

AN EMPIRICAL MODEL OF SEARCH WITH VERTICALLY DIFFERENTIATED PRODUCTS*

Matthijs R. Wildenbeest[†]

First draft: November 2006

Current version: August 2011

Forthcoming in *RAND Journal of Economics*

Abstract

This article provides a framework for studying price dispersion in markets with product differentiation and search frictions. We show under which assumptions we can obtain an equilibrium in which vertically differentiated firms mix prices over different supports. The model can explain the frequently changing prices reported in several empirical studies, but also why some firms have persistently higher prices than others. We show how to estimate the model by maximum likelihood using only prices. Estimates for grocery items in the United Kingdom reveal that most of the observed price variation is explained by supermarket heterogeneity rather than search frictions, whereas the estimated amount of search is low.

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

JEL Classification: C14, D83, L13

*This article is based on the last chapter of my dissertation. I am grateful to the editor, Philip Haile, and two anonymous referees for comments and suggestions that have substantially improved this article. In addition, I wish to thank José Luis Moraga-González for his valuable comments and suggestions. Mike Baye, Ambarish Chandra, Eric Rasmusen, Michael Rauh, Mike Waterson, Chris Wilson, and 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 article has benefited from presentations at HIOC 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 a feature observed in many markets. Recent empirical studies have documented substantial price dispersion in markets for grocery products (Lach, 2002; Caglayan, Filiztekin, and Rauh, 2008), mutual funds (Hortaçsu and Syverson, 2004), electronics sold online (Baye, Morgan, and Scholten, 2004), and books sold online (Hong and Shum, 2006). Several factors may have contributed to the relatively large differences in prices across sellers found in these articles. First, a large theoretical literature on consumer search behavior (e.g. Reinganum, 1979; Burdett and Judd, 1983; Stahl, 1989; Baye and Morgan, 2001; Janssen and Moraga-González, 2004) has shown that imperfect information about sellers' prices may lead to equilibrium price dispersion, even when products are homogeneous.¹ Second, since products differ in terms of characteristics, differences in prices might also be explained by product differentiation.

The substantial variation in prices at any given point in time found in the empirical literature is consistent with a product differentiation as well as with a search friction explanation. However, several studies have found that some sellers set on average higher prices than others, while at the same time prices change frequently (see for instance Lach, 2002; Baye, Morgan, and Scholten, 2004; Lewis, 2008). Although the first observation is consistent with a product differentiation explanation, unless the seemingly unsystematic changes in prices have gone together with changes in the quality of the products offered, these observed pricing patterns are difficult to explain using the product differentiation explanation alone. On the other hand, while search models can explain frequent and asymmetric price changes by equilibrium mixed strategies of homogeneous firms (e.g., Burdett and Judd, 1983; Stahl, 1989), the lack of firm heterogeneity means these models cannot explain why some firms have higher average prices than others.

This article offers a framework that helps explaining both the frequent and asymmetric price changes as well as the persistent differences in average prices found in the empirical literature on price dispersion. A suitable model should have both mixed strategies and firm asymmetries as main ingredients. Although taken separately these two lead to straightforward modeling approaches, a unified framework is more problematic to generate. This is, no doubt, due to the difficulty of generating a unique mixed strategy equilibrium when firms are differentiated. In this article we show which simplifications enable us to map a vertical product differentiation model into a standard homogeneous goods model, such that mixed strategies and firm asymmetries can be combined into

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

a single framework. Although this requires us to make several strong assumptions, our model is able to rationalize the observed pricing patterns better than product differentiation models and search models individually. In addition our model allows us to better measure the relative importance of product differentiation and search frictions as sources of price dispersion than existing models.

Following Armstrong and Vickers (2001) we use a framework in which firms compete directly in utility space. This allows us to reduce the relevant strategy space from two dimensions (quality and price) to a single dimension, which substantially simplifies the analysis. We assume a one-to-one mapping from prices to utilities. Consumers have the same preferences towards quality and are searching non-sequentially for the firm offering the highest utility level. We show that by assuming firms obtain quality input factors in perfectly competitive markets in combination with a constant returns to scale quality production function, we can obtain a symmetric equilibrium in which firms play mixed strategies in utility space. Price distributions will be different for firms offering different quality levels, even though in expectation firms have similar quality-adjusted prices. This means the search model in this article can explain the frequent and seemingly unsystematic price changes observed in several empirical papers, but unlike other existing search models it also offers an explanation why some firms have persistently higher or lower average prices than other firms.²

In the second part of this article 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 percent 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 percent of consumers visit two supermarkets at most.

²Although Wolinsky (1986) and Anderson and Renault (1999) allow for *horizontal* product differentiation, their models do not generate price dispersion. The search model of Armstrong, Vickers, and Zhou (2009) does have price dispersion in equilibrium due to either prominence or quality differences among firms, but does not allow for mixed strategies.

³Several 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 percent. Smith (2006) analyzes store location and size using a characteristics utility model estimated with individual consumer data. Although the focus of this article 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.

Average price-to-cost margins are estimated to be between 8 and 9 percent, which seems reasonable 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 includes prices of 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.

The results of this article contribute to the consumer search literature. Our model builds on the extensive theoretical literature on consumer search and price dispersion, but unlike most search models we allow for vertical product differentiation. In the theoretical part of the article we provide assumptions under which a potentially complicated model reduces to a model that matches the pricing dynamics we observe in the data and can be estimated using only a panel of prices. Our analysis relies on the following assumptions: consumers have the same preferences toward quality and perfectly competitive quality input markets in combination with a constant returns to scale quality production function. These simplifying assumptions make our framework fundamentally different from most standard models of vertical product differentiation but allow us to treat the residuals of a fixed effects regression of prices on a constant as the outcome of a common mixed strategy in the relevant strategy space and permit us to estimate search costs using a maximum likelihood procedure similar in nature to Moraga-González and Wildenbeest (2008).

Our article fits within the recent literature on the structural estimation of consumer search models. Hong and Shum (2006) estimate search cost distributions in a homogeneous 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 frequent and asymmetric price changes as we observe in our data. Moreover, in a mixed strategy equilibrium expected profits need to be the same for all prices in the support of

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

the equilibrium price distribution, 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 non-structural 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 article goes one step further by using the structure of the theoretical search model to estimate the underlying search cost distribution.

In the theoretical model we assume consumers search non-sequentially. Whether this is an appropriate assumption depends on the context. Our method of dealing with firm heterogeneity can nevertheless be applied to sequential search models as well. As shown by Hong and Shum (2006), an advantage of estimating a non-sequential vis-à-vis a sequential search model is that the estimation of the non-sequential search model can be done without making parametric assumptions on the distribution of search costs or using cost data. Although our main focus is on non-sequential search, as a robustness check we also obtain search costs using a sequential search model. Search costs are found to be relatively similar, which suggests that our estimates do not rely too much on the assumption of non-sequential search.

In terms of policy implications, Armstrong (2008) argues that especially when there are information frictions competition policy may occasionally harm some consumers. Indeed, we find evidence in our data that having more intensively searching consumers 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. Intuitively, the additional demand from the intensively searching consumers will make it more profitable for firms to focus on price comparing consumers. To restore the indifference condition required for the mixed strategy equilibrium, profits derived when focusing on non-searching consumers should go up as well. Since in this case firms are already offering the worst possible deals, the only way to increase profits derived from the non-searching consumers is to increase their share by decreasing the gains from search. This is done by setting prices that are less dispersed.

The structure of this article 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 stores offering a homogeneous good to imperfectly informed consumers. The homogeneous good is bundled with several store related services, which add value to the homogeneous good and allow stores to differentiate themselves in terms of quality.⁵ On the supply side there are N stores, indexed by j , selling goods at a unit cost r_j . On the demand side there is a continuum of imperfectly informed consumers demanding at most one unit of the good.⁶ We make the following assumption on the degree of heterogeneity among consumers:

Assumption 1 (consumer heterogeneity) *Consumers have the same preferences towards quality, but differ in their search costs.*

As in Hortaçsu and Syverson (2004), this allows us to assume consumers share a common utility function, i.e., consumers derive utility from consumption of the good sold by store j according to:

$$u_j = v_j - p_j, \tag{1}$$

where v_j is the valuation of buying the good at store j and p_j is the corresponding price. Consumers know their valuation for the good supplied by the different stores but do not observe prices. By engaging in costly search consumers can gain information about the prices of the goods and thus the utilities derived from the goods at a subset of the stores. Consumers then buy the product from the firm in their sample providing the highest utility level.

A consumer's search cost c is assumed to be a random draw from a common atomless distribution function $G(c)$ with support $(0, \infty)$ and positive density $g(c)$ everywhere.⁷ We assume consumers

⁵Examples of store related services are the availability of sufficient car parking space, baggers, sufficient number of cashiers, a convenient location, and flexible opening hours.

