L'université n'entend ni approuver ni désapprouver les opinions particulières de la candidate

Contents

List of Tables

List of Figures

Acknowledgments

It requires a different set of skills from those of an economist to put into words the importance that a certain group of people had in my life as a PhD student in Toulouse. I will never be able to make you justice in paper.

Pierre Dubois is an extraordinary supervisor. Always very supportive and encouraging, he taught me a lot about Economics and research. He gave me the impression of being part of a team, and made me work when I was so easily distracted. I don't know if I would have finished my PhD if it weren't for him...

Nothing would have been possible without Aude Schloesing. Not even renting an apartment. She is an amazing person, extremely helpful, efficient, caring. I am very thankful to her and I'll miss her!

CAPES and INRA provided the financial support that made this PhD possible.

Vincent Requillart and Alban Thomas played a major role in helping me find a job. Thank you very much!

I would also like to thank the people in the INRA-GREMAQ. I am especially thankful to Valérie Orozco and Christophe Bontemps for their help with the database and Stata. Olivia and Kathia were always very helpful and very sweet.

Friends make everything easier and funner. Thank you very much to my graduate school friends Mikel, Sandra, Tuba, Rita, Catherine, Marie Louise, Gaëtan, Sara, Kevin, Jose, Matteo, Juan Miguel, Florian, Alex, Yu, Ruxanda. The Brazilian gang is like having a family abroad! Muitos beijos e muito obrigada a Adriana, Felipe, Rodrigo, Rodrigo, Luis, Raquel, Luciano, Rafael, Martha e Klênio. Et bien sûr, les toulousains, qui m'ont fait me sentir chez moi à Toulouse: Clément, Guillaume, Laurianne, Nico,Florent, Mathieu, Amelie, Laurent, Milou.

Finally, I am very thankful to my parents and brother, who have always been very supportive, encouraging, and just plain awesome. And Julien, who helped me so much, in so many non academic ways, who puts up with all my mood swings, who believes I am "championne du monde". This thesis is dedicated to the four of you.

Abstract

The goal of this doctoral thesis is to contribute to the better understanding of individual behavior in what concerns food product choices, and its implications to the functioning of the market and to competition policy.

The first chapter focuses on purchase choices in a dynamic demand environment. When consumers stockpile, traditional static discrete-choice models overestimate long-term price responses. I develop a demand model with inventories and estimate the structural parameters fully accounting for consumers' unobservable heterogeneity, but without having to solve the dynamic programming. I find a significant quantitative difference between the price-elasticities yielded by the static and inventory model, pointing to the risks of making wrong policy recommendations based on short run measures.

The second chapter investigates an interesting phenomena reported by the recent literature, i.e., evidence of price decreases during demand peaks. I argue that the price decrease is an artefact of ignoring product differentiation. I develop a simple individual demand model which shows that at periods of exogenous high demand, consumers migrate towards cheaper lower quality products, pushing the average category price down. I test model implications and estimate structural demand for ice-cream purchases, which has a seasonal peak during the summer.

Chapter 3 deals with price dispersion and search costs. I draw a map of price dispersion in French supermarkets, showing that important price differences remain even after controlling for observable and unobservable store and market characteristics. I then perform reduced form tests to investigate the importance of search costs on explaining price dispersion. Finally, I present a model of consumer choice with sequential search costs and develop an empirical strategy to identify the parameters of the search costs distribution. Results indicate that search costs are high and constitute a major cause of price dispersion. Indeed, consumers obtain at most three utility quotes before purchasing the product, with the vast majority of consumers purchasing the first product drawn.

Introduction[1]

The study of Industrial Organization, and in particular of competition policy issues, requires deep understanding of the way consumers behave when making purchasing decisions. The way individuals react to and the amount of information they detain about prices, as well as other observable or unobservable characteristics of a product, affect firm behaviour and consequently the competition environment and have to be taken into account when making policy recommendations.

The goal of this thesis is to contribute to the better understanding of individual behaviour in what concerns food product choices, and its implications to the functioning of the market and to public policy.

In particular, I study consumers' price elasticities, quality choice, and search costs. In each of the three chapters that constitute this thesis, I develop not only a theoretical model explaining the economic phenomena under study, but also an empirical strategy to test implications and identify structural parameters of the model.

The first chapter deals with the estimation of price-elasticities in a dynamic environment, where forward looking consumers are allowed to hold inventories. Although static discrete choice models of individual demand (McFadden 1980, 1984) are simple and intuitive, they are inappropriate in a number of markets. Indeed, if consumers stockpile, then the static model will yield price elasticities that overestimate the long run elasticities, because the price elasticities in the static model will be picking up not only variation on consumption but also the short term variations in inventories. Upper biased measures of price elasticities may have serious

[1] Je remercie le département SAE2 de l'INRA pour l'accès aux données TNS Worldpanel (ex données Secodip) des années 1998-2000 lors de ma thèse dans l'UMR UT1-CNRS-INRA-EHESS, 1291, GREMAQ à l'Université de Toulouse 1.

consequences in terms of policy recommendation. It would lead regulators, for instance, to allow anti-competitive mergers.

I develop a dynamic model of demand with inventories, and estimate structural long run price elasticities which are consumer specific without having to solve the dynamic programming, which is not only a difficult numerical problem, but also very costly in terms of computer time. The main contributions of the chapter are therefore the simplicity of the estimation method and the flexibility with respect to consumers' unobserved heterogeneity. The cost is that product is assumed to be homogeneous.

The main proposition of the paper says that when consumers purchase at a high price, they only purchase to cover consumption not to stockpile. This proposition enables me to separate periods when consumers stockpile and periods when they don't. Relying on the purchase decision equation for periods when they don't stockpile and making an additional assumption on how families consume at periods when they do not purchase, I am able to identify the structural parameter needed to calculate the long run price elasticities in the inventory model.

I also derive testable implications of the model and find compelling empirical evidence to support it. I consider 6 product categories: pasta, tuna, coffee, butter, milk, yogurt. Product categories were chosen as to verify certain criteria. First, I chose products that were consumed at regular basis. I also wanted products that differed in terms of cost of stocking and storability. Finally, I chose product categories which are relatively homogeneous in order to decrease the risk of measurement error due to the broad definition of product.

Empirical results show that price elasticities yielded by a static model which ignores inventories overestimate long run responses by more than 50% and up to 200%. Those results are robust to different price expectation hypothesis. Hence, not considering a dynamic model of demand can significantly bias the results on price elasticities and consequently may lead to wrong policy recommendations.

The second chapter analyses consumer and pricing behavior during periods of exogenous increases in demand. Although we would a priori expect prices to increase during demand peaks, a number of recent studies brings evidence that for certain categories of products, demand peaks are actually periods of lower, not higher, prices.

I contribute to the emerging literature on the effect of demand peaks on prices by formalizing Nevo and Hatzitaskos (2006)'s argument that price sensitivity is higher during demand peaks due to a change in the brand-level preferences of consumers. I derive testable implications of the model, which are tested using micro data on ice-cream purchases. Furthermore, I structurally estimate demand and calculate cross-price elasticities, showing they are indeed higher during demand peaks.

The demand for ice-cream increases significantly during the summer, while the average price paid goes down. To study the quality choice shift, I consider quality as the only product differentiation dimension and I restrict the number of quality levels to two. The quality indicator considered is whether the brand of the product is a national (high quality) or a store (low quality) brand. Although I do not want to start a debate on the actual taste differentials, I claim that in France there is widespread belief that store brand products are of lower quality than national brand products. The perceived quality difference can also be inferred from prices: national brands are on average a lot more expensive than store brands. To check whether results are dependent on the restrictive product definition, tests are also performed considering other dimensions of differentiation (store, brand and product characteristics).

I find evidence that during periods of exogenous increase in demand, consumers are more price elastic. As a result, they tend to shift product choice towards cheaper products. The higher elasticity and the shift in demand drive prices down. The decrease in prices is however less important than the decrease measured by average aggregate prices paid. Since demand migrates towards lower cheaper products, ignoring product differentiation and looking only at average category prices leads to an overstatement of the importance of the price decrease. Results point out the importance of considering product differentiation.

The third and last chapter, written jointly with my thesis supervisor Pierre Dubois, focuses on price dispersion and search costs.

The first goal of the chapter is to draw a map of price dispersion for food products in France.

The second goal is to identify the source of the price dispersion that cannot be accounted for by store and market characteristics. In particular, we test if there is evidence that search costs are driving price differentials. We also study the effect of the opportunity cost of time on prices paid by consumers.

