, and Scholten, 2004; Lewis, 2007), the current literature on consumer search behavior lacks a theoretical foundation to explain both simultaneously. In this paper we will present a model that combines the two and as such is able to rationalize the observed pricing patterns better than product differentiation models and search models individually. More specifically, we examine the impact of search frictions on the competitiveness of a market for products that are differentiated in terms of quality. We show that in equilibrium firms randomly draw prices from asymmetric price distributions. Average prices are linked to vertical characteristics, which means the search model in this paper can explain random pricing patterns, but unlike other existing search models it also offers an explanation why some firms have persistently higher or lower average prices than other firms. To the best of our knowledge this feature is novel in the literature.[2]

In the second part of the paper we use the structure of the equilibrium model to estimate both search costs and the impact of vertical product differentiation on prices. We show that the model can be estimated by maximum likelihood using only price data. We apply the estimation method to price data from supermarkets in the United Kingdom.[3] The data covers the period between August and October 2008. The estimation results point out that the model does quite well in explaining observed prices of a basket of staple items across the four major supermarket chains in the UK. Our estimates indicate that around 61% of the variation in prices is explained by supermarket heterogeneity, while the rest of the variation is due to search frictions. Besides that we find that the amount of search is relatively low; about 91% of consumers search at most two times. Average price-to-cost margins are estimated to be between 8% and 9%, which seems reasonable

[2] Although Wolinsky (1986) and Anderson and Renault (1999) allow for horizontal product differentiation, their models do not generate price dispersion.

[3] Several studies have recently looked at competition in the UK supermarket industry. Smith (2004) estimates a model of consumer choice and expenditure and finds that mergers between the largest firms will lead to price increases of up to 7.4%. Smith (2006) analyzes store location and size using a characteristics utility model estimated with individual consumer data. Although the focus of this paper will be on supermarket choice as well, our study differs from previous ones in the sense that we will mainly concentrate on how supermarket choice relates to consumer search behavior. An advantage of our method is that we only need price data, although this means that unlike previous studies using more detailed data, we are unable to control for horizontal product differentiation.

for this sector.[4] We show that ignoring the vertical product differentiation component leads to an overestimation of search frictions, which can explain the relatively high search cost estimates found by others in the past.

Our data set also includes prices on organic items, which provides us with a natural case to investigate how search costs relate to consumer demographics: organic food purchasers tend to have distinct characteristics and several studies have shown that consumers of organic grocery items have on average higher incomes. We conduct an experiment in which we compare our search cost estimates to estimates obtained using a basket of only organic items and find that organic food purchasers have higher search costs on average. These difference might be explained by De los Santos (2008), who finds a significant negative relationship between income and the time spent searching.

Our paper fits within the recent literature on the structural estimation of consumer search models. Hong and Shum (2006) estimate search cost distributions in a homogenous good setting by maximum empirical likelihood using only price data. Moraga-González and Wildenbeest (2008) show that a maximum likelihood approach can improve on their results. Hortaçsu and Syverson (2004) also estimate search costs in a model of vertical product differentiation. An important difference is that price dispersion in Hortaçsu and Syverson (2004) is the result of firms playing pure strategies, while in our model it is the result of mixed strategies. This means our model is capable of explaining random pricing patterns as we observe in our data. Moreover, in a mixed strategy equilibrium profits need to be the same across firms, which gives an extra condition that can be used for the estimation of the model. The extra condition makes that here only price data is needed to estimate the model, while Hortaçsu and Syverson (2004) need both price and quantity data. This is important since in many settings the econometrician only observes prices.

In a related empirical paper, Lach (2002) studies existence and persistence of price dispersion using price data of four different products in Israel. Several predictions from search models are tested and he finds the patterns in the price data to be in line with these predictions. Lach (2002) controls for differences between firms in a similar way as we do here. In that sense, the analysis presented here shows that vertical product differentiation can be captured in a theoretical model in such a way that Lach's approach is theoretically justified. Moreover, our paper goes one step further by using the structure of the theoretical search model to estimate the underlying search cost distribution.

In terms of policy implications, Armstrong (2008) argues that especially when there are in-

[4]Smith (2004) reports gross margins that are between 11% and 14% for the year 2000.

formation frictions competition policy may occasionally harm some consumers. Indeed, we find evidence in our data that a policy to encourage people to visit price comparison sites will hurt most grocery shoppers. Using the estimated search cost distribution as a starting point, we find that increasing the share of consumers with very low search costs results in higher prices being charged by the supermarkets. More consumers with very low search costs makes it more costly for firms to set very low prices. Stores will start focusing more on the consumers with relatively high search costs, with higher prices and overall higher profit margins for the stores as a result.

The structure of this paper is as follows. In the next section we discuss the theoretical model. Section 3 continues with a method to estimate search costs using maximum likelihood. In Section 4 we apply the estimation method to price data from supermarkets. The last section concludes.

2 The model

We study a model of firms offering a differentiated good competing for incompletely informed consumers. On the supply side there are N firms, indexed by j, producing goods at a unit cost r_j. The goods are vertically differentiated, i.e., the goods can be ranked according to their characteristics and consumers all agree in the ranking. The model can be used to address two sources of product differentiation. The goods can be differentiated either because the goods themselves are heterogeneous, for example because the products have different features. It could also be that the good is homogeneous, but that the firms selling the product are differentiated. An example of this would be stores selling the same products but offering different service levels.

On the demand side there is a continuum of consumers demanding at most one good, all deriving the same utility from consumption of good j:

$$u_j(X_j, p_j; \beta) = \beta X_j - p_j + \xi_j, \tag{1}$$

where X_j are the observable characteristics of the good, p_j is the price of firm j's good, ξ_j are (by the econometrician) unobservable characteristics of good j and where the parameter β describes the relation between X_j and u_j. Since the coefficient of price is normalized to -1, utilities are measured in the same unit as prices. In what follows, let $v_j \equiv \beta X_j + \xi_j$; we shall refer to v_j as the valuation for the good produced by firm j. Consumers know their valuation for the good produced at the different firms but prices are only observed after searching. By engaging in costly search the consumers can gain information about the prices of the goods and thus the utilities derived from the goods at a subset of the firms. Consumers are characterized by their search cost c, which is a random draw from the distribution function $G(c)$, with density function $g(c)$. We assume consumers search non-sequentially, i.e., consumers determine before entering the market how many times to search.[5] Consumers then buy the product from the firm in their sample providing the highest utility level.

Firms and consumers play a simultaneous move game. The store characteristics are assumed to be drawn from some distribution and are fixed in the short run. We assume the difference between valuation and unit cost $v_j - r_j$ to be the same across firms, i.e., $x = v_j - r_j$, where x is the maximum possible margin that can be attained by the firms. This means that more favorable characteristics come at a higher cost. Moreover, by restricting v_j and r_j to be related in this way, as we will show below, firms are symmetric in the margin received at each offered utility level. In this way incentives for the firms are identical, which gives rise to the existence of a symmetric equilibrium in utility levels.

Valuations and unit costs are common knowledge. Therefore, an individual firm takes the pricing strategies of the other firms and the search behavior of consumers as given while setting its own price. An individual consumer takes the firm pricing strategies as given and decides on a number k of firms to visit in order to maximize utility. The fraction of consumers sampling k firms is denoted by μ_k.

We focus on symmetric equilibria in utility levels. A condition for a symmetric equilibrium in utility levels to exist is that some consumers should search once, while others should search more than once. The intuition for this is that if all consumers did compare prices, all firms would set a price equal to their unit cost, which implies that all firms would be offering the same utility level x. As a result, there is no reason to search. On the other hand, if no consumer would be willing to compare prices, firms would set their price equal to their valuation, which means that all firms would offer a utility level of zero. Consumers would not participate, because they have to pay a search cost c to enter the market.[6]

A second condition for a symmetric equilibrium in utility levels to exist is that the firms must play mixed strategies in setting their utility level. The proof for this is similar to the proof of Lemma 2 of Moraga-González, Sándor, and Wildenbeest (2008) and can be explained by the idea that offering slightly more utility gives a discrete jump in profits when dealing with consumers that compare utilities. Hence there are no atoms in the utility distribution. On the other hand, at utility levels close to zero only consumers searching once will buy, so in that case offering a lower utility

[5]The way consumers search is similar to the non-sequential search model of Burdett and Judd (1983).
[6]See also Lemma 1 of Moraga-González, Sándor, and Wildenbeest (2008).

increases profits. As a result, firms draw utilities from a common atomless utility distribution, which we denote $L(u)$, with a lower bound equal to zero.