⁶Anticipating the empirical analysis of grocery prices presented in Section 4, because supermarkets are multi-product firms and consumers are shopping for bundles of goods, the unit demand assumptions needs some additional justification. We will address this issue later in this article.

⁷In the non-sequential search model of Burdett and Judd (1983) the search cost distribution is degenerated at a particular search cost c , which could lead to a situation where zero or two active search equilibria exists. Tappata (2009) shows the restrictions needed on $g'(c)$ to get a unique symmetric equilibrium in a search all-or-nothing variant of Burdett and Judd (1983). Moraga-González, Sándor, and Wildenbeest (2010) show that if Burdett and Judd (1983) is extended to allow for search cost heterogeneity, uniqueness of a price dispersed equilibrium can be established for the case of two firms and uniform search costs. More general results on uniqueness are hard to obtain because the

search non-sequentially, i.e., consumers determine before entering the market how many times to search (as in Burdett and Judd, 1983). Non-sequential (or fixed-sample-size) search is often thought as a constrained version of sequential search. However, as shown by Morgan and Manning (1985) the optimal search rule allows searchers to choose both the sample size and whether to continue searching and as such includes both non-sequential search and sequential search as special cases. When the search outcome is observed with some delay, like in markets for labor, mortgages, and specialized inputs, non-sequential search is typically optimal, because it allows a searcher to gather information quicker than would have been possible with sequential search.⁸ Although our main focus is non-sequential search, as we will show in Section 4 our method of dealing with firm differentiation does not rely on the non-sequential search assumption and can be extended to sequential search.⁹

Our assumption that consumers have the same preferences toward quality implies that our model is fundamentally different from most standard models of vertical product differentiation (Mussa and Rosen, 1978; Gabszewicz and Thisse, 1979; Shaked and Sutton, 1982). Unfortunately, with only price data it is unlikely that one can separately identify preference heterogeneity and search cost heterogeneity. Even with quantity data, as noted by Hortaçsu and Syverson (2004), to empirically separate preference heterogeneity from search cost heterogeneity one needs to observe something that moves the search cost distribution independently of preferences for quality. Still, to some extent the distribution of search costs takes up the role of preference heterogeneity: even if prices are the same in our model, market shares will be different because search cost heterogeneity creates heterogeneity in choices.

Our objective is to characterize stores' optimal pricing strategies in environments where both search frictions and store heterogeneity are important. We therefore assume each store's quality level is fixed in the short run. In addition we make the following assumptions on quality:

Assumption 2 (quality) *Firms obtain quality input factors in perfectly competitive markets and the quality production function exhibits constant returns to scale.*

equilibrium cannot be computed explicitly, but simulations suggest the uniqueness result is more general and also applies to the model in this article.

⁸De los Santos, Hortaçsu, and Wildenbeest (2011) find that the non-sequential search protocol does a better job explaining observed search patterns for online books than the sequential search protocol, even though in that market there is no such delay in observing the search outcome.

⁹As shown by Hong and Shum (2006), one particular issue when estimating a homogeneous (mixed strategy) sequential search model is that it can only be estimated by either making parametric assumptions on the search cost distribution or by using marginal cost data, whereas neither one is necessary for estimation of the non-sequential search model.

To see how this helps the analysis, let q_j denote store j 's quality, which we assume to be a function of input factors \mathbf{y} according to a quality production function $q_j(\mathbf{y})$ with constant returns to scale. Assume consumers' valuation for a store offering quality level q_j has the additively separable structure $v(q_j) = x + q_j$, where x denotes consumers' valuation for the homogeneous good itself, so independent of store quality. Firms determine q_j such that the valuation-cost markup $v(q_j) - r(q_j)$ is maximized.¹⁰ In perfectly competitive quality input markets, factor prices are equal to the value of their marginal products; with constant returns to scale, by Euler's theorem, the total cost of quality inputs exhausts quality related output, i.e., $r(q_j) = q_j$. This, in turn, implies that the valuation-cost markup does not depend on store quality, i.e., $v(q_j) - r(q_j) = x + q_j - r(q_j) = x$.¹¹ By having v_j and r_j related in this way, as we will show below, firms are symmetric in the margin received at each offered utility level. This makes firms symmetric in the relevant strategy space, which allows us to focus on symmetric mixed strategy equilibria in utility levels.

Following Armstrong and Vickers (2001) we analyze the game by having firms compete directly in utility space. Firms and consumers play a simultaneous moves game. Valuations and unit costs are common knowledge. Therefore, an individual store takes the strategies of the other stores and the search behavior of consumers as given while setting utility. A firm j 's strategy is denoted by a utility distribution $L_j(u)$. An individual consumer takes the stores' strategies as given and decides on a number k of stores to visit in order to maximize utility. The fraction of consumers sampling k firms is denoted by μ_k .

Given the constant valuation-cost markup we can limit our attention to symmetric equilibria in utility levels, i.e., firms have a common utility distribution $L(u)$. The utility distributions are therefore i.i.d. across firms, which means that ex ante consumers view all firms as identical in expected utility terms and consumers search randomly among firms. A first condition that partially characterizes a symmetric equilibrium in utility space is that some consumers search once, while others search more than once (see Burdett and Judd, 1983). The intuition for this is that if all consumers did compare stores, all firms would set a price equal to their unit cost. As shown above the valuation-cost markup is constant in the choice of quality related input components, so even

¹⁰By Assumption 1 all consumers share the same utility function $u = v(q_j) - p(q_j)$, so offering a specific combination of quality and utility implies a price $p(q_j) = v(q_j) - u$. Therefore, the margin firm j makes when offering quality q_j such that a consumer's net utility is u , is $p(q_j) - r(q_j) = v(q_j) - r(q_j) - u$. This means that for a given utility level u (which determines consumer demand), firms decide on their quality levels such that $v(q_j) - r(q_j)$ is maximized.

¹¹See the appendix for an example for the case of a Cobb-Douglas quality production function with two quality related input factors. One implication of Assumption 2 is that unit cost is increasing in quality at a constant rate, while the theoretical literature on quality typically assumes unit cost is increasing in quality at an increasing rate (see for instance Mussa and Rosen, 1978).

though firms might be offering different service levels they would still be offering the same utility level $v_j - r_j = x$. As a result, there is no reason to search. On the other hand, if no consumer would be willing to compare stores, 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.

A second condition that partially characterizes a symmetric equilibrium in utility space is that stores have mixed strategies in utility levels. The reasoning for why this has to be the case follows earlier search papers (e.g., proposition 3 in Varian, 1980). Intuitively, if some utility level u is set with positive probability, there is a positive probability that another firm also sets a utility level u . Since some consumers sample more than one firm, offering slightly more utility $u + \varepsilon$ will give a discrete jump in profits with the same probability with which the other firms set u . For small ε this will be profitable, which means there can be no atoms in the equilibrium utility setting strategies. As a result, firms draw utilities from a continuous cumulative distribution function $L(u)$.

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 $k \cdot c$. 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 $E[\max\{u_1, u_2, \dots, u_k\}] - kc = E[\max\{u_1, u_2, \dots, u_{k+1}\}] - (k + 1)c$, or

$$c_k = E[\max\{u_1, u_2, \dots, u_{k+1}\}] - E[\max\{u_1, u_2, \dots, u_k\}]. \quad (2)$$

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

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

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:

$$(x - u) \sum_{k=1}^N \frac{k\mu_k}{N} L(u)^{k-1} = x \cdot \frac{\mu_1}{N}. \quad (3)$$

Unfortunately, this equation cannot be solved for $L(u)$, so the equilibrium distribution of utilities is only implicitly defined. Solving equation (3) for u gives

$$u = x \cdot \frac{\sum_{k=2}^N k\mu_k L(u)^{k-1}}{\sum_{k=1}^N k\mu_k L(u)^{k-1}}. \quad (4)$$

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

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

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

$$l(u) = \frac{\sum_{k=1}^N k \mu_k L(u)^{k-1}}{(x-u) \sum_{k=1}^N k(k-1) \mu_k L(u)^{k-2}}, \quad (6)$$

where $L(u)$ solves the equilibrium condition in equation (3).

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

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

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

$$c_k = \int_0^1 u(y)[(k+1)y - k]y^{k-1} dy. \quad (7)$$

Then using the same change of variable in equation (4) we can get rid of $u(y)$ in equation (7).