Finally, we estimate the distribution of consumers' search costs. Recovering the distribution of search costs is important because the existence of search costs affects competition policy issues. For instance, in the presence of search costs, firm entry does not necessarily improve welfare. Stahl (1989) shows that an increase in the number of firms in the market may actually decrease welfare depending on how search costs are distributed among the population of consumers. Also, if search costs are important, firms may retain considerable market power even in seemingly competitive situations, which in some markets may justify price regulations.

We consider a number of identical food products sold at different stores in France. Following Lach (2002), for each product considered, we regress the log of prices (expressed in differences from the period's average so that all difference is cross-sectional), pulled over time and store, on a chain-store fixed effect, a market fixed effect, a time-period effect, and a store size effect. The residual from such a regression can be considered as the price of a homogeneous good purged from store and market heterogeneity. It can therefore be interpreted as a measure of the distance (or deviation) from the Bertrand outcome.

We investigate the importance of search costs on explaining price dispersion in two ways. First, we study the effect of a decrease in search costs on price differentials. As suggested by the literature (see Warner and Barsky, 1995), we assume periods of high aggregate demand are periods of lower search costs per product since the fixed component of the search costs is divided by a longer list of items to be purchased. We regress alternative measures of price dispersion on seasonal dummies and dummies indicating periods of exogenous peak in demand, such as Christmas and weekends.

Second, we test the effect of consumers' opportunity cost of time on the prices they pay.

If search costs are relevant, consumers with a high cost of time have a less intense searching activity and thus pay higher prices on average. The opportunity cost of time is captured by income, number of children, age, and whether the consumer has a professional activity.

Finally, we present a model of consumer choice with sequential search costs and develop an empirical strategy to identify the parameters of the search costs distribution, which we estimate along with the parameters of the utility function. Products are considered to be heterogeneous with both vertical and horizontal attributes. Hence, consumers search for the product with the highest indirect utility instead of the lowest price.

As far as we are concerned, this is the first paper in the literature to identify search costs in a context of horizontally differentiated products. The horizontal dimension is particularly important when dealing with physical (not online) stores. In this context, the relative geographical location of the store, which is consumer specific, is clearly an important characteristic affecting choices, and ignoring this differentiation dimension will bias estimated parameters. Also, unlike previous methodologies, our's does not require any assumption on how firms set prices. This means that once we recover the demand parameter estimates, we can test between alternative models of firm behavior. In particular, we could test whether equilibrium prices are a result of pure or mixed strategy Nash equilibrium.

Reduced-form tests show that price dispersion is important in the French food market, even after controlling for unobserved store attributes. The price dispersion is also persistent over time. Stores frequently change positions in the cross-sectional distribution of prices, which is evidence in favor of firms playing mixed strategies. Moreover, there is evidence of a negative correlation between average price of the product and price dispersion, which is consistent with the idea that consumers have more incentives to search for high valued items since, due to a fixed cost component, search costs are relatively (to the high price of the product) lower in this case.

Periods of aggregate demand peaks, where search costs are expected to decrease, are also periods of lower price dispersion. This result indicates that search costs are a major cause of price dispersion. Furthermore, we find that prices paid increase with the opportunity cost of time, indicating that search costs are an important component of consumer behavior and that consumers who are time constrained search less intensively and end paying higher prices for identical products.

Results from the structural estimation show that consumers obtain at most three utility quotes before purchasing the product. The vast majority of consumers (more than 90%) do not search at all, purchasing the first product drawn.

The empirical analysis in all three chapters is performed using French home scan data. Household level information on store visits, purchases, and prices paid were collected during three years (1999, 2000, 2001) from a nationally representative survey. Data on household characteristics, including characteristics of the home and of the individuals composing the household, and store characteristics were also collected.

Chapter 1

Inventories, Unobservable Heterogeneity and Long Run Price Elasticities

1.1 Introduction

Consistently estimating demand is an empirical task of great importance in the area of Industrial Economics. Since information on production costs and wholesale prices is rarely available, the study of market structures requires the use of estimated preference parameters from which sample market shares and price cost margins can be recovered. The traditional empirical literature in the area has relied on standard static discrete choice models such as described in McFadden (1980, 1984). Examples are Berry, Levinsohn and Pakes (1995), who study the automobile industry, and Nevo (2001), Berto Villas-Boas (2007), and Bonnet and Dubois (2006) who estimate demand for ready-to-eat cereal, yogurt, and bottled water respectively, among many others.

While the static discrete-choice models are an interesting and relatively simple framework for studying individual demand, the recent literature casts doubts on the appropriateness of the resulting estimated price-elasticities for most demand applications. The study of mergers and market structure, for example, requires knowledge of long run responsiveness to price.

As noticed by Hendel and Nevo (2006a, 2006b), if consumer behavior includes stockpiling, for instance, then short and long run elasticities differ, and static models will yield short-run elasticities which overestimate the long-run measures.

A number of papers bring evidence that dynamics in the form of inventories are an important component of individual demand. If consumers stockpile then the decision to purchase today is affected by past prices and future expected prices. Reduced-form studies that test whether consumers stockpiling behavior is relevant include Boizot, Robin and Visser (2001) and Hendel and Nevo (2006a).

Hendel and Nevo (2006a) document purchasing patterns that are consistent with stockpiling behavior. More specifically, they find that duration since last sale positively affects aggregate quantity purchased, both during sales and non-sales periods. This is consistent with a model of stockpiling where consumers follow a s-S inventory rule. The longer the duration since previous sale, the closer consumers will on average be to the lower-bound inventory threshold, making purchase more likely. Furthermore, they find that indirect measures of storage costs are negatively correlated with the probability that households buy large quantities on sale. The authors also report a significant difference between sales and non-sales purchase in what concerns duration from previous purchase and duration to next purchase. Duration to previous purchase is shorter during sales periods than during non-sales. The idea is that during sales consumers will take advantage of the lower prices and buy at higher levels of current inventory, i.e., duration to previous purchase is shorter. Since inventories will be at higher levels, on average it will take longer for consumers to reach purchase threshold and duration to next purchase will therefore be longer.

A very similar model is used by Boizot et al. (2001). The only differences are that in the latter time is continuous and consumption is assumed to be exogenous and constant (actually, Hendel and Nevo also assume exogenous constant consumption when they estimate demand. The theoretical model itself however determines consumption endogenously). They perform reduced form empirical analysis, finding evidence that corroborates the consumer inventory theory.

More recently, some researchers have structurally estimated demand under stockpiling behavior of consumers. This is the case of Hendel and Nevo (2006b), Erdem, Imai, and Keane (2003), and Sun (2005).

Hendel and Nevo (2006b) structurally estimate a model of household demand for a storable product that incorporates the dynamics dictated by stockpiling behavior. Their goal is to assess and quantify the implications of stockpiling on demand estimation and, in particular, compare the resulting estimates to the ones obtained from standard static models. In the dynamic model, households purchase both for current consumption and for inventory building. Consumers increase inventory when the difference between current and expected future price is lower than the cost of holding inventory. To estimate the model, the authors use weekly scanner data on laundry detergents collected in nine supermarkets of a large U.S. mid-west city. The state space includes prices and advertising expenditures for all brands in all sizes of the products. Hendel and Nevo suggest an interesting method to reduce the state space: in their model, the probability of choosing a brand, conditional on quantity, does not depend on dynamic considerations. Therefore, a large number of parameters can be estimated from a static brand-choice model, without solving the dynamic program. Estimation follows an adjusted version of the "nested algorithm", as proposed by Rust (1987), where the value function is approximated by policy function iterations as suggested by Benitaz-Silva et al. (2000). Results suggest that ignoring dynamics has strong implications on demand estimation. The static model overestimates own price elasticities, underestimates cross-price elasticities to other products, and overestimates substitution to no purchase outside option. Resulting estimated price-cost margins using the figures yielded by standard static models will be biased downwards.

Close to the work of Hendel and Nevo is that of Erdem, Imai and Keane (2003). The main difference between the two papers is related to how they reduce the complexity of the state space. Erdem et al. assume that once consumption is determined, each brand in stock is consumed at a rate proportional to the share of that brand in storage. Their method is more flexible in modelling unobserved consumer heterogeneity but at a higher computational cost. Which method is more convenient will depend on the market or industry under study. In particular, for applications where the choice sets are large, the Erdem et al.'s method is difficult to apply.

Finally, Sun (2005) studies promotion effects on consumption, which is endogenous but not uncertain as in Hendel and Nevo (2006b). The shock to utility is assumed to be logistically distributed so that product choice probabilities are multinomial logit. To solve the dynamic program, Sun adopts simulated maximum likelihood techniques employing Monte Carlo methods (Keane, 1993) in addition to the interpolation method (Keane and Wolpin, 1994) to estimate parameters. The model is applied to individual purchases of packaged tuna and yogurt.