Consumer search behavior should be optimal. This means that for a consumer searching k times, the expected utility should be higher than the expected cost of searching kc. Moreover, the net benefit of searching k times should be higher than the net benefit of searching $k-1$ or $k+1$ times. Now define c_k as the search cost of the consumer indifferent between searching k and $k+1$ times. For this consumer $E[\max\{u_1, u_2, \ldots, u_k\}] - kc = E[\max\{u_1, u_2, \ldots, u_{k+1}\}] - (k+1)c$, or

$$c_k = E[\max\{u_1, u_2, \ldots, u_{k+1}\}] - E[\max\{u_1, u_2, \ldots, u_k\}]. \tag{2}$$

The share of consumers who search k times is then given by

$$\mu_k = \int_{c_k}^{c_{k-1}} g(c)dc = G(c_{k-1}) - G(c_k). \tag{3}$$

Now consider optimal firm behavior. Given expected consumer behavior μ_k and expectations on $L(u)$, the profit of firm j offering utility u_j is given by

$$\pi_j(u_j; L(u)) = (x - u_j) \sum_{k=1}^{N} \frac{k\mu_k}{N} L(u_j)^{k-1}.$$

Since $x - u_j = p_j - r_j$, the first part of this equation is the margin the store makes on its product. The second part represents the expected quantities sold, and is explained as the summation over all N consumer groups of the share of consumers searching k times multiplied by the probability that these μ_k consumers visit the firm (which is k/N) and by the probability that a firm selling the product at a utility level of u_j offers the highest utility out of k firms, which is $L(u_j)^{k-1}$.

Given the mixed strategies, in equilibrium a store should be indifferent between setting any utility in the support of $L(u)$. In addition, the lower bound of $L(u)$ should be equal to zero. This is because a firm offering a utility of zero will only sell to the consumers searching once, and surplus extracted from these consumers is maximized by setting $\bar{p}_j = v_j$ so that $\underline{u} = 0$. In this case the profit equation simplifies to $\pi(\underline{u}) = x\mu_1/N$. Setting this equal to the equilibrium profits in general gives the equilibrium condition for this model:

$$(x - u) \sum_{k=1}^{N} \frac{k\mu_k}{N} L(u)^{k-1} = x \cdot \frac{\mu_1}{N}. \tag{4}$$

Unfortunately, this equation cannot be solved for $L(u)$, so the equilibrium distribution of utilities is only implicitly defined. Solving equation (4) for u gives

$$u = x \cdot \frac{\sum_{k=2}^{N} k\mu_k L(u)^{k-1}}{\sum_{k=1}^{N} k\mu_k L(u)^{k-1}}. \tag{5}$$

Although the utility distribution is the same for each firm, since $u = v_j - p_j$, the price distribution is different across firms:

$$F_j(p) = \Pr[p_j \leq p] = \Pr[p \geq v_j - u_j] = \Pr[u_j \geq v_j - p] = 1 - L(v_j - p).$$

The maximum utility in the market can be found by setting $L(u) = 1$, which gives

$$\bar{u} = x \cdot \frac{\sum_{k=2}^{N} k\mu_k}{\sum_{k=1}^{N} k\mu_k}. \tag{6}$$

Individual firms choose a utility level to maximize expected profits given expected search behavior of the consumers and given the expected utility distribution function, so in equilibrium the first order condition with respect to u should be zero, i.e.,

$$\frac{\partial \pi}{\partial u} = \sum_{k=1}^{N} \frac{k\mu_k}{N} L(u)^{k-1} - (x-u) \sum_{k=1}^{N} \frac{k(k-1)\mu_k}{N} L(u)^{k-2} l(u) = 0.$$

Solving this expression for $l(u)$ gives the density function of utility

$$l(u) = \frac{\sum_{k=1}^{N} k\mu_k L(u)^{k-1}}{(x-u)\sum_{k=1}^{N} k(k-1)\mu_k L(u)^{k-2}}. \tag{7}$$

Using the characterization of the utility distribution equation (2) can be rewritten as a function of the utility distribution:

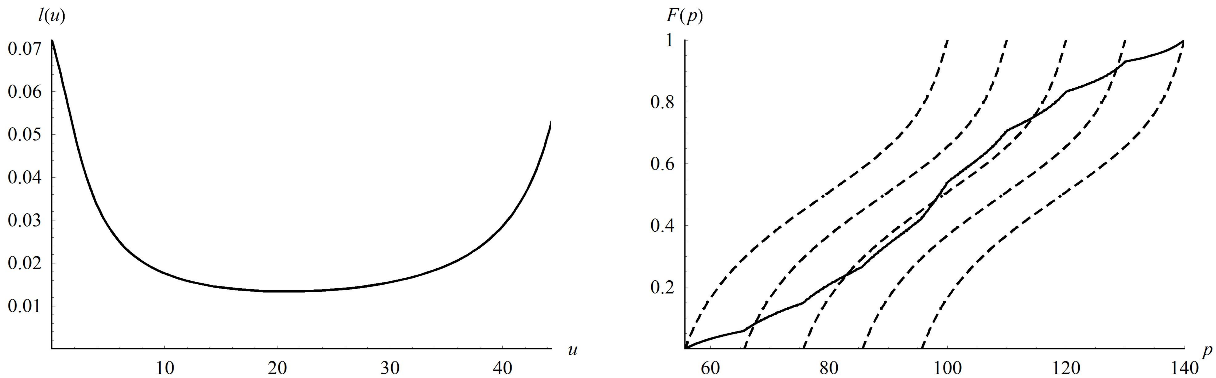
$$c_k = \int_{\underline{u}}^{\bar{u}} (k+1) u L(u)^k l(u) du - \int_{\underline{u}}^{\bar{u}} k u L(u)^{k-1} l(u) du.$$

By using the change of variable $y = L(u)$, we obtain $dy = l(u)du$. Plugging this into the equation above, transforming the lower limit into $y = L(\underline{u}) = 0$ and the upper limit into $y = L(\bar{u}) = 1$ and solving gives

$$c_k = \int_0^1 u(y)[(k+1)y - k] y^{k-1} dy. \tag{8}$$

Then using the same change of variable in equation (5) we can get rid of $u(y)$ in equation (8).

As an example, we calculate equilibrium when consumers search costs are drawn from a log-



(a) Utility pdf

(b) Price cdfs

Figure 2: Example equilibrium search model

normal distribution with parameters 0.5 and 5. Figure 2 gives plots of the equilibrium for 5 firms with valuations ranging from 100 to 140 and $x = 50$ so that marginal cost range from 50 to 90. In Figure 2(a) the equilibrium utility density is plotted. Most mass is at the extremes of the distribution, with slightly more mass at lower utilities than at higher utilities. This shows the tradeoff firms face: set a high utility to attract consumers who compare several offerings or set a low utility in order to maximize surplus extracted from consumers who do not search. In Figure 2(b) the equilibrium price cdfs are drawn; the dashed lines are the firms' individual price cdfs and the solid line is the price cdf for all the firms together. What is interesting to note is that the shape of the individual price cdfs is quite different from the shape of the price cdf of all firms together. This means that assuming all firms are selling the same homogeneous product when in fact they are not might likely lead to wrong estimates of the underlying search cost distribution. We will come back to this issue in the empirical section.

Note that price dispersion in Hortaçsu and Syverson (2004) is a result of firms playing pure strategies, while in the model presented here it is a result of mixed strategies. In a mixed strategy equilibrium profits need to be the same across firms, which gives an extra condition that can be used for the estimation of the model. As we will see in the next section, this extra condition makes that here only price data is needed to estimate the model, while Hortaçsu and Syverson (2004) need both price and quantity data.

3 Estimation

In this section we present a method to estimate the model presented in the previous section using only price data. Assume the prices N firms charge for the same good are observed for a certain period of time, the latter being indicated by the subscript t. There are two methods to calculate utilities from observed prices. In the first method v_j can be (superconsistently) estimated by taking the maximum observed price $\bar{p}_j$ for each firm j during the sampling period. Since $v_j \equiv \beta X_j + \xi_j$ we can rewrite equation (1) as $u_{jt} = v_j - p_{jt} = \bar{p}_j - p_{jt}$, so corresponding utilities for all observed prices can be calculated. The second method follows from rewriting the utility function as $p_{jt} = v_j - u_{jt}$. This equation can be estimated by carrying out a fixed effects regression of prices on a constant, i.e.,

$$p_{jt} = \alpha + \delta_j + \epsilon_{jt},$$

where α is a constant, δ_j are the firm fixed effects and ϵ_{jt} are the residuals. Note that with this specification, valuations v_j are estimated by $\alpha + \delta_j$ and utilities are calculated by taking the negative of the residuals ϵ_{jt}. Moreover, ϵ_{jt} is simply the price at time t for firm j minus the average price of firm j within the period, which means that $u_{jt} = -\epsilon_{jt} = p_j - p_{jt}$, where p_j is the average price for store j.[7] In both methods utilities are calculated by restricting the shape of the price distribution to be the same across firms (although they might have different means), but instead of using the maximum observed prices across firms the second method uses the average observed prices across firms to serve as a proxy for differences in valuations. Although the first method will give superconsistent estimates of the valuations, it is very sensitive to outliers, so we will follow the second method in what follows.

Notice that the proposed method bares resemblance with how heterogeneity in the structural auction literature is being dealt with. Haile, Hong, and Shum (2003) provide a test for common values in first-price sealed-bid auctions and show that under certain conditions equilibrium bids are additively separable into a common auction-specific component and an idiosyncratic component (see also Hu and Shum, 2008; and Bajari, Houghton, and Tadelis, 2007). The auction specific component is assumed to be function of observed auction characteristics, which, as shown by Haile *et al.* (2003), implies the residuals of a regression of observed bids on the covariates can be treated as normalized bids. While this method effectively deals with observed heterogeneity, our method is different in the sense that it deals with unobserved heterogeneity as well.