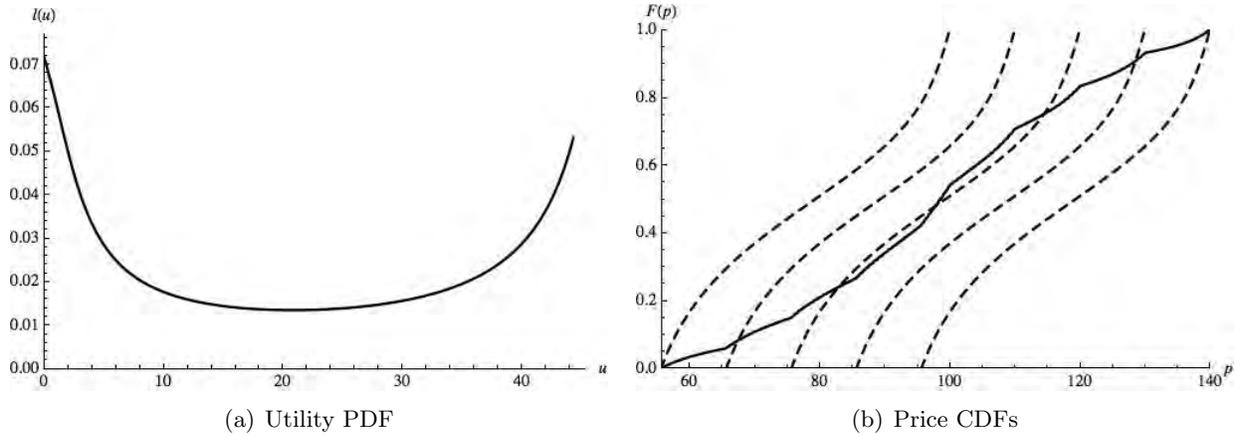


Figure 1: Example equilibrium search model

As an example, we calculate equilibrium when consumers search costs are drawn from a log-normal distribution with parameters 0.5 and 5. Figure 1 gives plots of the equilibrium for 5 firms with valuations ranging from 100 to 140 and $x = 50$ so that unit costs range from 50 to 90. In Figure 1(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 1(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. 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 .¹² In both methods utilities are calculated by restricting the shape of the price

¹²For the analysis the utility levels are 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

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 bears resemblance with how is being dealt with heterogeneity in the structural auction literature. 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 An, Hu, and Shum, 2010; Bajari, Houghton, and Tadelis, 2006). The auction specific component is assumed to be a function of observed auction characteristics, which, as shown by Haile, Hong, and Shum (2003), implies the residuals of a regression of observed bids on the covariates can be treated as normalized bids.

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 (6) can be used to estimate the search cost distribution by maximum likelihood. Since all firms are assumed to draw utilities 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 number of parameters appearing in the likelihood function can be reduced by solving the calculated upper bound of the utility distribution in equation (5) 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 (6). 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 , $k = 1, 2, \dots, N - 1$.

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.

valuations v_j .

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 primary focus is on low prices.

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 supermarket chains 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. Morgan and Manning (1985) have shown that in situations where there is a delay between the decision to search and search outcome non-sequential search is typically optimal, while sequential search is usually optimal in the absence of such a delay. A limitation of assuming non-sequential search in this particular setting is that for a typical consumer there is not much delay between search outcome and search decision when searching for grocery products, which suggests sequential search is the more adequate search protocol to model search activity for

this market. Still, to justify the non-sequential search 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 resembles non-sequential search.¹³

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 unit 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, whereas the model in Section 2 is a static one. The implicit assumption is that supermarkets play a stationary repeated game of finite horizon, which we interpret as every firm making a new draw from the equilibrium utility distribution in every new period. Variation in utility over time (and hence prices) therefore reflects the mixed strategy equilibrium. This also 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

¹³This is valid if (a) consumers act strategically in selecting where to do their secondary (top-up) shopping trips and (b) if prices do not change between the primary and secondary shopping trip. Unfortunately, both are difficult to verify. Around 70 percent of consumers do one main shopping trip a week, so any secondary shopping is likely to be done in the same week. Since we have only weekly price data it is not possible to see if prices have changed throughout the week. However, there is some information available on the nature of the top-up shopping trips. According to the 2000 Competition Commission supermarket investigation a typical shopper at one of the four major supermarkets spends one-third of her grocery expenditures at that supermarket, while around forty percent is spread out over the other three major supermarkets (see Table 11 of Appendix 4.1 of Competition Commission, 2000). Unfortunately, no distinction between secondary and primary shopping is made, although the report mentions that according to most of the major supermarkets a significant number of top-up customers is visiting their stores (up to nearly 50 percent of sales at Safeway, now Morrisons).

some supermarket, to some extent loyalty can also be accommodated in the current setting.

The model relies on the assumption of perfectly competitive quality input markets. Examples of store related services are flexible opening hours, baggers, and a sufficient number of cashiers. Supermarkets typically use unskilled labor to provide these services, for which a perfect competition assumption does not seem too far off.

The focus of this study will be on relatively homogeneous goods, but we allow for the possibility that supermarkets are differentiated in terms of the service they offer. Most theory models explain price dispersion by either mixed 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 move up and down their price distributions. 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. In addition our framework allows us to better measure the relative importance of product differentiation and search friction in explaining price dispersion.

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 mixed 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. We study what happens to equilibrium pricing and searching when there is an exogenous shift in search costs. We also investigate to what extent our findings rely on the non-sequential search assumption. Finally, we examine to what extent alternative models can explain pricing patterns observed in the data.

Description of the data and empirical issues

The data is collected using Tesco Price Check, a price comparison tool put by Tesco on its website.¹⁴ 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,

¹⁴See <http://www.tescopricecheck.com>. The exact nature of Tesco Price Check has changed somewhat over time; currently Tesco compares prices with only ASDA and only for products that appear on the customer's recent receipt.

Sainsbury's, and Morrisons. The survey Tesco uses for collecting data covers only superstores. All four have adopted a national pricing policy.¹⁵

Our data set covers a period of twelve weeks from September till October 2008.¹⁶ 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 to 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 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 this is that the 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. Another drawback of focusing on a fixed basket of goods is that we are ignoring price search for related products within a supermarket.¹⁷ Our estimates therefore only reflect price comparison between different supermarkets, so this should be kept in mind when interpreting the results.

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,

¹⁵This excludes smaller stores such as Tesco Metro and Express and Sainsbury's Local and Central.

¹⁶In the end of 2008 Tesco started comparing prices with only one competitor at a time. The prices of this competitor were used for a couple of weeks, whereafter Tesco typically switched to one of the other two, making the price comparison tool less suitable as a data source for our analysis.

¹⁷According to the consumer survey carried out by the Competition Commission in 2000, 30 percent of respondents usually or always compare grocery prices within the same supermarket, while 47 percent never or rarely does so.

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 baskets that can be constructed out of all available items. About 64 percent 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.¹⁸ To deal with this we instead create a representative basket by taking a list of twenty-four items used by comparison website mySupermarket.co.uk to track groceries expenditures.¹⁹ The price of the basket is then the sum of all individual prices of the twenty-four items.²⁰ Products are selected based on popularity and the basket consists of both branded items and store brands. The 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.²¹ 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.²² Table 1 gives an overview of the products selected as well as some summary 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.

¹⁸One way to deal with this is to randomly select 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 consist 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.

¹⁹We thank mySupermarket for sharing this list with us.

²⁰We do not have data on expenditure to control for expenditure weights. However, the items on the list are selected in such a way that they are representative of the weekly grocery needs of a typical consumer.

²¹See for example <http://news.bbc.co.uk/2/hi/business/7362676.stm>.

²²For 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.

Table 1: Summary statistics for items in the basket

Item	Mean Price (Std)	Minimum Price	Maximum Price	Coefficient of Variation ($\times 100$)
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

Note: The list is based on the basket of staple items mySupermarket.co.uk uses to track consumer groceries expenditures. Unless a brand name appears in the item description, products are store brands. Prices are in British pounds. For each item we have 48 observations.

During the sampling period Tesco Pricecheck only compared 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 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 this will be a parameter in the equilibrium utility distribution. Twelve supermarkets had an expected UK market share larger than 0.5 percent in 2007, which we will take as the number of firms competing.²³

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.

²³Note that estimating N is problematic. For positive μ_N the utility density function goes to infinity when N goes to infinity, which means the log-likelihood function goes to infinity as well.

As a results, supermarkets with better valued characteristics can ask higher prices on average.