Notice that a common aspect of the models above is that estimation of structural parameters require solving the dynamic program, which is a complicated numerical problem, as well as very costly in terms of computer time (see Rust, 1996). In this paper we develop a way of structurally estimating long run price responses of consumers under inventory behavior without having to solve the dynamic program. That is one of the main contributions of our work. We use a model of dynamic consumer demand similar to Hendel and Nevo (2006a)'s model, where an additional assumption on consumption permits not only to empirically test the validity of the model but also to estimate the structural parameter needed to calculate the long run price elasticities.

Our methodology is extremely flexible with respect to consumer unobserved heterogeneity: the main estimated parameter is household-specific, thus yielding household-specific price elasticities. Developing a method that fully accounts for heterogeneity is the second important contribution of this paper. We compare the long run and short run elasticities, finding that the short run measure overestimates long run elasticities by more than 80% on average, making it clear that considering consumer stockpiling behavior makes a significant quantitative difference in price elasticities. Results are robust to different price expectation hypothesis as well as different estimation methods.

We use a home scan survey. Household level information on store visits, purchases, and prices paid were collected during three years (1999, 2000, 2001) from a nationally representative survey. Data on household characteristics, including characteristics of the home and of the individuals composing the household, and store characteristics were also collected.

This work is organized as follows. In the next section, we present the inventory model and derive the testable implications, the purchase decision equations, and the long and short run demand price elasticities implied by the model. In section 3, we compare our methodology to that in Hendel and Nevo (2006b) and in Erdem et al. (2003). Data description as well as descriptive statistics for some of the variables used in the empirical analysis can be found in Section 4. Section 5 brings the econometric implementation and empirical results, while the sixth section checks the robustness of the results. Section 7 concludes.

1.2 The Model

Empirical studies of market competition need an unbiased estimate of the long-run demand price elasticity in order to calculate price cost margins when information on actual costs is not available. The usual discrete-choice static models of demand yield estimates of the short-term price elasticities, which will be different from the long-run elasticities if some dynamic component influences consumers' purchase choice. In particular, if consumers hold inventories, the static model will not yield the desired elasticities. To correct for this problem, we propose a dynamic demand model with random prices which takes into account the possibility that consumers stockpile. Making an assumption on consumption at periods without purchases, we are able not only to derive testable implications of the model but also to structurally estimate model parameters without having to solve the dynamic program. Finally, we derive the short and long-run demand price elasticities implied by the dynamic model and show they are indeed different.

1.2.1 Consumer Behavior with Inventories

The net per period utility of consumer i is equal to $u(c_{it})$, where u is an increasing and concave function of consumption at period t, c_{it}. At each period, consumer i must decide how much to purchase of a certain good, how much to consume, and how much to stock as inventory. The law of motion of inventories is:

$$y_{it} = q_{it} - c_{it} + x_{it-1} \tag{1.1}$$

where y_{it} is the end of the period level of inventories, and x_{it} is the beginning of the period level, i.e., before consumption c_{it} and purchases q_{it}. Notice that

$$y_{it-1} = x_{it}$$

The problem of the consumer i at any period t is:

$$\max_{\{c_{it},q_{it},y_{it}\}} E_t \left\{ \sum_{t=\tau}^{\infty} \delta^t \left[u\left(c_{it}\right) - \alpha_i p_t q_{it} - \Phi\left(y_{it}\right) \right] \right\} \tag{1.2}$$

$$s.t. \;\; y_{it} = q_{it} - c_{it} + x_{it} \quad (\lambda_{it})$$

$$y_{it} = x_{it-1}$$

$$q_{it} \geqslant 0 \quad (\Psi_{it})$$

$$y_{it} \geqslant 0 \quad (\mu_{it})$$

where α_i is the marginal utility of revenue, and the parameters between brackets are the Lagrange multipliers of each constraint. The function $\Phi\left(y_{it}\right)$ represents the cost of storing inventory, which is an increasing and convex function of inventories. Assume $\Phi\left(y_{it}\right) = \phi y_{it}^2$.

At the beginning of each period (before purchases) consumers learn the practised prices at the period. The Lagrangian of the problem is:

$$£ = \max_{\{c_{it},q_{it},y_{it}\}} E_t \left\{ \sum_{t=0}^{\infty} \delta^t \left[u\left(c_{it}\right) - \alpha_i p_t q_{it} - \phi y_{it}^2 + \lambda_{it}\left(q_{it} - c_{it} + x_{it} - y_{it}\right) + \Psi_{it} q_{it} + \mu_{it} y_{it} \right] \right\}$$

The first order conditions with respect to consumption at period t (c_{it}), purchase at period t (q_{it}), and end of the period inventory at period t (y_{it}), are, respectively:

$$\frac{\partial £}{\partial c_{it}} = 0 \Rightarrow u'\left(c_{it}\right) = \lambda_{it} \tag{1.3}$$

$$\frac{\partial \pounds}{\partial q_{it}} = 0 \Rightarrow \alpha_i p_t = \lambda_{it} + \Psi_{it} \tag{1.4}$$

$$\frac{\partial \pounds}{\partial y_{it}} = 0 \Rightarrow y_{it} = \frac{\delta E_t\left(\lambda_{it+1}\right)}{2\phi_i} - \frac{\lambda_{it}}{2\phi_i} + \frac{\mu_{it}}{2\phi_i} \tag{1.5}$$

We assume positive consumption every period and that consumers always expect future consumption to be positive. This assumption can be strong for certain product categories. But here, as we will see later, we only consider products that are consumed at regular basis and for which we expected households to consume a positive amount at every period.

Manipulating the first order conditions and given the assumptions above, we get our main result:

Proposition 1 *In periods with purchases ($q_{it} > 0$), if the price is higher than the discounted expected price for the following period ($p_t > \delta E_t\left(p_{t+1}\right)$), the utility maximizing end of the period inventory level is equal to zero ($y_{it} = 0$). Furthermore, in this case, next period purchased quantity (q_{it+1}) will be positive. On the other hand, if $p_t \leq \delta E_t\left(p_{t+1}\right)$, then $y_{it} > 0$ in periods with purchases and periods without purchases, and next period purchased quantity can be either positive or equal to zero.*

Proof. When $q_{it} > 0$ ($\Psi_{it} = 0$), we have: $y_{it} = \frac{\alpha_i(\delta E_t(p_{t+1}) - p_t)}{2\phi_i} - \frac{E_t \Psi_{it+1}}{2\phi_i} + \frac{\mu_{it}}{2\phi_i}$. Assume $y_{it} > 0$ when $p_t > \delta E_t\left(p_{t+1}\right)$. Then, $\mu_{it} = 0$ and $y_{it} = \frac{\alpha_i(\delta E_t(p_{t+1}) - p_t)}{2\phi_i} - \frac{E_t \Psi_{it+1}}{2\phi_i}$, which implies $y_{it} < 0$ (since $p_t > \delta E_t\left(p_{t+1}\right)$ and $\frac{E_t \Psi_{it+1}}{2\phi_i} \geqslant 0$), which contradicts $y_{it} > 0$. Thus $y_{it} = 0$ when $p_t > \delta E_t\left(p_{t+1}\right)$ and $q_{it} > 0$.

When $y_{it} = 0$, next period purchase is going to be equal to: $q_{it+1} = y_{it+1} + c_{it+1}$. The end of the period inventory level is non-negative, while consumption is assumed to be positive at every period. Therefore, if $y_{it} = 0$, $q_{it+1} > 0$.

Now, take periods when $q_{it} > 0$ ($\Psi_{it} = 0$) and $p_t \leq \delta E_t\left(p_{t+1}\right)$. Assume that $\delta E_t\left(p_{t+1}\right)$, in this case, $y_{it} = 0$ ($\mu_{it} > 0$). Then $y_{it} = \frac{\alpha_i(\delta E_t(p_{t+1}) - p_t)}{2\phi_i} - \frac{E_t \Psi_{it+1}}{2\phi_i} + \frac{\mu_{it}}{2\phi_i}$. Now, $E_t \Psi_{it+1}$ will be greater than zero if and only if the expected purchase at period $t+1$ (q_{it+1}) is equal to zero. When $y_{it} = 0$, the expected purchase next period is equal to: $E_t(q_{it+1}) = E_t\left(y_{it+1} + c_{it+1}\right) = E_t\left(y_{it+1}\right) + E_t(c_{it+1})$. Notice that this expectation is going to be greater than zero since we assume consumers always expect next period consumption to be positive and since end of the period inventories are non-negative. But if $E_t(q_{it+1}) > 0$, then $E_t\Psi_{it+1} = 0$ and $y_{it} = \frac{\alpha_i(\delta E_t(p_{t+1})-p_t)}{2\phi_i} + \frac{\mu_{it}}{2\phi_i}$, which implies $y_{it} > 0$ (since $p_t \leq \delta E_t(p_{t+1})$ and $\mu_{it} > 0$), contradicting $y_{it} = 0$. Therefore, it must be that $y_{it} > 0$ in periods when purchases are positive and the price is lower than the regular price.