The estimated utilities allow us to proceed as in Moraga-González and Wildenbeest (2008) to estimate the parameters of the model. The density function in equation (7) can be used to estimate the search cost distribution by maximum likelihood. Since all firms are assumed to draw utilities

[7]For the analysis the height of the utilities is not important, all what matters are the differences between the utilities. This means that our estimate of the search cost distribution does not change when the same constant is added to all valuations v_j.

from the same distribution, all the estimated utilities can simply be pooled. The log-likelihood function is then $LL = \sum_{i=1}^{M} \log l(u_i)$, where M is the total number of observations. The likelihood function can be concentrated by solving the calculated upper bound of the utility distribution in equation (6) for x as a function of the rest of the parameters, i.e.,

$$x = \bar{u} \cdot \frac{\sum_{k=1}^{N} k\mu_k}{\sum_{k=2}^{N} k\mu_k},$$

and by plugging this into equation (7). In addition all μ_k's have to add up to one, which can be used to get rid of μ_N. The likelihood function is then maximized with respect to the remaining parameters of the model, i.e., μ_k, $k = 1, 2, \ldots, N-1$.

Standard errors of the μ_k's are calculated by taking the square root of the diagonal entries of the inverse of the negative Hessian matrix evaluated at the optimum, while in order to calculate standard errors of the maximum possible margin x and the critical search cost values c_k the Delta method can be used.

4 Empirical analysis

As an illustration of the model and the estimation procedure, in this section we apply the estimation method to prices collected between August and October 2008 from supermarkets in the United Kingdom. The supermarket sector is typically a sector in which vertical product differentiation plays an important role. A consumer survey carried out by the UK's Competition Commission in 2000 finds that the most important determinants of store choice are, apart from the prices charged for the groceries, whether it is possible to do the weekly shopping under one roof, whether the store is within easy and convenient reach of home, product availability, the availability of sufficient car parking space, and flexible opening hours. Favorable characteristics increase the utility level of the typical visitor of a supermarket, but also come at a cost. Full-service supermarkets, focussing on quality, are in general more expensive than for example discounters, whose primarily focus is on low prices and not on service.

The application of our search model to supermarkets might need some additional justification. Search models have so far been structurally estimated using data from the mutual fund industry (Hortaçsu and Syverson, 2004), online book stores (Hong and Shum, 2006), and online stores selling memory chips (Moraga-González and Wildenbeest, 2008). All three markets have in common that the physical location of the firm or store selling the product is of lesser importance. Conventional food retailers however usually tend to operate in the offline world, which means that physical locations are in fact important. Although this implies that horizontal product differentiation issues might be relevant, allowing for horizontal product differentiation in addition to vertical product differentiation would complicate the analysis too much. We therefore ignore horizontal characteristics, so location should be interpreted as a vertical characteristic, which can be justified by the idea that some supermarkets in general have better locations than others.

Another assumption made in the model is that consumers search non-sequentially for the highest utility around. Non-sequential search implies that consumers determine before they start searching how many times to search.[8] To justify this assumption one could think of consumers using advertisements in for example newspapers to collect information about prices at different supermarkets, the use of price comparison sites on the Internet, or a situation where there are a lot of shops at the same distant place in town. Moreover, people typically dedicate one trip to purchase the bulk of their grocery needs (so-called *primary* shopping), while the remainder of shopping trips are used to complement the main trip (so-called *secondary* shopping). The supermarket picked for the secondary shopping is not necessarily the same as the one chosen for the primary shopping so consumers might use price information obtained during secondary shopping trips to determine where to do their primary grocery shopping – a situation which very much resembles non-sequential search.

In the analysis we do not explicitly take advertising into account. Through advertising consumers essentially get some price information for free, so ignoring advertising puts a lower bound on the estimated search costs. On the other hand, as will be explained in more detail below, our focus will be on a basket of goods instead of on individual items, so ignoring advertising could be justified on the basis of the argument that consumers are not so much interested in the prices of only a few advertised products, but only in the price of a basket of grocery products. Our focus on a basket of goods also helps to justify the inelastic demand assumption, since as long as the basket is in line with average weekly shopping expenditures, the usual buyer is expected to buy a single basket at a time.

The model is applied to a data set of prices that are collected over time, so the implicit assumption is that supermarkets play a stationary repeated game of finite horizon. This means that we are ignoring dynamic effects caused by for example loyalty cards, advertising, and switching costs. However, since part of the share of consumers searching only one time can also be interpreted as consumers being loyal to some supermarket, to some extent loyalty can also be accommodated in

[8]See Morgan and Manning (1985) for the optimality of sequential versus non-sequential search.

the current setting.

The focus of this study will be on relatively homogeneous goods, but I allow for the possibility that supermarkets are differentiated in terms of the service they offer. Most theory models explain price dispersion by either random pricing strategies of homogeneous firms or by pure strategies of heterogeneous firms. The model described in Section 2 combines the two: heterogeneous firms mix over price distributions with different support. As is shown below in some detail, in the data set average prices across stores are persistently different over time, but at the same time, stores randomize their prices. These observations make the model presented here a suitable theoretical framework to study price setting behavior of supermarkets in relation to search behavior of consumers, as traditional search models cannot explain both things at the same time.

The setup of the empirical analysis is as follows. In the next subsection, we start by giving a description of the data set. We check some of the implications of the model, like random pricing. Next, we estimate the model structurally. We compare estimates obtained using our basket of staple items with those of a similar basket that consists of only organic items. Finally, we study what happens to equilibrium pricing and searching when there is an exogenous shift in search costs.

4.1 Description of the data

The data is collected using Tesco Price Check, a price comparison tool put by Tesco on its website.[9] In addition to posting its own prices, each week Tesco collects over 10,000 prices in two branches of each of Sainsbury's and ASDA and three branches of Morrisons around the United Kingdom. Tesco, Sainsbury's, ASDA, and Morrisons are often called the big four; together they shared about 65% of the market in 2007. Tesco is the biggest in terms of grocery sales, followed by ASDA, Sainsbury's, and Morrisons. The survey Tesco uses for collecting data covers only superstores. All four have adopted a national pricing policy.[10]

Our data set covers a period of twelve weeks from September till October 2008. The data set consists each week of around 14,000 products. Because the purpose of the price comparison tool is to compare prices of Tesco with those of the other three supermarkets, all products in the data set are carried by Tesco and at least one of the three other supermarkets. In our analysis we focus on the primary shopping trip. According the Competition Commission's 2000 consumer survey around 70% of households do their main grocery shopping just once a week, which means that the majority of consumers is probably most interested in the total price of their primary shopping

[9] See http://www.tesco.com/todayattesco/pricecheck.shtml.

[10] This excludes smaller stores such as Tesco Metro and Express and Sainsbury's Local and Central.

basket and not so much in the prices of individual items. The focus will therefore be on a basket of regularly bought items. Another reason to focus on a basket instead of on individual items is that a supermarket is a multi-product firm so a single-product model as described in Section 2 is probably not the right model when investigating individual products. A drawback of looking at baskets of products instead of individual products is that behavior of consumers who go to different supermarkets for each different product is not captured. Disregarding these consumers can be justified by the survey evidence that the majority of people do their main grocery shopping just once a week.

Price differences for the basket across supermarkets allow us to identify the vertical production differentiation component, so it is important that the products included in our basket are carried by all four food retailers. In addition, to be able to identify search costs we need to observe prices for the items in our basket over time. Furthermore, we only include food and non-alcoholic beverage items, as classified by the most recent list of representative items that the Office of National Statistics uses to construct the CPI (see Wingfield and Gooding, 2008). This leaves us with more than a thousand products from which we can construct our shopping basket.

Including all food and non-alcoholic beverage items for which we have complete price information in the shopping basket would result in an average price for the basket that is more than £1,500, so to get a reasonable estimate of the search cost distribution we have to decrease the size of the basket to more realistic proportions. One problem is that by constraining the size of the basket we increase the number of potential shopping basket that can be constructed out of all available items. About 64% of consumers spend less than £50 on their weekly shopping for groceries at supermarkets (Competition Commission, 2000), and if we want to take this as the goal size of our basket this means that without additional constraints we have to make an arbitrary choice among millions of possible combinations of items.[11] To deal with this we instead use a list of twenty-four items used by comparison website mySupermarket.co.uk to track groceries expenditures.[12] The

[11]One way to deal with this is to randomly selection items out of the pool of all available items and use these to construct the basket. This can be repeated many times, where each randomly constructed basket is considered as one price observation. Although this seems like an intuitive way to deal with the selection issues, a problem is that this approach basically assumes that for a given store the prices of all randomly created baskets are drawn from the same underlying distribution. Since baskets consists of underlying individual items, this requires a level of dependence and coordination in pricing which seems highly unlikely, if possible at all. According to the central limit theorem a lack of dependence will make the price distribution of the randomly constructed baskets converge to a normal distribution. Although these normal distributions will have different means, and as such will give a clear ranking of the supermarket chains in terms of prices which can be used to identify the store heterogeneity aspect of the model, it not useful for estimating search costs. A normal distribution will appear no matter how consumer search, which means the price distribution does not contain any information on consumer search behavior. This identification problem makes it impossible to infer search costs of consumers using only observed prices.