Even though unit costs can be different across supermarket because of differences in service levels, in our model it is implicitly assumed that wholesale price for the items in the basket are similar across supermarkets. However, even if a wholesaler charges the same wholesale price to each supermarket, the net price paid will included fixed and variable promotional discounts, settlement discounts, overrides, and other payments, and is therefore likely to be different across retailers. According to a supplier pricing analysis carried out by the Competition Commission, there are differences in wholesale prices paid across retailers, but these differences are mostly across different groups of retailers. For instance, the four largest grocery retailers in our empirical exercise combined achieve the lowest wholesale prices and pay prices that are on average around 5 percent below the mean of all retailers (see Appendix 5.3 of Competition Commission, 2008). Retailer-specific information on relative wholesale prices is confidential, but the Competition Commission does report that among the largest retailers Tesco pays the lowest wholesale prices on average, although this does not mean they pay the lowest wholesale price for every item. More detailed information is available in the 2000 investigation (see Appendix 11.5 of Competition Commission, 2000): for the suppliers' top five lines Tesco paid on average 3.8 percent less than the average prices paid by retailers, while this was 2.3 percent less for Sainsbury, 2.2 percent less for ASDA, and 0.2 percent less for Morrisons. Except for Morrisons, the differences between prices charged by suppliers are relatively small for the grocery chains in our analysis.²⁴

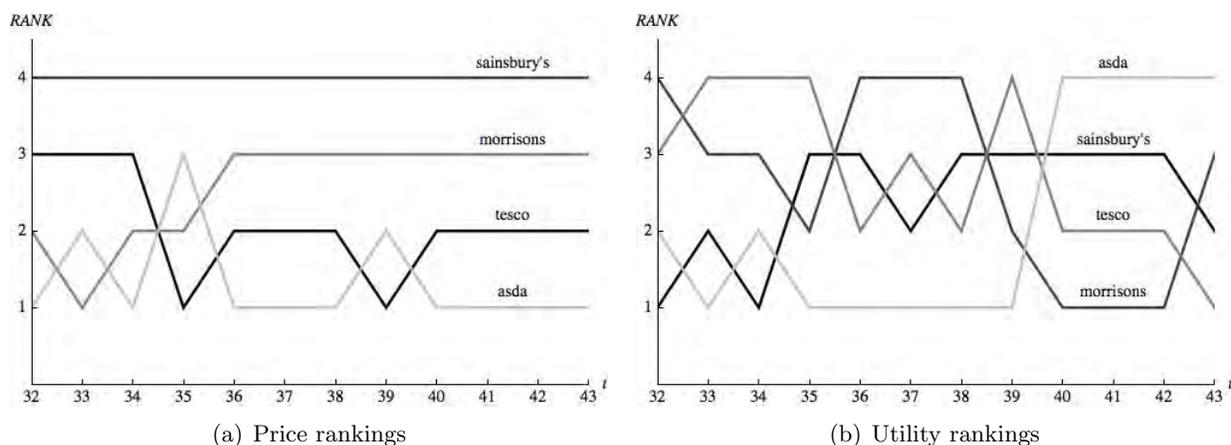


Figure 2: Rankings over time

²⁴Although this suggests Morrisons had somewhat lower buying power than the other three major retailers, Morrisons obtained Safeway in 2004, which is likely to have improved Morrisons overall position towards suppliers (Safeway paid 1.2 percent less than the average prices paid by retailers in 2000).

According to the model prices are 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 fluctuating. At the other extreme stores are so much different that the supports do not overlap at all and price rankings do not change. Figure 2(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 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 price rankings should show more variation. Notice that these findings do not contradict our model, since the model allows for non-fluctuating price rankings. Nevertheless, since in our model firms have a common mixed strategy in utility, a clear prediction of the model is that although price rankings may be constant, firms should be moving up and down the store ranking in terms of utility. Figure 2(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.²⁵ Although rankings are not changing every week, there is now more variability. 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.

Table 2: Quartile spent time in

Quartile	Prices				Utilities			
	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

Note: In percentages.

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 rankings, but usually not every week. For example, 42 percent of the price observations in the first quartile were for stores that had a price in this quartile for one successive week. Likewise, 25

²⁵See Table 4 on p. 24 for the regression results.

percent of prices in the first quartile belong to stores that were in this quartile for three successive weeks. Especially the stores that have a price within the third or fourth quartile stay there for many weeks: at least 67 percent 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.

Table 3: Durations by quartile

duration	Prices				Utilities			
	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

Note: In percentages.

If stores have mixed strategies in utilities, and therefore in prices, there should not be any serial correlation. To formally test for this we have calculated autocorrelation functions (ACF) for each basket. Except for ASDA's price series, the calculated autocorrelations are not significantly different from zero.²⁶ This suggests serial correlation is not a serious issue in the data, although the relatively short panel means any formal test will have little power.²⁷

Finally, to formally test whether the sample of calculated utility levels for a given chain is drawn from the empirical utility CDF of all chains together, we use a Kolmogorov-Smirnov test (K-S test).²⁸ All K-S values are below the 95 percent critical value of 1.36, which means we

²⁶The ACF values are 0.44 for Tesco, 0.54 for Sainsbury's, 0.15 for Morrisons, and 0.73 for ASDA. If the autocorrelation is within $\pm 1.96/\sqrt{T}$, where $T = 12$ is the number of price observations over time for each supermarket chain, it is not significantly different from zero at (approximately) the 5 percent significance level.

²⁷To see if the test can actually detect serial correlation in similar sample sizes as we are looking at, we have done a small Monte Carlo simulation, using 10,000 replications. Our results indicate that if prices are generated by a simple AR(1)-process, i.e., $p_t = p_{t-1} + \varepsilon_t$, and if we use the suggested cutoff value of $1.96/\sqrt{12} \approx 0.57$, in about 52 percent of the replications we wrongfully reject that there is no serial correlation. If we add a time-trend to the specification, i.e., $p_t = p_{t-1} + 0.2t + \varepsilon_t$, the percentage in which we fail to reject serial correlation drops to approximately 16 percent. At more conservative cutoff values these percentages go down, for instance at 0.45 (which means according to ACF values we cannot reject no serial correlation for two of the four supermarkets) without a time-trend the percentage of replications for which we wrongfully reject serial correlation is approximately 33 percent, while with the time trend included this happens in about 7 percent of the replications.

²⁸The K-S values are calculated as $\sqrt{T} \cdot \tau$, where $T = 12$ is the number of observations over time for a given

cannot reject that the calculated utility values for each chain are drawn from the empirical utility CDF constructed by using the utility levels of all chains. Taking prices instead we can reject that observed prices are drawn from the overall empirical price CDF for two out of four grocery stores. This suggests our approach of assuming a common utility distribution better fits the data than assuming a common price distribution.

Estimation of search costs

We use the representative basket of twenty-four regularly bought items to estimate search costs. Figure 3(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 mixed 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.

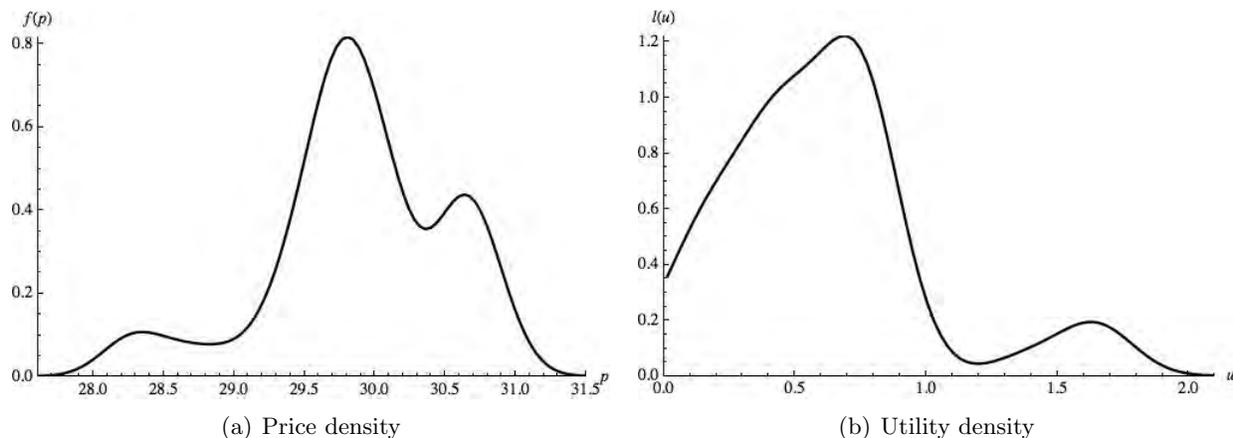


Figure 3: Price and utility density basket

Table 4 gives the results of the fixed effects regression. Specification (A) has only store dummies included, while specification (B) takes time fixed effects into account as well. In both cases the R^2 is quite high: 0.61 and 0.75 respectively. The R^2 tells us something about the relative importance of differentiation versus search frictions. If we do not take store heterogeneity into account when estimating search costs 100 percent of the variation in prices is attributed to search. According to

supermarket chain and τ is the maximum difference between the empirical CDF of utility levels and the calculated utilities levels for a given chain.

specification (A) store dummies take up 61 percent of the variation in prices, so with store dummies added the percentage of variation in prices that we can attribute to search drops from 100 percent to 39 percent.²⁹ 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 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.³⁰ 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.

Table 4: Regression results

	(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

Note: Standard errors in parenthesis. In all specifications the dependent variable is the price for the basket across chains. 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.