Now, let $q_{it} = 0$ ($\Psi_{it} > 0$) and $p_t \leq E_t(p_{t+1})$. Assume $y_{it} = 0$. Then $\mu_{it} > 0$ and $E_t\Psi_{it+1} = 0$, and $y_{it} = \frac{\alpha_i(\delta E_t(p_{t+1})-p_t)}{2\phi_i} + \frac{\Psi_{it}}{2\phi_i} + \frac{\mu_{it}}{2\phi_i}$, which implies $y_{it} > 0$, contradicting the initial hypothesis $y_{it} = 0$. Thus $y_{it} > 0$ when $q_{it} = 0$ and $p_t \leq \delta E_t(p_{t+1})$.

The next period purchase is going to be equal to: $q_{it+1} = y_{it+1} - x_{it+1} + c_{it+1}$, which is equal to zero if $y_{it+1} \geqslant y_{it} - c_{it+1}$, and greater than zero otherwise. ■

Although prices are random, we assume consumers always expect prices to return to their regular level p_r (including the discount[1]), which is equal to the mean price they pay for the good. At first view, this may seem a strong assumption, given that stores are known to frequently offer discounted prices. However, notice that the regular price, defined as the mean price paid by the consumer across all her purchase occasions, already incorporates discount prices and their probabilities. Therefore, what may be controversial in our hypothesis of expected next-period price being always equal to regular price is the assumption that future prices expectations are independent of price realizations today[2]. Our assumption is valid if consumers are unable to predict the timing of sales. Indeed, a number of theoretical studies show that in an important number of situations sales are necessarily random. See, for example, Braido (2009) and references therein.

Anyway, we do not need the assumption that price expectations are equal to the observed average price. For instance, we could assume price expectations are rational and estimate a Markov price process. The rational expectation hypothesis would then imply that expected prices would be equal to the predicted price process. Although in general this may be a preferred

[1] Since the time period we consider is short (one week), we let the discount rate be very close to 1. So $E_t(p_{t+1}) = \delta^{-1}p_r \simeq p_r$, where p_r is the regular price.

[2] The standard assumption (see Hendel and Nevo, 2006) is that price expectations are first order Markov.

solution, we only do that in the robustness check section of the paper. We chose an alternative route of action because we are not sure how much we can trust our price process estimates since we do not observe prices, only prices paid by the households. But if our model is to be applied to a database which brings enough information to assure consistent estimates of the price process, there is no reason why the rational expectation hypothesis should not be used.

Finally, notice that the absence of "precautionary" stocking is not implied by the assumptions on expectations. Rather, it is due to the assumption that preferences are quasi-linear, which is quite a standard assumption in the literature without which it would be much harder to estimate model parameters. The quasi-linearity of preferences imply that the marginal utility of consumption is separable from the marginal utility of income and linear in prices. Hence, there is no concavity which would trigger precautionary behavior.

From the proposition above, we know that, given purchases today at non-discounted prices ($p_t > p_r$), consumers will certainly purchase next period. On the other hand, if they purchase today at a lower than regular price, they will hold positive inventories at the end of the period and they may not have to purchase next period. Thus, the probability of purchasing next period is higher when the consumer purchases today at higher prices. This is the idea behind the first implication of the model:

Implication 1 Conditional on inventories, duration until next purchase is higher at discounted price periods.

Purchases are going to be positive when $q_{it} = y_{it} - x_{it} + c_{it} > 0$. Hence, the level of inventory at the beginning of the period, x_{it}, that triggers purchase is $\tilde{x}_{it} = y_{it} + c_{it}$, which is lower at discounted prices than at regular or higher prices since both y_{it} and c_{it} decrease with current prices. Thus:

Implication 2 Conditional on inventories, duration from last purchase is lower at periods of discounted price.

Moreover, the higher the marginal cost of holding inventories ϕ_i, the lower the chosen inventory level y_{it}, and the lower $\tilde{x}_{it}$ that triggers purchases. Therefore, the lower ϕ_i, the higher the frequency of purchases. If we compare two households, l and j, such that $\phi_l > \phi_j$, we should expect household l to purchase more frequently than household j:

Implication 3 Conditional on inventories, households with a high marginal cost of holding inventories purchase more frequently than households with low marginal costs.

Of course, we do not observe the marginal cost of holding inventories. However, we can make some reasonable assumptions on the ordering of the marginal costs. Take the purchase of butter, for example, which must be stocked in the refrigerator. Then stocking butter is certainly more costly for a household who has a refrigerator than for a household who does not. In general, an important part of the cost of stockpiling is the space cost. The less space a household has available for stocking, the higher the marginal costs of stocking. Thus, households that live in bigger houses probably have lower marginal costs of holding inventories. We can therefore test Implication 3 using some observable characteristics of the households' homes as indirect measures of space availability.

Notice that testing the above implications is equivalent to testing the dynamic model of consumer decision. The alternative hypothesis is the static model, where duration is independent of prices because a price variation will be completely translated into consumption variation, and will not affect decisions in other periods.

1.2.2 The Purchase Decision

From Proposition 1, we know that $\Pr(y_{it} = 0 \mid q_{it} > 0, p_t > p_r) = 1$. Thus

$$E_t\left(q_{it} \mid q_{it} > 0, p_t > p_r, y_{it} = 0\right) = E_t\left(q_{it} \mid q_{it} > 0, p_t > p_r\right)$$

Furthermore, if $y_{it} = 0$ then $q_{it} = c_{it} - x_{it}$, where $c_{it} = h(\alpha_i p_t)$ and $h = u'^{-1}$. Therefore, conditional on purchases and prices being higher than regular:

$$\begin{aligned} E_t\left(q_{it} \mid q_{it} > 0, p_t > p_r\right) &= E_t\left(c_{it} - x_{it} \mid q_{it} > 0, p_t > p_r\right) \qquad (1.6) \\ &= E_t\left(h\left(\alpha_i p_t\right) - x_{it} \mid q_{it} > 0, p_t > p_r\right) \end{aligned}$$

When $p_t \leq p_r$ and $q_{it} > 0$ ($\Psi_{it} = 0$) on the other hand, y_{it} is always positive ($\mu_{it} = 0$). In this case, we have $q_{it} = y_{it} - x_{it} + c_{it}$, where $y_{it} = \frac{\alpha_i(p_r - p_t)}{2\phi_i} - \frac{E_t \Psi_{it+1}}{2\phi_i}$ and $c_{it} = h(\alpha_i p_t)$. Hence

$$E_t\left(q_{it} \mid q_{it} > 0, p_t \leq p_r\right) = E_t\left(\frac{\alpha_i\left(p_r - p_t\right)}{2\phi_i} - \frac{E_t \Psi_{it+1}}{2\phi_i} + h\left(\alpha_i p_t\right) - x_{it} \mid q_{it} > 0, p_t \leq p_r\right) \tag{1.7}$$

Now, assume that at each t we observe the purchased quantity q_{it} with an error v_{it}. Substituting the observed quantity $q_{it}^{*} = q_{it} - v_{it}$ into (1.6) and (1.7), we get:

$$\begin{aligned} E_t\left(q_{it}^{*} \mid q_{it} > 0, p_t > p_r\right) &= E_t\left(q_{it} - v_{it} \mid q_{it} > 0, p_t > p_r\right) \\ &= E_t\left(c_{it} - x_{it} - v_{it} \mid q_{it} > 0, p_t > p_r\right) \end{aligned} \tag{1.8}$$

and

$$\begin{aligned} E_t\left(q_{it}^{*} \mid q_{it} > 0, p_t \leq p_r\right) &= E\left(q_{it} - v_{it} \mid q_{it} > 0, p_t \leq p_r\right) \\ &= E_t\left(\frac{\alpha_i\left(p_r - p_t\right)}{2\phi_i} - \frac{E_t \Psi_{it+1}}{2\phi_i} + h\left(\alpha_i p_t\right) - x_{it} - v_{it} \mid q_{it} > 0, p_t \leq p_r\right) \end{aligned} \tag{1.9}$$

The observable variables in (1.8) are q_{it}^{*}, the regular price p_r, and the price at period t, p_t. Although the beginning of the period level of inventories x_{it} is not directly observable, we know, from the law of motion of inventories that:

$$\begin{aligned} x_{it} &= x_{i1} + \sum_{n=1}^{t-1} q_{in} - \sum_{n=1}^{t-1} c_{in} \\ &= x_{i1} + \sum_{n=1}^{t-1} q_{in}^{*} - \sum_{n=1}^{t-1} c_{in} - \sum_{n=1}^{t-1} v_{in} \end{aligned}$$

Moreover, we know from Proposition 1 that at periods t' with purchases ($q_{it'} > 0$) and price higher than regular price ($p_{t'} > p_r$), the chosen end of the period inventory of consumer i is going to be equal to zero ($y_{it'} = 0$), and thus at the beginning of the next period, inventories ($x_{it+1'}$) will also be equal to zero. Or, more formally, if at $t_0(i) < t, q_{i0(i)} > 0$ and $p_{0(i)} > p_r$, then $y_{i0(i)} = 0$ and $x_{i1(i)} = 0$. Therefore, we will consider household i's period zero as the period $t_1(i)$ immediately following period $t_0(i)$,which is the first period when prices are higher than regular and household i purchases. In that way, we are sure that x_{i1} is equal to zero and we can write:

$$x_{it} = \sum_{n=t_1(i)}^{t-1} \left(q_{in}^* - c_{in} - v_{in}\right) \tag{1.10}$$

where we observe q_{it}^* for all t and can thus calculate $\sum_{n=t_1(i)}^{t-1} q_{in}$.