[12]We thank mySupermarket for sharing this list with us.

basket includes items like tea bags, milk, eggs, pasta, minced beef, corn flakes, and rice. These are all staple items and therefore likely to be of great importance for the financial well-being of the food retailers. Moreover, the fact that these items are tracked by mySupermarket.co.uk and picked up by the popular press at regular intervals makes it likely that the supermarkets are especially interested in pricing strategies for these particular items.[13] For some of the products on the mySupermarket.co.uk shopping list we do not have a complete series of prices across all four supermarkets. To deal with this we have replaced those products with similar items so that in the end our basket comes very close to the one used by mySupermarket.co.uk.[14] Table 1 gives an overview of the products selected as well as some sampling statistics. Although prices for a few products have not changed over the sampling period, for most items in the basket there is variation in prices over time and across stores.

Item	Mean Price (Std)	Minimum Price	Maximum Price	Coefficient of Variation (×100)
thick sliced white loaf 800g	0.73 (0.01)	0.72	0.75	1.80
bananas loose	0.78 (0.03)	0.77	0.85	3.65
golden delicious apples class 1 loose	1.38 (0.10)	1.00	1.49	7.17
mixed peppers 3 pack	1.32 (0.13)	1.00	1.38	10.25
cucumber portion	0.34 (0.02)	0.25	0.37	6.22
iceberg lettuce each class 1	0.73 (0.08)	0.37	0.84	10.23
tomatoes 6 pack	0.76 (0.14)	0.50	0.99	18.72
maris piper potatoes 2.5kg pack	1.96 (0.10)	1.48	1.99	5.20
whole milk 3.408ltr/6 pints	2.15 (0.06)	2.12	2.25	2.57
free range eggs medium box of 6	1.36 (0.00)	1.36	1.36	0.00
english butter salted 250g	0.93 (0.02)	0.89	0.94	2.36
cathedral city mild cheddar 400g	3.14 (0.19)	2.66	3.31	6.20
beef mince 500g	2.11 (0.34)	1.00	2.25	16.01
wafer thin smoked ham 500g	3.00 (0.05)	2.97	3.19	1.52
garden peas 142g	0.29 (0.00)	0.29	0.29	0.00
baked bean in tomato sauce 420g	0.34 (0.04)	0.31	0.42	12.97
dolmio original bolognese sauce 500g	1.32 (0.05)	1.00	1.34	3.60
strawberry jam 454g	0.69 (0.06)	0.45	0.79	8.92
silver spoon half spoon sugar 500g	0.92 (0.06)	0.84	0.97	6.11
cornflakes 500g	0.95 (0.00)	0.95	0.95	0.00
fusilli pasta twists 500g	0.75 (0.04)	0.69	0.79	5.94
basmati rice 1kg	1.80 (0.10)	1.74	1.99	5.67
80 teabags 250g	1.29 (0.04)	1.19	1.42	3.07
pure orange juice smooth 1 litre	0.85 (0.09)	0.58	0.88	10.31

Notes: The list is based on the basket of staple items mySupermarket.co.uk uses to track consumer groceries expenditures. Prices are in British pounds. For each item we have 48 observations.

Table 1: Simple statistics for items in the basket

Tesco Pricecheck only compares prices of products at the big four supermarket chains in the UK. Although most consumers will have more than four options for their grocery shopping this

[13] See for example http://news.bbc.co.uk/2/hi/business/7362676.stm.

[14] For a few product-store pairs a small number of weeks is missing. Since it is important to have a balanced panel we have replaced these missing observations by the price of the week before.

is not necessarily a constraint. Our model assumes all supermarket chains play the same mixed strategy in utilities, so price data from one supermarket over time is already enough to estimate search costs. However, we do need information on what supermarkets perceive as the number of competitors, since that is a parameter in the equilibrium utility distribution. Twelve supermarkets had an expected UK market share larger than 0.5% in 2007, which we will take as the number of firms competing.

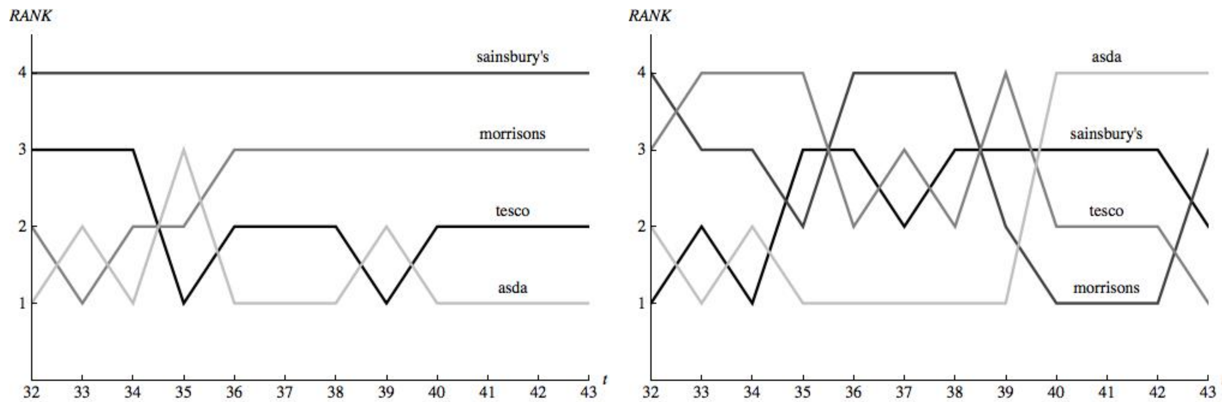
Although not all items in the basket are branded, all items are similar across stores in terms of physical characteristics. Moreover, for generic products like eggs, milk, apples, and cucumbers it is likely that consumers do not care much about the (in store) brand. Nevertheless, since our model explicitly models quality differences between supermarkets, perceived quality differences between in store brands, as well as differences in other store characteristics will show up in our estimates. As a results, supermarkets with better valued characteristics can ask higher prices on average.

4.2 Implications of the model

To investigate whether the right model is used to study search behavior of consumers in this setting, we will first check if some of the implications of the model, like random pricing and persistent differences in average price across stores, are observed in the data. As in Lach (2002), we concentrate on the relative position of the stores in terms of price rankings, and how these positions evolve over time, both before and after correcting for vertical product differentiation.[15]

According to the model prices are randomly drawn from a distribution. Each store will have its own price distribution to draw from and depending on the degree of firm heterogeneity there will be more or less overlap in the supports of these distributions. At one extreme there is no firm heterogeneity, the supports completely overlap, and price rankings are completely random. At the other extreme stores are so much different that the supports do not overlap at all and price rankings do not change. Figure 3(a) shows how the price rankings of the stores evolve over time for the basket. Although Sainsbury's always had the highest prices for the basket and is therefore persistently ranked fourth in terms of prices, there is variability in the rankings of the other supermarkets. Most of the variety comes from Tesco and ASDA as Morrisons is stably

[15] In recent work it has been observed that a typical grocery product is sold at a regular price for a number of time periods, whereas only once in a while the product is sold at a discount price (see, e.g., Pesendorfer, 2002; and Hosken and Reiffen, 2004). An implication of this is that current prices depend on past prices. This seems to be at odds with the implication of the search model in this paper that prices are random draws from some distribution, since this means that prices are not predictable and that there is no correlation over time. However, an important difference between the papers mentioned above and this paper is that while they focus on single products, the focus here is on baskets of goods.



(a) Price rankings (b) Utility rankings

Figure 3: Rankings over time

ranked third in the second half of the sampling period. This suggests that a search model in which consumers place the same value on each supermarket is not appropriate, since in such a model rankings should be completely random. Notice that these findings do not contradict our model, since the model allows for non-fluctuating price rankings. Nevertheless, a clear prediction of the model is that although price rankings may be constant, rankings in terms of utility should be random. Figure 3(b) gives the utility rankings of the stores over time for all items together, where utilities are calculated as in Section 3, i.e, the negative of the residuals of a regression of prices on store dummies.[16] Clearly, this looks much more like a random pattern. This can also be seen in Table 2 where we give the percentage of prices and utilities in each quartile of respectively the price and utility distribution for each of the supermarkets in our data. While in terms of prices the supermarkets tend to spend most time in one quartile only, in terms of utilities it is more spread out.

	prices				utilities			
Quartile	Tesco	Sainsbury's	Morrisons	ASDA	Tesco	Sainsbury's	Morrisons	ASDA
q_1	16.7	0.0	8.3	75.0	16.7	25.0	8.3	50.0
q_2	58.3	0.0	25.0	16.7	58.3	16.7	41.7	16.7
q_3	25.0	0.0	66.7	8.3	25.0	25.0	16.7	0.0
q_4	0.0	100.0	0.0	0.0	0.0	33.3	33.3	41.7

Notes: In percentages.