Figure 3(b) gives a kernel estimate of the utility density function, where the utilities are the negatives of the residuals of the fixed effects regression in specification (A). Figure 3(b) shows that the utility density is right-skewed, which tells us that although it is possible to encounter relative

²⁹If we take time fixed effects into account, 71 percent of the variation in prices that is not related to time fixed effects is due to firm fixed effects.

³⁰The data is taken from Appendix 3.1 of the final report of the Competition Commission’s (2008) Groceries Market Investigation.

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.

Table 5: Estimation results

	(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
K-S $F(p)$	1.12	1.13	1.10	1.10
K-S $L(u)$	1.27	1.37	1.16	0.64

Note: 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).

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 percent. 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 percent, while only 8 percent of consumers compare all prices. A similar picture arises when the estimated search cost CDF is graphed, as in Figure 4(a). Most consumers search only once, so for these consumers we can only infer that their search cost should have been at least 27 pence in order to rationalize their behavior. Similarly, search costs for the 29 percent of consumers who are comparing prices should have been at most 27 pence. As shown in Figures 3(a) and 3(b) as well as Table 4 around sixty percent of the already low price dispersion for the basket across supermarkets disappears after correcting for store differences, which means that in order to find it worthwhile to shop around consumers should have very low search costs. These consumers might enjoy grocery shopping or have low opportunity cost of time. Note that these patterns

in search behavior are roughly in line with survey findings of the Competition Commission—57 percent of respondents never or rarely compare grocery prices across stores, whereas only 19 percent of respondents always or usually compare prices across different supermarkets. 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%.³¹

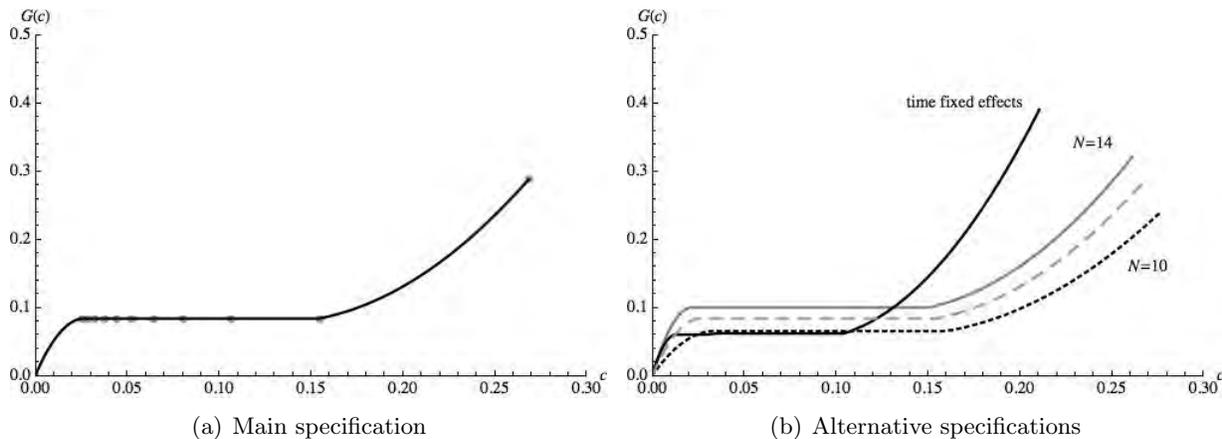


Figure 4: Estimated search cost distribution

As can be seen in Figure 5(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 K-S test are put in Table 5. Since the K-S $F(p)$ value in the first column is below the 95 percent critical value of the K-S statistic we cannot reject that the prices are drawn from the estimated price distribution.³² Of course, given that around 61 percent of the variation can be explained by store fixed effects, a substantial part of the fit in Figure 5(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 5(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

³¹Smith (2004) reports gross margins that are between 11 and 14 percent and that revenue minus all store costs as a percentage of revenue is between 6 and 10 percent for the four supermarkets for the year 2000.

³²See also Footnote 28 on p. 23. In Table 5 K-S $F(p)$ is calculated as $\sqrt{m} \cdot \tau$, where m is the number of observations and τ is the maximum absolute difference over all prices between the estimated price CDF and the empirical price CDF. The 95 percent critical value of the Kolmogorov Smirnov Statistic is 1.36 (see Massey, 1951). Note that this is a conservative estimate since some of the parameters that enter the test are estimated (see Lilliefors, 1967)—the appropriate probability of a Type I error will therefore be smaller than suggested by the standard cutoff values of the Kolmogorov Smirnov statistic (around 1 percent instead of 5 percent).

from the corresponding K-S $L(u)$ value in Table 5, which is below the appropriate critical value.

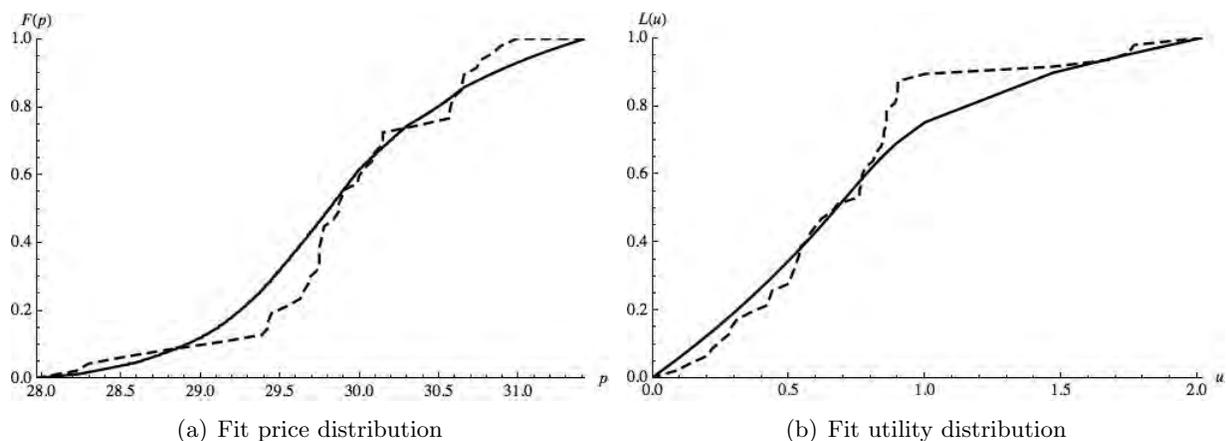


Figure 5: Fit price and utility distributions

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 4(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 that much either. The black solid curve in Figure 4(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.

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 6(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 percent 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 homogeneous 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. This can also be seen in Figure 7, where we have plotted the estimated search cost distributions using data from only one firm at a time. As shown in Figure 7(a), estimates obtained by estimating the model with vertical product differentiation separately for each firm are roughly consistent with one another.³³ However, if we use prices instead of utility levels, Figure 7(b) shows the estimated search cost CDFs are very different for the four supermarket chains. This not only illustrates correcting for quality differences is important if we want to pool observations across supermarkets, but also helps to support the underlying homogeneity assumption in terms of utility levels.

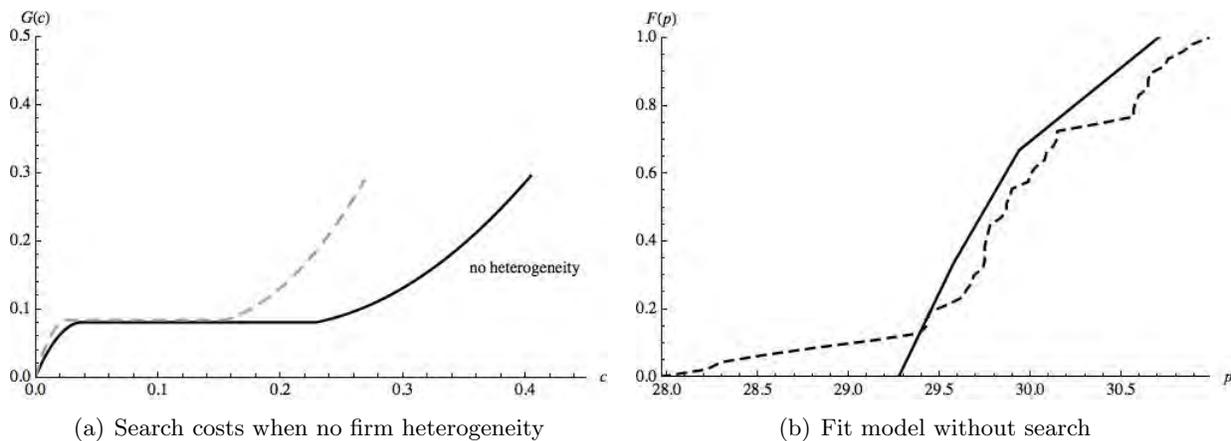


Figure 6: Estimated search costs without firm heterogeneity and fit without search

Finally, in Figure 6(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 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.

³³Note that the search cost estimation stage is done using data from a single firm, but in order to estimate differences in valuations across firms using the fixed effects regression we need to use data from all supermarket chains together.