In what concerns consumption, its utility maximizing level at periods $\tilde{t}$ with purchases $(q_{i\tilde{t}} > 0)$, $\tilde{t} \in \{1, 2...\}$, is $c_{i\tilde{t}} = h\left(\alpha_i p_{\tilde{t}}\right)$. However, for periods $\bar{t}$ without purchases $(q_{i\bar{t}} = 0)$, we only know that $c_{i\bar{t}} = x_{i\bar{t}} - y_{i\bar{t}}$, which we cannot use in (1.10) to calculate beginning of the period inventories. We assume, therefore, that at periods without purchase, consumption will be equal to consumption at regular prices , that is $c_{i\bar{t}} = h\left(\alpha_i p_r\right)$.

Utility Specification

Assume $u\left(c_{it}\right) = -\left(1/\rho\right)\exp\left[-\rho c_{it}\right]$, where ρ is a positive parameter, so that $h\left(\alpha_i p_t\right) = -\frac{1}{\rho}\ln \alpha_i - \frac{1}{\rho}\ln p_t$. Let $T_{i1}^{t-1} \in \{t_1(i), ..., t-1\}$ be the set of periods where consumer i purchased, and $T_{i0}^{t-1} \in \{t_1(i), ..., t-1\}$, the set of periods where i does not purchase. Then:

$$\sum_{n=t_1(i)}^{t-1} c_{in} = -\frac{1}{\rho}\left(T_i^{t-1}\ln\left(\alpha_i\right) + \sum_{n \in T_{i1}^{t-1}} \ln p_n + T_{i0}^{t-1}\ln p_r\right) \tag{1.11}$$

where T_i^t is the total number of periods for household i $(T_i^t = T_{i1}^{t-1} + T_{i0}^{t-1} + 1 = t - t_1(i))$.

Substituting (1.10) and (1.11) into the purchased quantity equations (1.8 and 1.9) yields, respectively:

$$E_t\left(q_{it}^* \mid q_{it} > 0, p_t > p_r\right) = -\frac{1}{\rho}T_i^{t-1}\ln\left(\alpha_i\right) - \frac{1}{\rho}\left[\ln p_t + \sum_{n \in T_{i1}^{t-1}} \ln p_n + T_{i0}^{t-1}\ln p_r\right] - \sum_{n=t_1(i)}^{t-1} q_{in}^* + E_t\left(\sum_{n=t_1(i)}^{t} v_{in} \mid q_{it} > 0, p_t > p_r\right)$$

and

$$E_t\left(q_{it}^* \mid q_{it} > 0, p_t \leq p_r\right) = \frac{\alpha_i}{2\phi_i}\left(p_r - p_t\right) + \frac{1}{\rho}T_i^t \ln\left(\alpha_i\right) + \frac{1}{\rho}\left[\ln p_t + \sum_{n \in T_{i1}^{t-1}} \ln p_n + T_{i0}^{t-1}\ln p_r\right] - \sum_{n=t_1(i)}^{t-1} q_{in}^* + E_t\Psi_{it+1} + E_t\left(\sum_{n=t_1(i)}^{t} v_{in} \mid q_{it} > 0, p_t \leq p_r\right)$$

Let $Q_{it}^* = \sum\limits_{n=t_1(i)}^{t} q_{in}^*$. Moving $\sum\limits_{n=t_1(i)}^{t-1} q_{in}^*$ to the left hand side, we get:

$$E_t\left(Q_{it}^* \mid q_{it} > 0, p_t > p_r\right) = -\frac{1}{\rho}T_i^{t-1}\ln\left(\alpha_i\right) - \frac{1}{\rho}\left[\ln p_t + \sum_{n \in T_{i1}^{t-1}} \ln p_n + T_{i0}^{t-1}\ln p_r\right] + E_t\left(\sum_{n=t_1(i)}^{t} v_{in} \mid q_{it} > 0, p_t > p_r\right) \tag{1.12}$$

and

$$E_t\left(Q_{it}^* \mid q_{it} > 0, p_t \leq p_r\right) = \frac{\alpha_i}{2\phi_i}\left(p_r - p_t\right) + \frac{1}{\rho}T_i^t \ln\left(\alpha_i\right) + \frac{1}{\rho}\left[\ln p_t + \sum_{n \in T_{i_1}^{t-1}} \ln p_n + T_{i0}^{t-1}\ln p_r\right] + E_t\Psi_{it+1} + E_t\left(\sum_{n=t_1(i)}^{t} v_{in} \mid q_{it} > 0, p_t \leq p_r\right) \tag{1.13}$$

We assume that the errors v_{it} are mean independent of q_{it}, p_t, T_{i0}^{t-1}, T_{i1}^{t-1}, and p_r for all t, which implies that $E_t\left(\sum_{n=t_1(i)}^{t} v_{in} \mid q_{it} > 0, p_t \leq p_r\right) = E_t\left(\sum_{n=t_1(i)}^{t} v_{in}\right) = 0$.

This means that we are able to estimate the marginal utility of income α_i and the parameter ρ using (1.12). Unfortunately, we cannot estimate ϕ_i in (1.13) because of $E_t\Psi_{it+1}$, which is an unknown function varying on period t and on household i. However, α_i is the most important parameter to be estimated when we are ultimately interested in price elasticities because, as will be seen in the next subsection, the marginal utility of income is the model parameter needed to calculate the long run price elasticity.

Instead of a CARA utility, Hendel and Nevo (2006) consider a logarithm utility function ($u(c_{it}) = \ln(c_{it})$). But in our case the logarithm utility function is too restrictive in terms of price elasticities because in our model they are always equal to 1. Examples of alternative utility functions can be found in Sun (2005), where utility is a second-degree polynomial of consumption, and Erdem et al. (2003), where utility is linear in consumption.

1.2.3 Price Elasticities: Long Run versus Short Run

If consumers stockpile, short run and long run price elasticities will differ. The long run price-elasticity should only take into account the effect of a price variation on consumption, not in purchases in a certain period, since part of the variation in purchases in a certain period will be due to variation in stocks[3]. The long run price elasticity is therefore the price elasticity of consumption in the inventory model, where purchase and consumption are not the same. What we call the short run price elasticity, on the other hand, measures the responsiveness of purchases to variation in prices. It can be calculated as the price elasticity of purchases in the inventory model, or as the price elasticity of demand in a static model where purchases and consumption are equal at every period.

In this subsection, we compare short (purchases) and long run (consumption) price responses in the inventory model, showing that these two measures are not the same. We also show the expressions for the short run price elasticities yielded by the static model of demand.