Table 2: Quartile spent time in

To study this issue in more detail, Table 3 gives information about the time a firm spends in each quartile of every week's price distribution. Stores change their relative position in the price

[16]See Table 4 on p.19 for the regression results.

rankings, but usually not every week. For example, 42% of the price observations in the first quartile were for stores that had a price in this quartile for one successive week. Likewise, 25% of prices in the first quartile belong to stores that were in this quartile for three successive week. Especially the stores that have a price within the third or fourth quartile stay there for many weeks: at least 67% of stores for more than six weeks. Among supermarkets pricing in the first and second quartile there is more fluctuation; prices are on average less than three successive weeks in one of these quartiles and all stores keep their prices in these quartiles for at most four weeks. Table 3 also looks at durations by quartile for utilities. Especially for the higher quartiles the mean duration is lower compared to the same figures for the price distributions. In none of the quartiles supermarkets price more than five subsequent weeks.

	prices				utilities			
duration	q_1	q_2	q_3	q_4	q_1	q_2	q_3	q_4
1 week	41.7	25.0	8.3	0.0	33.3	75.0	25.0	16.7
2 weeks	0.0	16.7	0.0	0.0	0.0	0.0	33.3	0.0
3 weeks	25.0	25.0	25.0	0.0	25.0	25.0	0.0	50.0
4 weeks	33.3	33.3	0.0	0.0	0.0	0.0	0.0	33.3
5 weeks	0.0	0.0	0.0	0.0	41.7	0.0	41.7	0.0
6+ weeks	0.0	0.0	66.7	100.0	0.0	0.0	0.0	0.0
mean	2.5	2.7	6.2	12.0	3.2	1.5	3.0	3.0
median	3	3	8	12	3	1	2	3
max	4	4	8	12	5	3	5	4

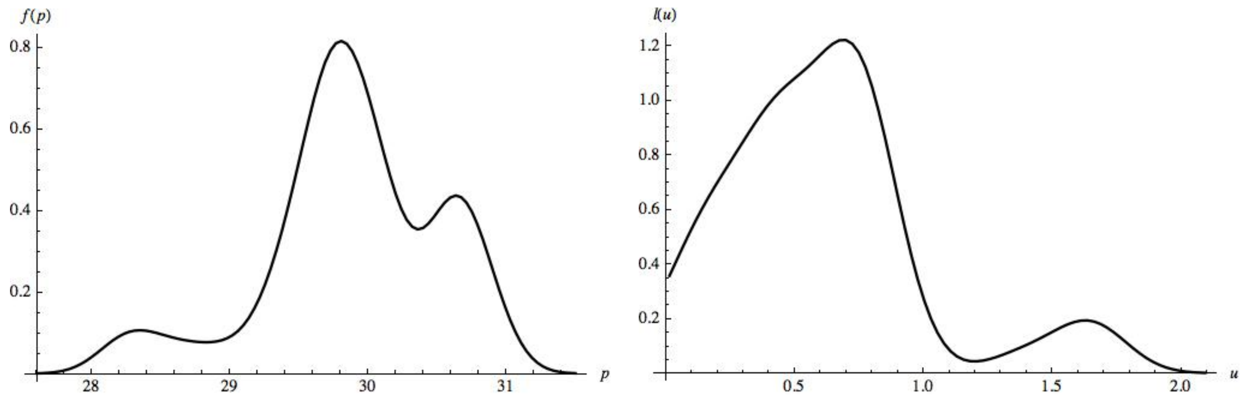
Notes: In percentages.

Table 3: Durations by quartile

4.3 Estimation of search costs

We use the basket of twenty-four regularly bought items to estimate search costs. Figure 4(a) gives a kernel estimate of the price density using prices of the basket of all supermarkets during the twelve week sampling period. According to the search model presented in Section 2 the price dispersion shown in this graph is explained as a combination of quality differences between stores and random pricing strategies. Because of the way utilities are defined in the model, utilities are essentially prices controlled for quality differences between stores. As described in the previous section we use a regression of prices on store dummies to derive utilities.

Table 4 gives the results of the fixed effects regression. Specification (A) has only store dummies included, while specification (B) takes time effects into account as well. In both cases the R^2 is quite high, which means store heterogeneity explains a large part of the variation in the data. To see whether the store fixed effects are jointly significant, an F-test is performed. As can be seen in Table 4, the p-value for the F-test is equal to zero for both specifications, which suggests that store fixed effects indeed matter. In specification (C) we replace the firm dummies with observed firm



(a) Price density (b) Utility density

Figure 4: Price and utility density basket

characteristics like the estimated share of delicatessen in the sales mix, the estimated share of petrol sales in the sales mix, and the average store size.[17] We have picked these variables because the share of delicatessen in the sales mix seems a reasonable proxy for the level of luxuriousness of the chains, and because according to the Competition Commission's 2000 consumer survey, consumers appreciate a large range of grocery products to choose from, as well as extra facilities such as a petrol station. As shown in the table, the regression results indicate that although the average store size is not significantly different from zero, the other two variables can explain a large part of the variation in prices across stores. Moreover, they move in the expected directions.

	(A)	(B)	(C)
Constant	29.88 (0.06)	29.88 (0.06)	26.98 (0.51)
Tesco	-0.30	-0.30	-
Sainsbury's	0.83	0.83	-
Morrisons	-0.06	-0.06	-
ASDA	-0.60	-0.60	-
Share of delicatessen in sales mix	-	-	0.85 (0.17)
Share of petrol in sales mix	-	-	0.07 (0.02)
Average store size (1,000 m^3)	-	-	0.01 (0.07)
N	48	48	48
R^2	0.61	0.75	0.61
Adjusted R^2	0.58	0.65	0.58
p-value F-test	0.00	0.00	0.00

Notes: Standard errors in parenthesis. Estimated specification (A) is with only cross-section fixed effects, specification (B) is with period fixed effects as well. Specification (C) is with observed characteristics only.

Table 4: Regression results

Figure 4(b) gives a kernel estimate of the utility density function, where the utilities are the

[17]The data is taken from Appendix 3.1 of the final report of the Competition Commission's 2008 Groceries Market Investigation (see Competition Commission, 2008).

negatives of the residuals of the fixed effects regression in specification (A). Figure 4(b) shows that there is substantial utility dispersion. The utility density is right-skewed, which tells us that although it is possible to encounter relative high utility levels, it happens with small probability. This gives some indication that the share of consumers searching intensively will not be very large in this market.

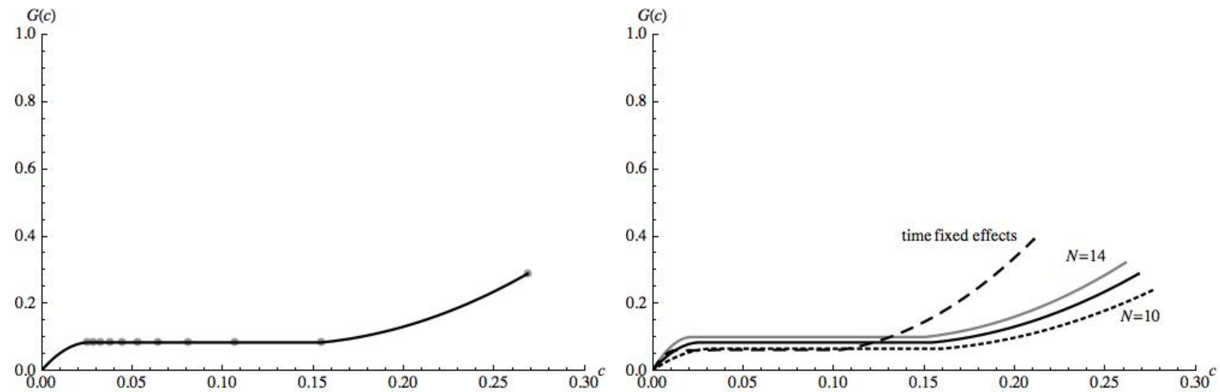
	(1)	(2)	(3)	(4)
N	12	10	14	12
# obs	48	48	48	48
μ_1	0.71 (0.19)	0.76 (0.15)	0.68 (0.20)	0.61 (0.18)
μ_2	0.20 (0.08)	0.18 (0.08)	0.22 (0.07)	0.33 (0.09)
μ_3	0.00	0.00	0.00	0.00
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
μ_{N-1}	0.00	0.00	0.00	0.00
μ_N	0.08 (0.11)	0.06 (0.07)	0.10 (0.13)	0.06 (0.07)
$v_j - r_j$	3.04	3.55	2.77	2.15
LL	16.95	18.55	15.66	13.29
$KS\ F(p)$	1.12	1.13	1.10	1.10
$KS\ L(u)$	1.27	1.37	1.16	0.64

Notes: Estimated standard errors in parenthesis. Column (1) gives our main results. In columns (2) and (3) we change the number of firms to respectively $N = 10$ and $N = 14$. Column (4) gives results using utilities corrected for time fixed effects (see specification (B) in Table 4).