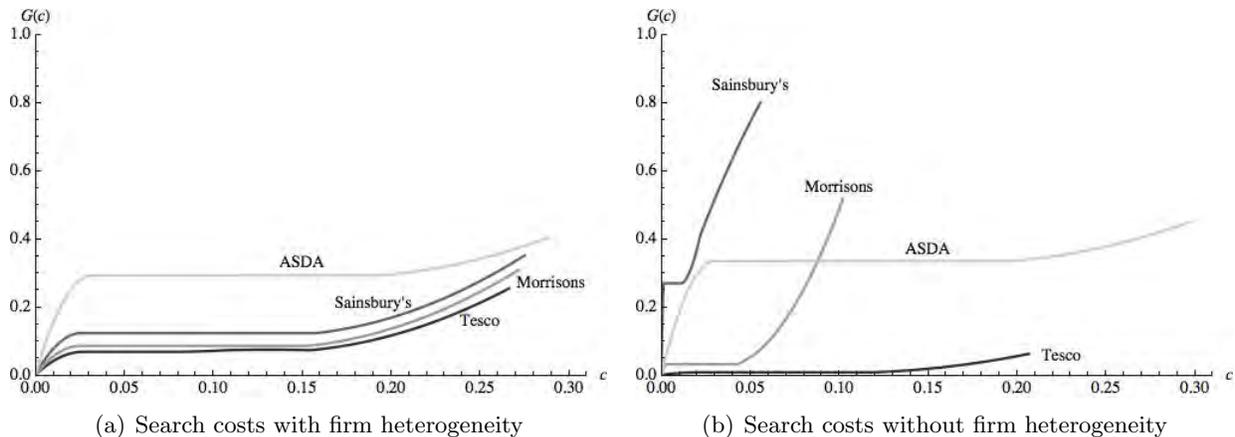


Figure 7: Estimated search costs using data from one firm only

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 education 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.

Table 6: Summary statistics for items in the organic basket

Product	Mean Price (Std)	Minimum Price	Maximum Price	Coefficient of Variation ($\times 100$)
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

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

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 to explain a larger part of the variation in the data; for the organic basket 79 percent of the variation in prices is explained by the firm dummies while this is only 61 percent 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. To make a fair comparison we have re-estimated the search cost distribution for the original basket using the same subset of stores as for the organic basket. Al-

though 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 wealthier. 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.

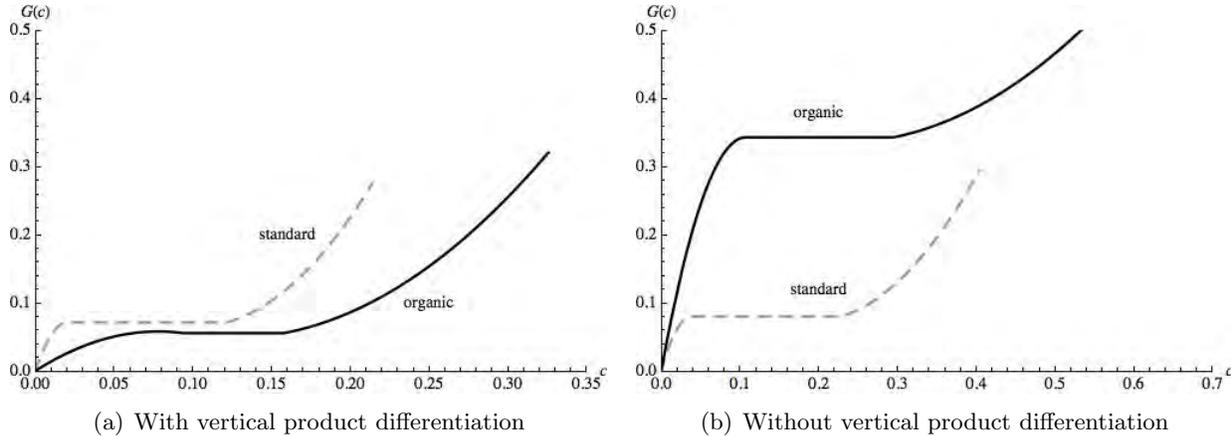


Figure 8: Estimated search costs organic versus standard basket

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.

The effects of a change in search costs

In a recent article Armstrong (2008) argues that especially when there are information frictions, competition policy may occasionally harm some consumers. For instance, if some fraction of consumers uses 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. 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.

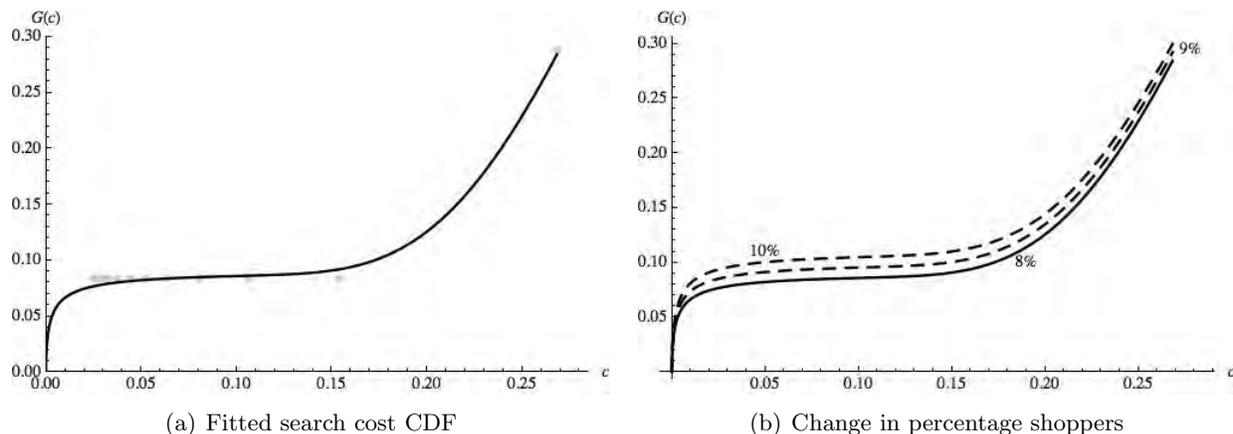


Figure 9: Simulated search cost CDFs

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.³⁴

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

	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%)
$E[\max\{u_1, \dots, u_N\}]$	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%)

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

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 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 as follows. An increase in the share of intensively searching consumers means firms will make more profits from these consumers. However, the equilibrium condition is such that a firm should be indifferent between focusing on the searching and non-searching consumers. When maximizing surplus from the non-searching

³⁴We 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.

consumer firms are already offering the worst possible deals, so the only way to increase profits from the non-searching consumers and restore the indifference condition is to try to increase the share of consumers who are not searching: some consumers searching twice should find it optimal to start searching once. To decrease the gains from searching for the consumers searching twice firms have to offer less dispersed deals than before. As a result firms start putting more mass on higher prices, while at the same time decreasing the upper bound of the utility distribution. As reported in Table 7 this will make it optimal for some consumers to shift from searching twice to searching once. As a result the profits of the stores increase by as much as close to sixteen percent in case of a two percent-point increase in shoppers.

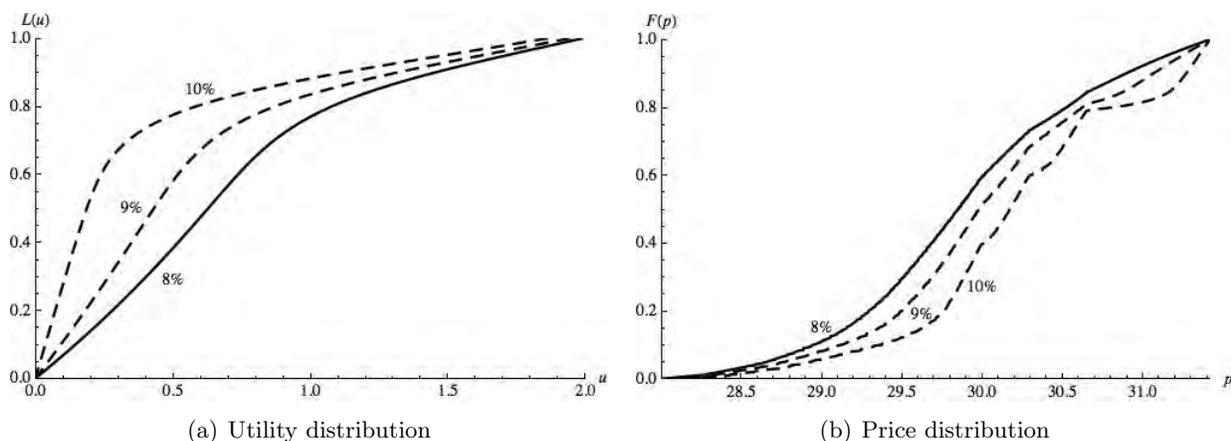


Figure 10: Effects of change in search costs

Sequential search

Throughout the analysis we have assumed consumers search non-sequentially, which, depending on the context, may not be the optimal search protocol for consumers. Fully estimating a sequential search model is beyond the scope of this article, but to see to what extent our findings rely on the non-sequential search assumption, we do a calibration exercise using a model based on the (homogeneous goods) sequential search model developed in Hong and Shum (2006). In the appendix we extend this model to allow for quality differentiation in a similar way as for the non-sequential search model in Section 2. As shown in the appendix, for fixed unit cost values r_j we can solve for the quantiles of the search cost distribution according to

$$G(c_i(u)) = \frac{\bar{u} - u}{x - u},$$

where $c_i(u) = \int_u^{\bar{u}} [1 - L(u)] du$ is the cutoff search cost for a consumer with a reservation utility corresponding to utility u .