Short Run The short run price-elasticity of demand will capture the effect of a variation in prices on the purchased quantity, which in the presence of stockpiling behavior is not necessarily equal to consumption. Hence, the short run price elasticity is actually the price elasticity of purchase in the inventory model and, as shown in the Appendix, it is equal to

$$\begin{aligned}
\epsilon_{it}^{SR_d} = & -\frac{h(\alpha_i p_t)}{(V_{it}-v_{it})}\Pr(p_t>p_r)\left[\Phi(V_{it}\mid p_t>p_r)+(V_{it}-v_{it})\Phi'(V_{it}\mid p_t>p_r)\right] \\
& -\frac{p_t}{V_{it}-v_{it}}\left(\frac{\alpha_i}{\theta_i}+\frac{1}{\alpha_i p_t^2}\right)\Pr(p_t\le p_r)\left[\Phi(U_{it}\mid p_t\le p_r)+(U_{it}-v_{it})\Phi'(U_{it}\mid p_t\le p_r)\right] \\
& +p_t\left[\Phi(V_{it}\mid p_t>p_r)-\Phi(U_{it}\mid p_t\le p_r)\right]\frac{d\Pr(p_t>p_r)}{dp_t} \qquad (1.14)
\end{aligned}$$

[3] See Hendel and Nevo (2006a, 2006b). A similar argument is developed in Feenstra and Shapiro (2001).

where Φ is the cumulative distribution function of $\sum_{n=t_1(i)}^{t} v_{in}$, and V_{it} and U_{it} are, respectively:

$$V_{it} = h(\alpha_i p_t) - x_{it}$$

and

$$U_{it} = \frac{\alpha_i}{2\phi_i}(p_r - p_t) - E_t \Psi_{it+1} + h(\alpha_i p_t) - x_{it}$$

As can be seen below, the expression for the short run (purchases) price elasticity is different from the expression for the long run elasticity of demand (consumption) under stockpiling behavior.

Unfortunately, we are unable to compute or estimate the elasticity in (1.14). However, we are able to estimate the short run elasticity implied by a static model of demand behavior. We would like to compare the measures thus obtained with the long run elasticities resulting from the inventory model. The static (short run) elasticities are obtained from a model identical to the inventory model described above, with the exception that dynamics are now ignored. Thus, consumers choose today whether to purchase and how much to purchase taking into account only the current price and the utility realization shock. Furthermore, in the static model, quantity purchased is equal to quantity consumed since there is no stockpiling. Considering $u(c_{it}) = -(1/\rho)\exp[-\rho c_{it}]$ leads to the following first order condition:

$$c_{it} = -\frac{1}{\rho}\ln\beta_i - \frac{1}{\rho}\ln p_t \tag{1.15}$$

where β_i is the marginal utility of income in the static model and c_{it} equals quantity purchased. Equation (1.15) implies that the price elasticity in the static model is thus:

$$\epsilon_{it}^{SR_s} = \frac{1}{\ln\beta_i + \ln p_t} \tag{1.16}$$

The parameter β_i can be identified in (1.15). The estimated parameters can then be plugged into (1.16) to obtain measures of the short price elasticities.

Long Run Inventories are a form of intertemporal substitution but in the long run, everything which is purchased will be consumed, since it is from consumption that the individual extracts utility. Therefore, the long run purchased quantity, or the long run demand, depends only on consumption, not on inventories. Hence, a measure of the long run price elasticity should take into account only the effect of prices on consumption. We shall consider the effect of prices on consumption even when there are no purchases because we assume consumption is positive in every period.

When $u(c_{it}) = -(1/\rho)\exp[-\rho c_{it}]$, the long run price elasticity (the consumption price elasticity in the inventory model) of individual i at period t is:

$$\begin{aligned}\epsilon_{it}^{LR} &= \frac{dc_{it}}{dp_t}\frac{p_t}{c_{it}} \\ &= \frac{1}{\rho p_t}\frac{p_t}{\frac{1}{\rho}(\ln\alpha_i + \ln p_t)} \\ &= \frac{1}{\ln\alpha_i + \ln p_t}\end{aligned} \tag{1.17}$$

Thus, to calculate the long run price elasticity implied by the inventory model, we only need to identify one of the parameters of the model, the α_i. The estimated values can then be directly plugged into (1.17) to obtain household specific measures of the long run price elasticities.

Notice that the price elasticities of the dynamic model have exactly the same functional form as in the static model (equation 1.16). What will differ between the short and long run measures is the estimated coefficients for the marginal utility of income ($\hat{\beta}_i$ in the static model and $\hat{\alpha}_i$ in the dynamic model).

1.3 Comparison with other Methods

In this section, we compare our work to other studies that structurally estimate parameters of the inventory model, namely Erdem et al. (2003) and Hendel and Nevo (2006b)[4]. We are mainly

[4] We chose to compare our work to those two paper because we believe they represent the state of the art in what concerns the study of consumer inventory behaviour. Another paper that structurally estimate demand parameters is Sun (2005). Reduced form estimates can be found in Ailawadi and Neslin (1998), and Boizot et al. (2001).

interested in comparing the characteristics of the different models in what concerns simplicity of the estimation method, flexibility of consumer heterogeneity, and product differentiation.

Hendel and Nevo's inventory model is very similar to ours: per period utility is a concave function, there is no stock out or purchase costs, holding inventory is costly, and prices are random. An important difference is that they consider brand choice. Product differentiation takes place only at the time of purchase. Literally, product differences affect the behavior of the consumer at the store but do not give different utilities at the time of consumption. This assumption reduces the state space because instead of the whole vector of brand inventories, only the total quantity in stock matters. Hence, they are able to separate the product (brand) and quantity decisions. Their approach leads to an important computational simplification, which is the main contribution of the paper. However, their model is very restrictive in terms of observable consumer heterogeneity, and it does not allow for unobservable heterogeneity, which would break down the complete separation of brand and quantity.

A less important difference between Hendel and Nevo's model and ours is related to the random term. In their model, the randomness is included as a preference shock . That is, per period utility is a function not only of consumption but also of an additive random shock, $u(c_{it} + v_{it})$. Equations to be estimated are exactly the same whether we consider a preference shock on consumption as they do or a measurement error on purchase as we do, only interpretation changes. However, in our model, if we consider v_{it} to be a preference shock on consumption, at each period the decision to purchase a positive quantity will depend on that period's preference shock. Therefore, in the purchase decision equations (1.12) and (1.13), the random component (which includes past shocks) would be correlated to the number of periods the household decided to purchase (T_{i1}) and with the number of periods the household decided not to purchase (T_{i0}), creating an endogeneity problem.

In Erdem et al.'s model, the consumption function is linear and consumers have an exogenous stochastic per period usage requirement for the good, which is only revealed after the purchase decision is made. Thus consumers run a risk of stocking out, which is costly, if they maintain an inadequate inventory to meet the usage requirement. Notice that the usage rate assumption means that consumption is independent of prices in the short run. However, if prices remain high for a long period, consumption will adjust accordingly through more frequent stock outs.

To reduce the complexity of the state space, they assume that once quantity to be consumed is determined, each brand in storage is consumed at a rate which is proportional to the share of that brand in storage. Together with the assumption that brand differences enter linearly in the utility function, it implies that only the total inventory and a quality weighted inventory matter as state variables.

Finally, they incorporate consumer heterogeneity by allowing for 16 types of consumers which differ in terms of taste for the brands and in terms of usage rates.

The approach in Erdem et al. is computationally more complicated than that on Hendel and Nevo and consumption is exogenous. However, it allows for some degree of unobservable heterogeneity.

The main drawback of our model is not considering product differentiation. Another weakness is the ad hoc hypothesis on consumption when there is no purchase. However, these restrictions are counterbalanced by an extremely flexible consumer heterogeneity structure and a very simple and fast way of computing structural estimates. Furthermore, in what concerns consumption, our assumption is not stronger than the usage rate hypothesis of Erdem et al. Indeed, in our case, consumption always responds to prices at periods with purchases. The way consumption in periods without purchases react to prices is similar to the Erdem et al.'s, i.e., adjustment happens following long term price changes (for instance, if prices increase and remain high for a long time, consumption without purchases will go down through the increase in regular prices).

1.4 Data

The database is a representative survey of households distributed across all regions of France. We use information on three years: 1999, 2000, and 2001. Each household was given a scanner with which to register every food product purchased. For each product purchased, we have information on its brand and characteristics, including price and pack size, the date of the purchase and the brand of the retailer where it was purchased. We also have comprehensive information on household demographics, and on home characteristics, such as if the household has a storage room, a bathroom, a fridge, pets etc.

In the database, one observation is one purchase made by the household. For each product category under study, we consider a sub-sample of households that purchased that product category at least once in the three years. The product categories that we study are milk, coffee, canned tuna, pasta, yogurt, and butter. Table 1.1 through 1.6 bring descriptive statistics on these product categories, including number of households that purchased the product at least once during the three-year time span, total quantity purchased, average quantity per purchase, average interpurchase duration, and average price paid per pack-size.