Table 5: Estimation results

The calculated utilities are used in the maximum likelihood procedure described in Section 3. The estimation results are presented in Table 5. In Column (1) we give the estimated parameters using the negative of the residuals of specification (A) in Table 4. The estimated share of consumers searching once is 0.71 and highly significant. The estimated share of consumers searching twice is with a estimate coefficient of 0.20 and a standard error of 0.08 significantly different from zero as well. The percentage of consumers searching for all stores around, although insignificant, is about 8%. All other μ_k's are not significantly different from zero. What is striking is that consumers either search for prices at one or two, or at all chains. The estimated share of consumers searching once or twice is around 91%, while only 8% of consumers compare all prices. A similar picture arises when the estimated search cost cdf and pdf are graphed, as in Figure 5(a). The flat part in the lower part of the search cost distribution indicates that consumers either have search cost of more than £0.15, or they have search cost almost equal to zero. Finally, Table 5 shows that the estimated maximum price-cost margin $v_j - r_j$ for the basket is £3.04, which implies average price-to-cost margins between 8% and 9%.[18]

[18]Smith (2004) reports gross margins that are between 11% and 14% and that revenue minus all store costs as a



(a) Main specification

(b) Alternative specifications

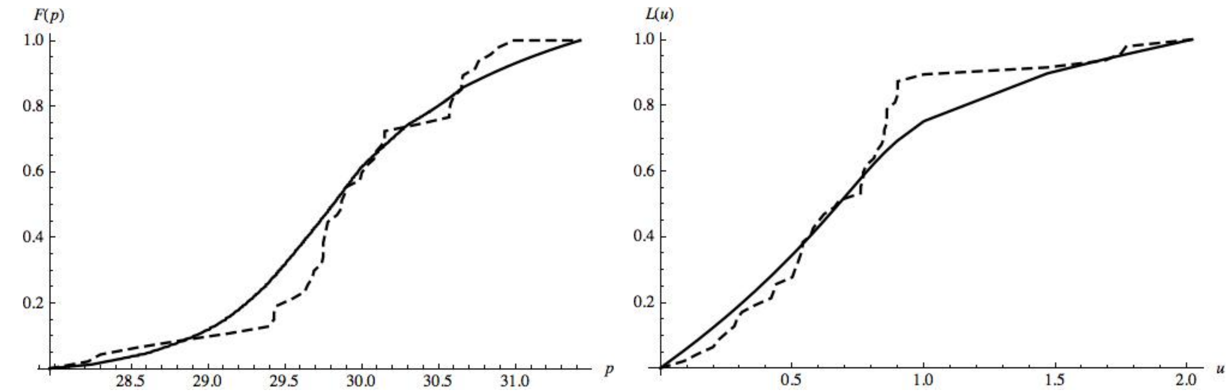
Figure 5: Estimated search cost distribution

As can be seen in Figure 6(a) the model does quite well in explaining the data since the estimated price cumulative distribution function, as indicated by the solid line, is quite close to the empirical one. The results of a more formal Kolmogorov-Smirnov test are put in Table 5.[19] Since the KS $F(p)$ value in the first column is below the 95%-critical value of the KS-statistic, which is 1.36, we cannot reject that the prices are drawn from the estimated price distribution. Of course, given that around 61% of the variation can be explained by store fixed effects, a substantial part of the fit in Figure 6(a) is due to non-search related causes. Since the utility distribution is derived by controlling for store fixed effects, in principle the fit of the utility distribution is a better indicator for determining to what extent search matters. Figure 6(b) shows the estimated utility cdf compared to the calculated utility cdf. As can be seen in this graph, the estimated utility distribution is close to the calculated utility distribution. That the model does quite well in explaining utilities can also be concluded from the corresponding KS $L(u)$ value in Table 5, which is below the critical value of 1.36.

To check for the robustness of the results to different specifications of the utility function we have also estimated the model using a different number of firms. Column (2) in Table 5 gives the estimated parameters for $N = 10$ whereas Column (3) gives the results for $N = 14$ firms. In addition we have plotted the estimated search cost distributions for $N = 10$ and $N = 14$ in Figure 5(b). The results do not change significantly by changing the number of firms. We have also estimated the model using utilities calculated from the time fixed effects specification (B) in Table 4. Column (4) of Table 5 shows that controlling for time fixed effects does not change the estimates

percentage of revenue is between 6% and 10% for the four supermarkets for the year 2000.

[19] In this table KS $F(p)$ is calculated as $\sqrt{m} \cdot \tau_m$, where m is the number of observations and τ_m is the maximum absolute difference over all prices between the estimated price cdf and the empirical price cdf.



(a) Fit price distribution

(b) Fit utility distribution

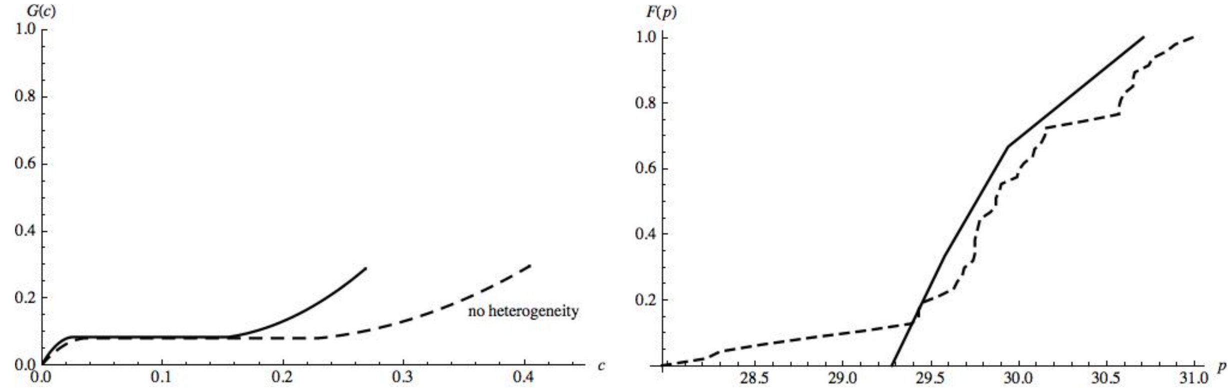
Figure 6: Fit price and utility distributions

that much either. The dashed curve in Figure 5(b) gives the corresponding estimated search cost distribution. Although there are differences in the magnitude of search costs, the results do not change much in a qualitative sense.

4.4 Relative importance of vertical product differentiation and search

A question of interest is how important vertical product differentiation and search are in explaining price dispersion. To answer this question we compare the estimates of the search cost model with vertical product differentiation to estimates from a model without vertical product differentiation. Figure 7(a) gives the estimated search cost cdf if it is assumed the stores are homogeneous instead of vertically differentiated. What is striking is that estimated search costs are now much higher. Given that around 61% of the variation in prices can be attributed to differences between stores and that this is no longer captured in different valuations across stores but in the prices itself, the gains from searching are much higher in the homogeneous search model. To be able to explain observed prices, the population of consumers should have higher search cost on average and should search less than in the search model with vertical product differentiation. Note that the homogenous search model does only slightly worse in explaining the observed prices, but as reported earlier, it fails to explain patterns at a more detailed level, so on those grounds the homogeneous products model can be rejected for this data set.

Finally, in Figure 7(b) we have plotted the empirical price distribution together with the fitted price distribution assuming firm heterogeneity is the only rationale for observed differences in prices. What is striking is that the model does a poor job in explaining high prices and especially



(a) Search cost when no firm heterogeneity

(b) Fit model without search

Figure 7: Estimated search costs without firm heterogeneity and fit without search

low prices. This is not surprising since in a model without search, deviations from the stores' averages prices cannot be explained, unless firm characteristics are changing over time. However, given the relatively short sampling period it seems unlikely that this explains the pricing pattern we observe.

4.5 Organic groceries

Our data set includes prices on organic items, which provides us with a natural case to investigate how search costs relate to consumer demographics: organic food purchasers tend to have distinct demographics and several studies have shown that consumers of organic grocery items have on average higher incomes. If the search behavior of organic food purchasers is affected by this we would expect to see this back in our search cost estimates. To test for this we conduct an experiment in which we compare the search cost estimates using the twenty-four non-organic items discussed above to estimates obtained using a basket of only organic items.

Organic food has quickly become more popular the last few years and is now a multi-billion dollar industry. Although organic farming is growing rapidly, it still accounts for only a small percentage of overall farming. Several studies have shown that organic food purchasers have distinct demographic profiles. In an overview of the empirical literature on organic food consumers Hughner *et al.* (2007) find that a consistent finding across studies is that consumers of organic food are female, have children living in the household, and are older. There is mixed evidence on the effects of income and eduction on organic purchase behavior.

Since organic food purchases tend to have distinct demographics, one would expect them to have different search costs as well. To test if this is indeed the case in our data we have created an organic basket by replacing each item in the original basket with an organic equivalent. Only for one of the items – sugar – we could not find an organic equivalent, so we kept non-organic sugar in the organic basket. For all other items we could find an organic item which more or less resembled the original item. In Table 6 we provide some summary statistics for individual items in the organic basket. Especially prices of produce items seem to be more dispersed than their non-organic equivalents. Overall, the average coefficient of variation for the items in the organic basket is about one percent point higher than that of the standard basket, which means the gains from searching are higher for the organic items. Only Tesco and Sainsbury's carry all items in the organic basket, so we have to focus on these two stores only. However, for the identification of underlying search costs we only need variation over time and not necessarily across stores, so we do not need the prices for other stores to get an estimate of the search cost distribution. We set the number of stores equal to five to take into account that not all supermarket chains in the UK sell organic products.