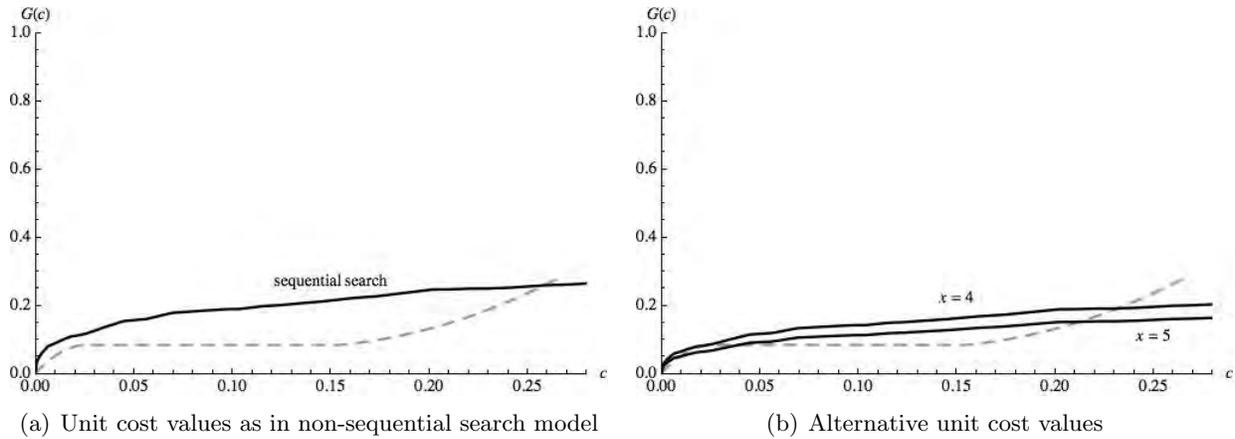


Figure 11: Estimated search cost distribution sequential search

Figure 11 summarizes our findings. The solid curve in Figure 11(a) gives the estimated search cost distribution when fixing r_j at the estimated values for the non-sequential search model, while the dashed curve in Figure 11(a) is the estimated non-sequential search cost distribution. Although the sequential search cost CDF does not have a similar flat part as in the non-sequential search cost CDF, both models give approximately the same share of consumers with search costs above 27 pence, as well as the share of consumers with search costs close to zero. Notice that the estimated search cost distribution depends on the specific values taken for r_j , and the estimates we get using the non-sequential search model might not correspond to the true values for r_j . An alternative is to estimate the model by making parametric assumptions on the search cost distribution—since Hong and Shum (2006) find lower unit cost values for their parametrically estimated homogeneous goods sequential search model than those for their non-sequential search model, as a robustness check we also obtain search cost CDFs using lower values for r_j . In Figure 11(b) we take unit cost values such that the valuation-cost markup x is either 4 or 5 instead of 3.04, which shows that the search cost CDF does not rely too much on the specific values assumed for r_j .

Alternative models

According to our theoretical framework price variation reflects both store heterogeneity and strategies of stores to deliberately change prices over time in order to price discriminate between searchers and non-searchers. Our method takes time-invariant store heterogeneity into account by ascribing

it to differences in store quality, and attributes the remainder of the variation in prices to mixed pricing strategies. Unfortunately our data does not allow us to empirically distinguish between time-varying store heterogeneity and mixed pricing strategies: both can explain the price dispersion we observe at any point in time as well as the ranking changes we observe over time. This means unobserved shocks like retailer specific inventory policies, advertising campaigns that are implemented with different timing, and asymmetric cost shocks might have attributed to the shifts in store rankings we observe in the data. However, our results do not rely on whether we can distinguish between mixed pricing strategies and unobserved shocks: the crucial feature of the model is that players are uncertain about the other players' choices. This uncertainty can either arise because of mixed pricing strategies, or because of uncertainty about the choice of a pure strategy. For instance, if there is a small amount of incomplete information in the form of unobserved shocks, the results of the model still hold and our search cost estimates will be the same, even though the equilibrium will be in pure strategies (purified).

Apart from time-varying store heterogeneity several alternative theories can potentially explain the pricing dynamics we observe in our data. For instance, in a setting in which firms have pure strategies, price changes might simply reflect changes in wholesale prices. However, unless these changes appear asymmetrically across retailers, such a model can easily be rejected since the ranking changes we observe for the basket cannot be explained. Moreover, the average correlation coefficient across chains for prices of specific items in the basket is 0.23, which indicates little correlation across supermarkets. This suggests that common changes in wholesale prices cannot be solely responsible for price changes.

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 Pesendorfer, 2002; Hosken and Reiffen, 2004). To explain these pricing dynamics, Pesendorfer (2002) presents an intertemporal model of demand accumulation, in which low-valuation consumers buy and store only when prices are low, while they consume from their own inventory when prices are high. In this model prices are equal to the willingness to pay of high-valuation consumers most of the time, while periodically products are sold at a randomly drawn sales price, with support of the sales price distribution below the willingness to pay of low-valuation consumers. The model thus predicts that prices are high most of the time, whereas only occasionally products are sold at a discount price. Using ketchup sales data from supermarkets in Springfield, Missouri, Pesendorfer (2002) indeed finds that this is the case for the market studied. However, an important difference

with our article is that while Pesendorfer focuses on a single product, we focus on a basket of goods. This is especially important since Pesendorfer's model applies only to goods that can be stored, while our basket consists of both perishable and non-perishable grocery items.

5 Conclusions

This article has 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. We have shown that by assuming consumers have the same preferences towards quality, and firms obtain quality input factors in perfectly competitive markets in combination with a constant returns to scale quality production function, we can obtain a symmetric equilibrium in which firms play mixed strategies in utility space. Because valuations and unit costs are different across firms, firms have different price distributions. This means firms mix their prices, but over different supports, so that average prices are different across firms over time, something that so far could not be explained by existing search models.

We have 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 has been applied to data from the four biggest supermarkets in the United Kingdom in the period August till October 2008. We find that around 61 percent 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.

Our method has required us to make several strong assumptions. Some of these assumptions can be easily relaxed, like our assumption that consumers search non-sequentially. Other assumptions, for instance our assumption that consumers are homogeneous in their quality preferences, are more

difficult to relax, unless richer data is available. Micro-level data on search behavior as well as quantity data will facilitate the estimation of both preference and search cost parameters. We hope that our findings will help to shape future work on how consumer search behavior affects differentiated product markets.

Appendix

Example constant valuation-cost markup

Consider the case of stores producing quality from input factors capital K and labor L , according to a Cobb-Douglas quality production function $q_j = zK^\alpha L^{1-\alpha}$. Let p_r be the price of K and p_w the price of L . For a given utility level u firms decide on their quality levels such that $v(q_j) - r(q_j) = x + zK^\alpha L^{1-\alpha} - p_r K - p_w L$ is maximized. Taking the first order condition with respect to K and L gives $p_r = \alpha z K^{\alpha-1} L^{1-\alpha}$ and $p_w = (1 - \alpha) z K^\alpha L^{-\alpha}$. Replacing p_r and p_w gives $v(q_j) - r(q_j) = x + zK^\alpha L^{1-\alpha} - \alpha z K^\alpha L^{1-\alpha} - (1 - \alpha) z K^\alpha L^{1-\alpha}$, which simplifies to $v(q_j) - r(q_j) = x$.