Table 1.1: Descriptive Statistics - Butter

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
number of households	6695	-	-	-	372869
total qty (kg)	151743.488	-	-	-	372869
qty/purchase (kg)	0.407	264.394	0.025	10.000	372869
avg duration (days)	15.99	26.91	0	1057	366174
price (€/kg)	4.848	0.72	0.899	119.520	372869
price/pack (€)	4.848	0.03	3.872	5.564	372869

Table 1.2: Descriptive Statistics - Yogurt

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
number of households	6814	-	-	-	646560
total qty (kg)	759114.624	-	-	-	646560
mean day (kg)	1.174	686.825	0	16.000	
avg duration (days)	9.60	17.62	0	903	639746
price (€/kg)	2.028	0.61	0.107	17.623	646532
price/pack (€)	2.028	0.35	1.204	4.025	646560

The choice of product categories was made according to three criteria. First of all, we chose products that are consumed in a regular basis. Our model does not apply for products that are infrequently consumed. Second, we chose products that differ in terms of storage costs. While butter and yogurt need to be stocked in a refrigerated area, this is not the case for tuna, coffee,

Table 1.3: Descriptive Statistics - Coffee

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
number of households	6548	-	-	-	231609
total qty (kg)	119226.560	-	-	-	231609
qty/purchase (kg)	0.514	374.540	0	12.000	231609
avg duration (days)	24.17	37.42	0	1071	225061
price (€/kg)	7.211	6.17	0.610	16.891	231597
price/pack (€)	7.211	0.91	1.387	3.765	231609

Table 1.4: Descriptive Statistics - Milk

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
number of households	6741	-	-	-	432267
total qty (kg)	2372314.368	-	-	-	432267
qty/purchase (kg)	5.488	4890.45	0	216.000	432267
avg duration (days)	14.43	20.07	0	1078	425526
price (€/kg)	0.640	0.21	0.061	16.007	432264
price/pack (€)	0.640	0.05	0.503	0.69	432267

Table 1.5: Descriptive Statistics - Pasta

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
number of households	6834	-	-	-	330253
total qty (kg)	231704.048	-	-	-	330253
qty/purchase (kg)	0.702	496.101	0	16.000	330253
avg duration (days)	18.62	32.16	0	973	323419
price (€/kg)	1.753	0.71	0.229	24.392	330047
price/pack (€)	1.753	0.02	1.418	2.256	330253

Table 1.6: Descriptive Statistics - Tuna

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
number of households	6598	-	-	-	124859
totql qty (kg)	38255.404	-	-	-	124859
qty/purchase (kg)	0.306	206.130	0.054	8.760	124859
avg duration (days)	40.06	66.5	0	1026	118261
price (€/kg)	6.921	2.26	0.335	89.244	124859
price/pack (€)	6.921	0.26	4.695	8.187	124859

and pasta. Butter and yogurt are thus more costly to stock than tuna, coffee or pasta. Milk is more costly to stock than tuna since it requires more space per pack etc. Third, we chose products that differ in terms of storability, or how long or well a product can be stored. Pasta and tuna can be stored for a longer period than coffee which can be stored for a lot longer than yogurt, for instance[5].

Moreover, we took into account potential measurement errors arising from the fact that we use a broad definition of product. We consider each category as a single product, capturing the fact that different brands are substitutes (although not necessarily perfect substitutes). If, however, the consumption of one of the category brands is, for a certain household, independent from the consumption of another brand, then by treating both brands as substitutes, we introduce measurement error in the definition of inter-purchase duration and underestimate the true effects. To try to avoid these, we chose categories which are relatively homogeneous, increasing the probability that different brands will be substitutes. An exception is, perhaps, yogurt. Yogurt is sold in different brands and pack sizes, but the main differentiation is between plain yogurt and non plain. It is not clear that all households will regard plain and, for instance, fruit yogurt as substitutes. In general, we expect more homogenous product categories to present stronger evidence consistent with the dynamic model than less homogenous product categories.

Table 1.7 through 1.12 present, for each product category sample, descriptive statistics of the household characteristics included in estimations as controls for observable heterogeneity.

Table 1.7: Descriptive Statisticts of Characteristics of Households who buy Butter

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**	**N**
home has a cellar	0.21	0.43	0	1	372869
house (1) vs apartment (0)	0.71	0.45	0	1	372507
car ownership	0.94	0.24	0	1	372869
household size	3.15	1.40	1	9	372869
responsible for purchases is a man	0.03	0.17	0	1	372869

[5] It would have been interesting to apply the model to a product that is not at all storable or that has an infinite storage cost. However, we could not find product categories presenting those characteristics.

Table 1.8: Descriptive Statisticts of Characteristics of Households who buy Yogurt

Variable	Mean	Std. Dev.	Min.	Max.	N
home has a cellar	0.21	0.41	0	1	646560
house (1) vs apartment (0)	0.67	0.47	0	1	645931
car ownership	0.94	0.24	0	1	646560
household size	3.21	1.39	1	9	646560
responsible for purchases is a man	0.03	0.16	0	1	646560

Table 1.9: Descriptive Statisticts of Characteristics of Households who buy Coffee

Variable	Mean	Std. Dev.	Min.	Max.	N
home has a cellar	0.23	0.42	0	1	231609
house (1) vs apartment (0)	0.71	0.45	0	1	231395
car ownership	0.94	0.24	0	1	231609
household size	3.07	1.38	1	9	231609
responsible for purchases is a man	0.03	0.17	0	1	231609

Table 1.10: Descriptive Statisticts of Characteristics of Households who buy Milk

Variable	Mean	Std. Dev.	Min.	Max.	N
home has a cellar	0.19	0.39	0	1	432267
house (1) vs apartment (0)	0.66	0.47	0	1	431405
car ownership	0.93	0.26	0	1	432267
household size	3.23	1.43	1	9	432267
responsible for purchases is a man	0.03	0.17	0	1	432267

Table 1.11: Descriptive Statisticts of Characteristics of Households who buy Pasta

Variable	Mean	Std. Dev.	Min.	Max.	N
home has a cellar	0.19	0.4	0	1	330253
house (1) vs apartment (0)	0.70	0.46	0	1	329878
car ownership	0.95	0.22	0	1	330253
household size	3.47	1.39	1	9	330253
responsible for purchases is a man	0.02	0.15	0	1	330253

Table 1.12: Descriptive Statisticts of Characteristics of Households who buy Tuna

Variable	Mean	Std. Dev.	Min.	Max.	N
home has a cellar	0.19	0.39	0	1	124859
house (1) vs apartment (0)	0.68	0.47	0	1	124732
car ownership	0.94	0.24	0	1	124859
household size	3.28	1.4	1	9	124859
responsible for purchases is a man	0.03	0.18	0	1	124859

1.5 Econometric Implementation and Empirical Results

1.5.1 Model Implications

The first two implications derived from the model relate duration (from last purchase and to next purchase) and prices. *Implication 1* states that duration from last purchase is lower if the price today is lower than regular. *Implication 2*, on the other hand, states that duration until next purchase is higher if the current price is lower than the regular price. Therefore, in a regression of duration from last purchase (until next purchase) on a dummy indicating if current price is lower then the regular price, the coefficient of the dummy should be negative (positive).

Regular price is defined as the mean price per pack size paid by the household during the three years considered. Therefore, each household i has a regular price p_{ris}, where s is the pack size. By defining regular price per household, we hope to be partly controlling for the relevant consumption set of the household, since what matters for the consumer when deciding to purchase and to stock is a lower or higher price than the one she is used to pay, not the mean price paid by all households. Suppose for instance that consumer i never buys store brands, which are usually cheaper. Then the price of store brand products should not affect her purchase and quantity decisions because store brands do not enter her consumption set. Analogously, if consumer i never shops at store A (too far away from home, for example) then the average price of products in store A should not affect her decisions, for the same reason as prices in northern France, for instance, should not enter the regular price index of consumers living in the South of France.

The definition of regular price at the household level, however, has its problems. Households that have a high cost of stocking have a harder time trying to coincide purchases and sales. Their level of inventories is on average lower and they have to purchase more frequently, making it harder for them to wait for the next low price. Therefore, on average, the regular price of high stock cost households will be higher, meaning there is a correlation, by definition, of regular price and the household's costs of stockpiling. We believe this problem is not very important since empirical results remain basically the same when we use a regular price which is not household specific and thus free of the correlation with household costs.

We consider pack size in the definition of regular price because we want to differentiate situations when the household purchased at a lower price because the product was really on a sale, from situations when households purchased at a lower price only because they bought larger packs. Since quantity discounts are extremely frequent, the price per quantity paid when a larger pack is purchased is lower than the price per quantity paid when small packs are purchased, even if the product category was not on sale. The use of the regular price defined by household and size permits to separate sales from quantity discounts.

In the Appendix (Table 1.19), we show results on a regression of regular prices per household and pack size on household characteristics. The regular price decreases with family size, and increases with the age of the head of the household. Households where a woman is the shopper have higher regular prices than household where a man shops. Having a child of 6 or less years of age positively affects regular prices paid. Interestingly, regular prices paid for butter, yogurt, and pasta decrease with household income whereas regular prices paid for milk, coffee, and tuna increase with income. Finally, having a cellar and having a car, which can be considered as indication of low cost of stockpiling, have a positive effect on the regular price, which we interpret as evidence that the correlation between household costs and regular prices is not very important.