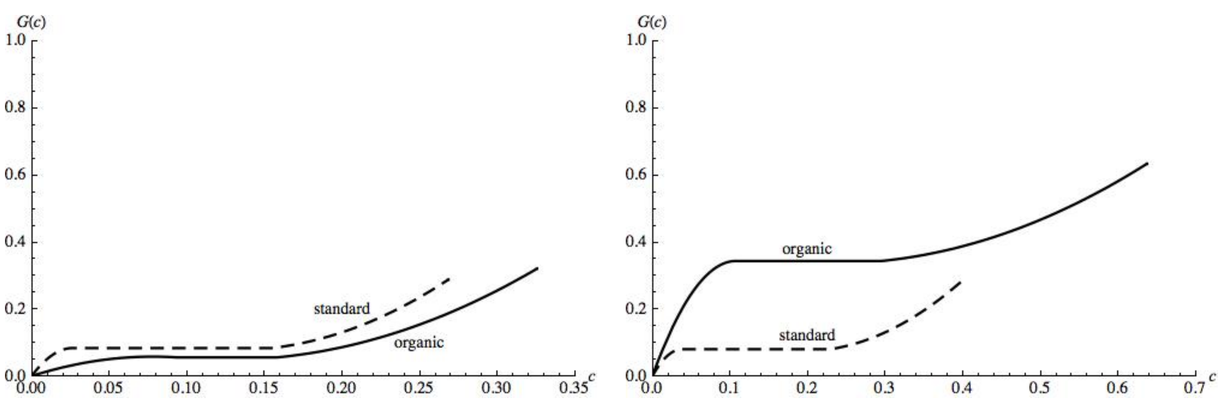
Product	Mean Price (Std)	Minimum Price	Maximum Price	Coefficient of Variation (×100)
organic thick sliced wholemeal bread 800g	0.99 (0.08)	0.89	1.09	8.42
organic bananas bunch of 6	1.39 (0.25)	1.00	1.59	17.96
organic gala polybag apple	2.24 (0.26)	1.99	2.49	11.40
organic sweet pointed peppers	1.81 (0.26)	1.58	2.09	14.29
organic whole cucumber	0.93 (0.09)	0.74	0.99	9.96
organic watercress	1.32 (0.20)	1.00	1.49	15.03
organic baby plum tomatoes 250g	1.66 (0.24)	1.27	1.99	14.70
organic red potatoes 2.5kg	1.91 (0.26)	0.97	2.15	13.78
organic whole milk 3.408 litre	2.39 (0.02)	2.38	2.50	1.03
organic eggs medium box of 6	1.82 (0.00)	1.82	1.82	0.00
organic butter 250g	1.17 (0.07)	1.12	1.27	6.07
organic farmhouse medium cheddar 320g	2.69 (0.23)	2.18	2.95	8.94
organic beef mince 500g	2.98 (0.02)	2.95	2.99	0.62
organic wafer thin ham 100g	2.79 (0.10)	2.69	2.89	3.66
organic petits pois 750g	2.56 (0.09)	2.54	2.99	3.59
organic baked beans 420g	0.45 (0.00)	0.45	0.45	0.00
dolmio organic bolognese sauce 500g	2.17 (0.18)	1.92	2.29	8.22
fairtrade organic strawberry conserve 340g	1.39 (0.00)	1.39	1.39	0.00
silver spoon half spoon sugar 500g	0.88 (0.04)	0.84	0.97	4.59
organic cornflakes 500g	0.83 (0.15)	0.44	0.89	18.23
organic fusilli 500g	0.84 (0.01)	0.84	0.85	0.60
organic basmati rice 500g	1.56 (0.14)	1.39	1.68	9.11
organic 80 teabags 250g	1.41 (0.00)	1.41	1.42	0.14
grove fresh pure organic orange juice 1 litre	2.49 (0.00)	2.49	2.49	0.00

Notes: Prices are in British pounds. For each item we have 24 observations.

Table 6: Simple statistics for items in the organic basket

To obtain utilities we again take the negatives of the residuals of a regression of prices on firm dummies. A major difference with the results in the previous subsection is that firm heterogeneity seems the explain a larger part of the variation in the data; for the organic basket 79% of the variation in prices is explained by the firm dummies while this is only 61% for the non-organic basket.

Figure 8(a) gives the estimated search cost distribution for the organic basket as well as search costs for the original basket. Although the differences are small, the curves show that estimated search costs are higher for the organic basket than for the basket we used in the previous section. The estimated share of people searching for all stores around is about two percent point smaller for the organic basket, while a slightly higher percentage of people searches only once for the non-organic basket. Even though these estimated shares do not differ a lot across the two baskets, because the organic basket is more expensive on average the overall estimated search cost distribution puts more weight on higher search cost values for the organic basket. One of the demographics mentioned above that could explain the higher search costs for organic food purchasers is the age difference between the groups of consumers: older consumers tend to be richer. Using a data set of actual consumer search behavior for online book purchases, De los Santos (2008) finds a significant negative relationship between household income and time spent searching, which suggests a positive relation between search costs and income. In addition De los Santos' (2008) finds weak evidence that households with children present as well as within the 40-54 age group spend less time searching online. As mentioned above, consumers with these demographics tend to be over-represented among the organic food purchasers, which helps to explain differences in estimated search costs between the organic and non-organic basket.



(a) With vertical product differentiation

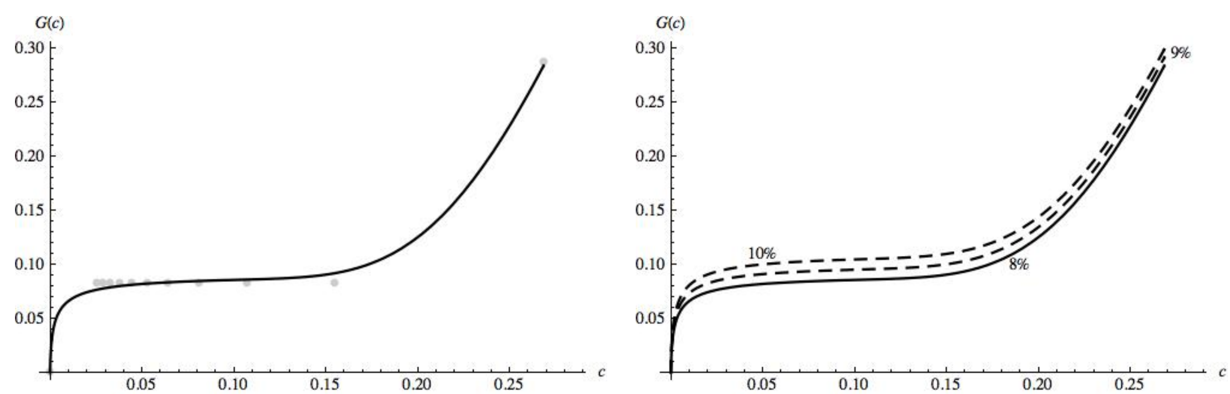
(b) Without vertical product differentiation

Figure 8: Estimated search costs organic versus standard basket

In Figure 8(b) we compare estimated search cost distributions in case we would have ignored heterogeneity among the two stores for which we have sufficient data on the items in the organic basket. Since a much bigger share of the variance is now explained by quality differences between stores, it is not surprising that we end up with very different results. In fact, estimated search costs are now much lower for the organic basket than for the non-organic basket. This illustrates that ignoring store heterogeneity might lead to potentially misleading results.

4.6 The effects of a change in search costs

In a recent paper Armstrong (2008) argues that especially when there are information frictions competition policy may occasionally harm some consumers. For instance, if some fraction of consumers use price comparison tools and observe all prices while others are uninformed, the average price paid by the uninformed shoppers might rise. In addition Armstrong argues that in some settings uninformed consumers exert a negative externality on the informed consumers. Therefore a policy to increase the use of price-comparison sites might not have a major impact on price levels. In this section we take a look at these issues by studying the effects of an exogenous shift in search costs on the equilibrium utility and price distributions for the standard basket. More specifically, we let the share of consumers with very low search costs (the *shoppers* or informed consumers) increase from eight to respectively nine and ten percent, while keeping the other structural parameters in the model fixed. Such a shift in the search cost distribution could for example occur as a result of more people using price comparison sites such as Tesco Price Check or mySupermarket, thereby reducing their search costs to very low levels.



(a) Fitted search cost cdf

(b) Change in percentage shoppers

Figure 9: Simulated search cost cdfs

To be able to calculate the new equilibrium after a change in the search cost distribution we first need to obtain a smooth estimate of the search cost distribution. For this purpose we fit a mixture of log-normal distributions to the estimated search cost points for the standard basket.