Sequential search

Hong and Shum (2006) show that in their (homogeneous) sequential search model, firms set prices according to the following indifference condition (see equation (8) of Hong and Shum, 2006):

$$(\bar{p} - r) \cdot \alpha = (p - r) \cdot (1 - H_p(p)),$$

where \bar{p} is the upper bound of the price distribution, α is the measure of consumers with reservation price equal to \bar{p} , and $H_p(p)$ is the distribution of reservation prices in the population. We can proceed by adding vertical differentiation in a similar way as in Section 2, i.e., by using the indirect utility function $u_j = v_j - p_j$ and a constant valuation-cost markup $v_j - r_j = x$ we can rewrite the indifference condition as

$$(x - \underline{u}) \cdot \alpha = (x - u) \cdot H_u(u),$$

where $H_u(u)$ is the distribution of reservation utilities and \underline{u} is the lower bound of the utility distribution. At the upper bound of the utility distribution $H_u(\bar{u}) = 1$, which we can use to solve the indifference condition, evaluated at $u = \bar{u}$, for α , i.e., $\alpha = (x - \bar{u}) / (x - \underline{u})$. Plugging this into

the indifference condition and solving for $H_u(u)$ gives

$$H_u(u) = \frac{x - \bar{u}}{x - u}. \quad (\text{A1})$$

A consumer who is indifferent between searching and not searching after having observed a utility draw u_i has search cost

$$c_i = \int_{u_i}^{\bar{u}} [1 - L(u)] du. \quad (\text{A2})$$

Using $H_u(u_i) = 1 - G(c_i)$ together with equations (A1) and (A2) we can solve for the quantiles of the search cost distribution, i.e.,

$$G\left(\int_u^{\bar{u}} [1 - L(u)] du\right) = \frac{\bar{u} - u}{x - u}. \quad (\text{A3})$$

The fixed effects regression of prices on a constant will give us an estimate of v_j as well as u , which means that if one has information on r_j we can obtain x . From this we can calculate the search cost distribution G evaluated at the cutoffs of the search cost distribution using equation (A3), where the cutoffs are calculated for all utilities (residuals from the fixed effects regression) using equation (A2).

References

- AN, Y., HU, Y., AND SHUM, M. “Estimating First-Price Auction Models with Unknown Number of Bidders: a Misclassification Approach.” *Journal of Econometrics*, Vol. 157 (2010), pp. 328–341.
- ANDERSON, S. AND RENAULT, R. “Pricing, Product Diversity, and Search Costs: A Bertrand-Chamberlin-Diamond Model.” *RAND Journal of Economics*, Vol. 30 (1999), pp. 719–735.
- ARMSTRONG, M. “Interactions between Competition and Consumer Policy.” *Competition Policy International*, Vol. 4 (2008), pp. 97–147.
- ARMSTRONG, M. AND VICKERS, J. “Competitive Price Discrimination.” *RAND Journal of Economics*, Vol. 32 (2001), pp. 579–605.
- ARMSTRONG, M., VICKERS, J., AND ZHOU, J. “Prominence and Consumer Search.” *RAND Journal of Economics*, Vol. 40 (2009), pp. 209–233.

- BAJARI, P., HOUGHTON, S., AND TADELIS, S. “Bidding for Incomplete Contracts: An Empirical Analysis.” Working Paper No. 12051, NBER, Cambridge, 2006.
- BAYE, M. AND MORGAN, J. “Information Gatekeepers on the Internet and the Competitiveness of Homogeneous Product Markets.” *American Economic Review*, Vol. 91 (2001), pp. 454–474.
- BAYE, M., MORGAN, J., AND SCHOLTEN, P. “Price Dispersion in the Small and in the Large: Evidence from an Internet Price Comparison Site.” *Journal of Industrial Economics*, Vol. 52 (2004), 463–496.
- BAYE, M., MORGAN, J., AND SCHOLTEN, P. “Information, Search and Price Dispersion.” In T. Hendershott, ed., *Handbook on Economics and Information Systems* Amsterdam: Elsevier, 2006.
- BURDETT, K. AND JUDD, K. “Equilibrium Price Dispersion.” *Econometrica*, Vol. 51 (1983), pp. 955–969.
- COMPETITION COMMISSION. “Supermarkets: A Report on the Supply of Groceries from Multiple Stores in the United Kingdom.” London: The Stationery Office, 2000.
- COMPETITION COMMISSION. “Market Investigation into the Supply of Groceries in the UK.” London: Competition Commission, 2008.
- DE LOS SANTOS, B. “Consumer Search on the Internet.” Working Paper No. 08-15, NET Institute, New York, 2008.
- DE LOS SANTOS, B., HORTAÇSU, A., AND WILDENBEEST, M. “Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior.” *American Economic Review*, (forthcoming).
- CAGLAYAN, M., FILIZTEKIN, A., RAUH, M. “Inflation, Price Dispersion, and Market Structure.” *European Economic Review*, Vol. 52 (2008), pp. 1187–1208.
- GABSZEWICZ, J. AND THISSE, J. “Price Competition, Quality and Income Disparities.” *Journal of Economic Theory*, Vol. 20 (1979), pp. 340–59.
- HAILE, P., HONG, H., AND SHUM, M. “Nonparametric Tests for Common Values in First-Price Sealed-Bid Auctions.” Working Paper No. 10105, NBER, Cambridge, 2003.

- HONG, H. AND SHUM, M. “Using Price Distributions to Estimate Search Costs.” *RAND Journal of Economics*, Vol. 37 (2006), pp. 257–275.
- HORTAÇSU, A. AND SYVERSON, C. “Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: a Case Study of S&P 500 Index Funds.” *Quarterly Journal of Economics*, Vol. 119 (2004), pp. 403–456.
- HOSKEN, D. AND REIFFEN D. “Patterns of Retail Price Variation.” *RAND Journal of Economics*, Vol. 35 (2004), pp. 128–146.
- HUGHNER, R., MCDONAGH, P., PROTHERO, P., SHULTZ, C., AND STANTON, J. “Who Are Organic Food Consumers? A Compilation and Review of Why People Purchase Organic Food.” *Journal of Consumer Behaviour*, Vol. 6 (2007), pp. 94–110.
- JANSSEN, M. AND MORAGA-GONZÁLEZ, J. “Strategic Pricing, Consumer Search and the Number of Firms,” *Review of Economic Studies*, Vol. 71 (2004), pp. 1089–1118.
- LACH, S. “Existence and Persistence of Price Dispersion: An Empirical Analysis.” *Review of Economics and Statistics*, Vol. 84 (2002), pp. 433–444.
- LEWIS, M. “Price Dispersion and Competition with Differentiated Sellers.” *Journal of Industrial Economics*, Vol. 56 (2008), pp. 654–678.
- LILLIEFORS, H. “On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown.” *Journal of the American Statistical Association*, Vol. 62 (1967), pp. 399–402.
- MORAGA-GONZÁLEZ, J., SÁNDOR, Z., AND WILDENBEEST, M. “Nonsequential Search Equilibrium with Search Cost Heterogeneity.” Working Paper No. 869, IESE Business School, University of Navarra, 2010.
- MORAGA-GONZÁLEZ, J. AND WILDENBEEST, M. “Maximum Likelihood Estimation of Search Costs.” *European Economic Review*, Vol. 52 (2008), pp. 820–848.
- MASSEY, F. “The Kolmogorov-Smirnov Test for Goodness of Fit.” *Journal of the American Statistical Association*, Vol. 46 (1951), pp. 68–78.
- MORGAN, P. AND MANNING, R. “Optimal Search.” *Econometrica*, Vol. 53 (1985), pp. 923–944.

- MUSSA, M. AND ROSEN, S. "Monopoly and Product Quality." *Journal of Economic Theory*, Vol. 18 (1978), pp. 301–317.
- PESENDORFER, M. "Retail Sales: A Study of Pricing Behavior in Supermarkets." *Journal of Business*, Vol. 75 (2002), pp. 33–66.
- REINGANUM, J. "A Simple Model of Equilibrium Price Dispersion." *Journal of Political Economy*, Vol. 87 (1979), pp. 851–858.
- SMITH, H. "Supermarket Choice and Supermarket Competition in Market Equilibrium." *Review of Economic Studies*, Vol. 71 (2004), pp. 235–263.
- SMITH, H. "Store Characteristics in Retail Oligopoly." *RAND Journal of Economics*, Vol. 37 (2006), pp. 416–430.
- STAHL, D. "Oligopolistic Pricing with Sequential Consumer Search." *American Economic Review*, Vol. 79 (1989), pp. 700–712.
- SHAKED, A. AND SUTTON, J. "Relaxing Price Competition Through Product Differentiation." *Review of Economic Studies*, Vol. 49 (1982), pp. 3–13.
- TAPPATA, M. "Rockets and Feathers: Understanding Asymmetric Pricing." *RAND Journal of Economics*, Vol. 40 (2009), pp. 673–687.
- VARIAN, H. "A Model of Sales." *American Economic Review*, Vol. 70 (1980), pp. 651–669.
- WINGFIELD, D. AND GOODING, P. "CPI and RPI: the 2008 basket of goods and services." *Economic & Labour Market Review*, Vol. 2 (2008), pp. 25–31.
- WOLINSKY, A. "True Monopolistic Competition as a Result of Imperfect Information." *Quarterly Journal of Economics*, Vol. 101 (1986), pp. 493–512.