The fitted search cost density derived is

$$\hat{g}(c) = 0.91 \cdot \text{lognormal}(c, -1.07, 0.31) + 0.09 \cdot \text{lognormal}(c, -6.00, 2.26).$$

In Figure 9(a) the fitted curve and the estimated points of the search cost distribution are plotted together. We model the change in the percentage of shoppers by adding consumers to the lower end of the search cost distribution. In Figure 9(b) it is shown how the fitted search cost distribution compares to the distributions with the extra shoppers added.[20]

	Basket estimated	Basket fitted	Basket 9% shoppers	Basket 10% shoppers
N	12	12	12	12
# obs	48	48	48	48
μ_1	0.71 (0.19)	0.73	0.78	0.84
μ_2	0.20 (0.08)	0.18	0.12	0.05
μ_3	0.00	0.01	0.01	0.00
μ_4	0.00	0.00	0.00	0.00
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
μ_{N-1}	0.00	0.00	0.00	0.00
μ_N	0.08 (0.11)	0.08	0.09	0.10
$v_j - r_j$	3.04	3.04	3.04	3.04
$E[u]$	0.75	0.71	0.56 (–20.9%)	0.37 (–47.7%)
$E[\max\{u_1, u_2\}]$	1.01	0.97	0.80 (–17.3%)	0.58 (–40.2%)
$E[\max\{u_1, \ldots, u_N\}]$	1.64	1.61	1.50 (–6.9%)	1.33 (–17.0%)
$E[p]$	29.84	29.88	30.02 (+0.5%)	30.21 (+1.1%)
$E[\pi]$	0.18	0.18	0.20 (+7.0%)	0.21 (+15.7%)

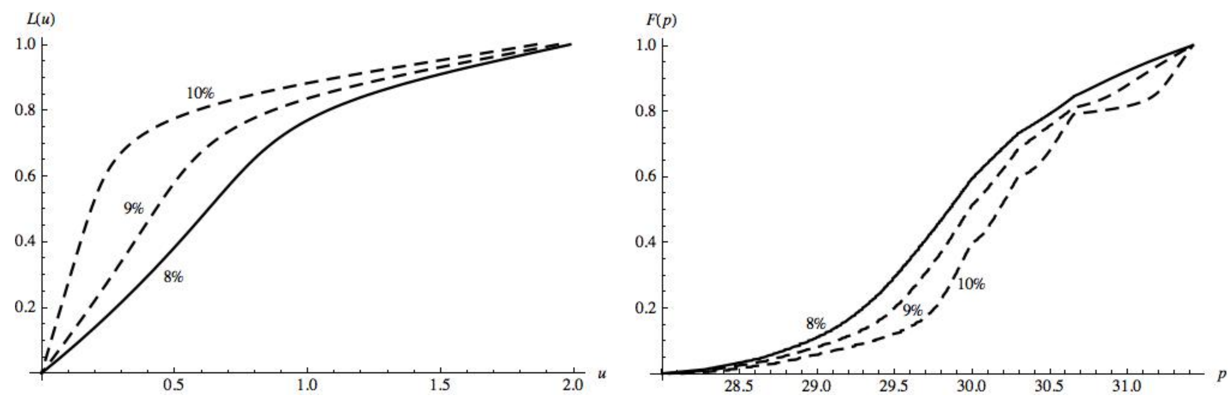
Notes: Column 1: estimated standard errors in parenthesis. Columns 3 and 4: percent changes relative to the fitted equilibrium in parenthesis.

Table 7: The effects of a change in search costs: estimated and simulated parameter values

Using the fitted and the modified search cost distributions, we estimate the effects of a change in the percentage shoppers. The results are reported in Table 7. In addition, Figure 10 gives the simulated price and utility distributions using the fitted and modified search cost distributions. As can be seen in the graphs, a higher share of consumers with low search costs leads to a lower expected utility and less competitive pricing. For example, a one percent-point increase in shoppers leads to an expected utility level which is almost twenty-one percent lower. The expected utility levels encountered by people searching more than once is less affected, although the expected utility level for people searching N times still goes down by almost seven percent. As shown before a large share of the variation in prices is explained by store heterogeneity, which means the effect on prices will not be as large as for utility levels, but still prices are expected to go up by a half percent

[20]We have obtained these search cost distributions by changing the mixture proportions from 0.91-0.09 to respectively 0.90-0.10 and 0.89-0.11.

for the one percent-point increase and more than one percent for the two percent-point increase in shoppers. This counter intuitive result can be explained by a change in focus of the stores: an increase in the share of intensively searching consumers makes it more costly to offer attractive deals. As a result stores start focusing more on the consumers with high search costs. Since a lot of them do not compare prices, it is optimal for the firms to start putting more mass on higher prices. As reported in Table 7 this will make it optimal for some consumers to shift from searching twice to searching once, leading stores to offer even less attractive deals. As a result the profits of the stores increase by as much as almost sixteen percent in case of a two percent-point increase in shoppers.



(a) Utility distribution

(b) Price distribution

Figure 10: Effects of change in search costs

5 Conclusions

This paper presented a non-sequential search model that allows for vertical product differentiation. Firms offering distinct products at different prices can be seen as competing in terms of utilities. It is shown that a symmetric mixed strategy equilibrium in utility levels exists. Firms draw utilities from a common utility distribution, but because valuations and unit costs are different across firms, firms have different price distributions. A result of this is that firms randomize their prices, but over different supports, so that mean prices are different across firms over time, something which so far could not be explained by existing search models.

It is shown how to estimate the model using price data only. Utilities are calculated by taking the negative of the residuals of a fixed effects regression of prices on store dummies. The calculated utilities then serve as an input to a maximum likelihood estimation procedure in order to estimate the underlying search cost distribution.

The method is applied to data from the four biggest supermarkets in the United Kingdom in the period August till October 2008. We find that around 61% of the observed variation in prices is due to firm specific effects. The model does reasonably well in explaining observed prices for a basket of twenty-four staple items. Estimates indicate that most consumers search only once or twice, which is consistent with findings of the Competition Commission. Moreover, a comparison with a basket of similar organic items indicates that organic food purchasers in general have higher search costs. Finally, we illustrate how the estimated search cost distribution can be used to simulate how changes in the share of consumers with low search costs affects equilibrium behavior of consumers and supermarkets. We find that an inflow of consumers with very low search costs leads to lower expected utility levels and higher average prices.

References

[1] Simon P. Anderson and Régis Renault: "Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model," *RAND Journal of Economics* 30, 719-35, 1999.

[2] Mark Armstrong: "Interactions between Competition and Consumer Policy," *Competition Policy International* 4, 97-147, 2008.

[3] Patrick Bajari, Stephanie Houghton, and Steve Tadelis: "Bidding for Incomplete Contracts: An Empirical Analysis," Mimeo, 2007.

[4] Michael R. Baye, 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* 52, 463-96, 2004.

[5] Michael R. Baye, John Morgan, and Patrick Scholten: "Information, Search and Price Dispersion," in: Terrence Hendershott (Editor), *Handbook on Economics and Information Systems*, Elsevier, 2006.

[6] Kenneth Burdett and Kenneth L. Judd: "Equilibrium Price Dispersion," *Econometrica* 51, 955-69, 1983.

[7] Competition Commission: "Supermarkets: A Report on the Supply of Groceries from Multiple Stores in the United Kingdom," The Stationery Office, London, 2000.

[8] Competition Commission: "Market Investigation into the Supply of Groceries in the UK," 2008.

[9] Babur De los Santos: "Consumer Search on the Internet," Mimeo, 2008.

[10] Philip Haile, Han Hong, and Matthew Shum: "Nonparametric Tests for Common Values in First-Price Sealed-Bid Auctions," NBER Working Paper #10105, 2003.

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

[12] Ali Hortaçsu and Chad Syverson: "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: a Case Study of S&P 500 Index Funds," *Quarterly Journal of Economics* 119, 403-56, 2004.

[13] Daniel Hosken and David Reiffen: "Patterns of Retail Price Variation," *RAND Journal of Economics* 35, 128-46, 2004.

[14] Yingyao Hu and Matthew Shum:"Estimating First-Price Auction Models with Unknown Number of Bidders: a Misclassification Approach," Mimeo, 2008.

[15] Renée Shaw Hughner, Pierre McDonagh, Andrea Prothero, Clifford J. Shultz II, and Julie Stanton: "Who Are Organic Food Consumers? A Compilation and Review of Why People Purchase Organic Food," *Journal of Consumer Behaviour* 6, 94-110, 2007.

[16] Maarten C. W. Janssen and José L. Moraga-González: "Strategic Pricing, Consumer Search and the Number of Firms," *Review of Economic Studies* 71, 1089-118, 2004.

[17] Saul Lach: "Existence and Persistence of Price Dispersion: An Empirical Analysis," *Review of Economics and Statistics* 84, 433-444, 2002.

[18] Matthew Lewis: "Price Dispersion and Competition with Differentiated Sellers," *Journal of Industrial Economics* 56, 654-78, 2008.

[19] José Luis Moraga-González, Zsolt Sándor, and Matthijs R. Wildenbeest: "Nonparametric identification and Estimation of Search Costs," Mimeo, 2008.

[20] José Luis Moraga-González and Matthijs R. Wildenbeest: "Maximum Likelihood Estimation of Search Costs," *European Economic Review*, 52, 820-48, 2008.

[21] Peter Morgan and Richard Manning: "Optimal Search," *Econometrica* 53, 923-44, 1985.

[22] Martin Pesendorfer: "Retail Sales: A Study of Pricing Behavior in Supermarkets," *Journal of Business* 75, 33-66, 2002.

[23] Jennifer F. Reinganum: "A Simple Model of Equilibrium Price Dispersion," *Journal of Political Economy* 87, 851-58, 1979.

[24] Howard Smith: "Supermarket Choice and Supermarket Competition in Market Equilibrium," *Review of Economic Studies* 71, 235-63, 2004.

[25] Howard Smith: "Store Characteristics in Retail Oligopoly," *RAND Journal of Economics* 37, 416-30, 2006.

[26] Dale O. Stahl: "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review* 79, 700-12, 1989.

[27] Damon Wingfield and Philip Gooding: "CPI and RPI: the 2008 basket of goods and services," *Economic & Labour Market Review* 2, 25-31.

[28] Asher Wolinsky: "True Monopolistic Competition as a Result of Imperfect Information," *Quarterly Journal of Economics* 101, 493-512, 1986.