We define a price discount as a price 5% or more lower than the regular price. We could have defined discounted price simply as any price lower than regular. With the five percent margin, however, we try to avoid confounding regular price fluctuations to actual discounts. The 5% margin is of course quite arbitrary, but we have performed the tests with different discounted price definitions (lower than regular, at least 2% lower than regular, at least 10% lower than regular, and at least 15% lower than regular), and the results are qualitatively the same (coefficient signs do not change). Table 1.13 below shows the proportion of total quantity purchased during sales[6].

Table 1.13: Proportion of Quantity Purchased during Sales

Product	Proportion
Milk	0.332
Coffee	0.340
Tuna	0.470
Pasta	0.608
Butter	0.340
Yogurt	0.404

One main concern when studying the correlation between interpurchase duration and price is household heterogeneity with respect to inventory costs. If household heterogeneity is important, and uncrontrolled for, the estimated coefficient will be biased. To correct for this, we include household fixed effects, as well as household characteristics that are potentially correlated with inventory costs, such as if the household has a car or not, and variables that are proxies for stock space availability. In general, the bigger the home, the more space available for stocking inventories, and the lower the cost of inventories. We use two measures of space availability: if the household has an extra room for storage (a cellar), and if the household lives in a house or in an apartment.

Results for the estimation of the effect of discounted price on interpurchase duration are presented in Table 1.14. The first and third columns show results for simple OLS regressions where the dependent variable is, respectively, duration since last purchase and duration until next purchase. Coefficients displayed in the second and fourth columns are from regressions with fixed effect where the dependent variables are again duration from last purchase and until next purchase. The three last columns present, respectively, the number of households and the number of observation included in the regressions with and without household fixed effects. All regressions include controls for region of residence, family size, presence of a child (of 16 or less and 6 or less years of age), household income, car ownership, age and education of the head of the household, and gender of the person responsible for purchases. They also include space

[6] Proportion of total quantity purchased per product during whole time span which was made in periods when prices (per pack) were at least 5% lower than regular price.

availability controls, namely, whether the family lives in a house or an apartment and whether there is a special room for storage.

Results highly corroborate the inventory model. Almost every coefficient presents the correct sign. The only exceptions are pasta and yogurt for which there is no empirical evidence that duration from last purchase is negatively affected by lower prices today (although duration to next purchase is higher if the product is on sale today). Notice that, in general, evidence in favor of the model is stronger once household fixed effects are included (for instance, in the regressions of duration from last purchase for milk and butter, and duration until next purchase for tuna, the coefficient has the right sign and is significant only when we include fixed effects), which underline the importance of household unobservable heterogeneity.

Table 1.14: Effect of discounted price on duration from last purchase and duration until next purchase

Product	Coefficient Estimates: duration on discount prices						
	last(1)	last (2)	next (1)	next (2)	Nb hh	N (1)	N (2)
Milk	0.058	-0.254	0.688	0.359	6729	424674	425526
	(0.89)	(4.08)**	(10.60)**	(5.81)**			
Coffee	-0.098	-0.211	0.356	0.423	6478	224857	225061
	(0.59)	(1.55)	(2.14)*	(3.10)**			
Tuna	-1.062	0.686	-0.589	1.504	6474	118144	118261
	(2.74)**	(1.94)	(1.52)	(4.26)**			
Pasta	0.300	0.026	1.127	0.955	6819	323054	323419
	(2.73)**	(0.25)	(10.24)**	(9.17)**			
Butter	0.128	-0.422	1.412	0.672	6645	365822	366174
	(1.37)	(5.19)**	(15.14)**	(8.28)**			
Yogurt	0.032	0.232	0.197	0.372	6798	639128	639746
	(0.71)	(5.73)**	(4.40)**	(9.19)**			

Notes: (i) Absolute value of t statistics in parentheses, (ii) * significant at 5%, ** significant at 10%; (iii) (1) regression without fixed effects, and (2) with fixed effects; (iv) controls are: whether the household has a car, family size, region of residence, income, whether the person responsible for purchases is a male, presence of a child of 16 or less years of age, presence of a child of 6 or less years of age, and educational level of the head; (v) the column "Nb hh" (5th column) presents the number of households in the sample, and the two last columns, the number of observations in the regression without and with fixed effects, respectively.

To investigate the relationship between inventory costs, as proxied by space availability at home and frequency of purchases (*Implication 3*), we run regressions for each product subsample in which the dependent variable is the number of times the household purchased the product during the three year time span. In this exercise, the focus is on frequent consumers. Therefore, for each product category, we consider a subsample of households that purchased the product at least 12 times during the three years (approximately once every three months). The regressions include controls for income level, household size, age and education of the person of reference, gender of the person responsible for the purchases, presence of a child of 6 or less years of age, presence of a child of 16 or less years of age, as well as the region where the household lives. The variables that proxy for available space are the same as used above, i.e., dummy variables indicating if the household has a cellar, and if the home of the household is a house or an apartment[7]. We also include a dummy indicating whether the household owns a car. Although a car is not necessarily correlated with space availability, we conjecture that having a car decreases storage costs simply because it decreases the cost of bringing home large quantities of the product. Estimated coefficients are in Table 1.15.

Table 1.15: Effect of Space Availability on Frequency of Purchases

	Milk	Coffee	Tuna	Pasta	Butter	Yogurt
Cellar	-7.882	-2.189	-2.163	-5.294	0.916	-6.013
	(5.61)**	(1.95)	(2.43)*	(4.32)**	(0.64)	(2.67)**
House	-4.455	0.734	-0.220	0.821	2.444	-2.234
	(3.34)**	(0.66)	(0.26)	(0.70)	(1.75)	(1.04)
Car	-9.511	-0.210	-1.256	-0.775	-0.582	-2.860
	(3.90)**	(0.10)	(0.75)	(0.34)	(0.22)	(0.73)
Other Controls				Yes		
Obs	6292	4920	3632	6019	5721	6414

Notes: (i) Dependent variable is frequency of purchase per household (ii) Absolute value of t statistics in parentheses, (iii) * significant at 5%, ** significant at 10%; (iv) Other controls are: income level, household size, region, age and education of person of reference, gender of person responsible for purchases, presence of child of 6 or less years of age, presence of child of 6 or less years of age.

[7] Ideally, we would include a measure of total space availability (i.e., the total size of the home) but the data do not include this information. Alternatively, we could use census data and compute square footage of living area by zip code, as in Bell and Hilber (2005). However, we believe that may be substantial heterogeneity in home sizes within a zip code in France and preferred to stick to the extra room and house or apartment measures. We could also have included other space availability measures, such as wether the home has a garden, a dog (as in Hendel and Nevo, 2003), a bathroom etc., but decided not not to in order to avoid multicollinearity.

Clearly, having a cellar negatively affects the frequency of purchases of all product categories, except for butter. Living in a house instead of an apartment and owning a car also seems to negatively affect the frequency of purchase, although the negative coefficient is significant only for milk. Interestingly, among the product categories considered, milk is the one that occupies the most space (i.e., one standard-size pack of milk is larger than one standard-size pack of coffee, or butter, or tuna) and that weighs the most (thus the importance of owning a car when deciding to purchase for storage).

1.5.2 Structural Analysis

Identification and Estimation

In this section, we structurally estimate demand for the case where price is higher than regular price (equation 1.12). Unfortunately, we cannot estimate the structural parameter ϕ_i because its identification relies on the estimation of equation (1.13), which includes an unknown and unobservable function, namely $E_t\Psi_{it+1}$.

Rewriting (1.12), we get the equation to be estimated:

$$E_t\left(Q_{it}^* \mid q_{it} > 0, p_t > p_r\right) = -\frac{1}{\rho}T_i^{t-1}\ln\left(\alpha_i\right) - \frac{1}{\rho}\left[\ln p_t + \sum_{n \in T_{i1}^t}\ln p_n + T_{i0}^{t-1}\ln p_r\right] \tag{1.18}$$

Identification of the parameters of the model is standard. Variation over time of prices and the periods with and without purchases ensure the semi-parametric identification of the model parameters. This means that we can identify and estimate α_i and ρ without specifying the distribution function of ε_{it}. We do so running an OLS.

To allow for consumer-specific marginal utility of income (α) we interact T_i^{t-1} in (1.18) with a set of dummy variables indicating the household. In this way, instead of getting one estimated α per product category, we get as many α estimates as there are households in each product category sample. Individual specific parameters are an important contribution of our work and can be obtained thanks to the large number of observations in the dataset.

Remember that we do not observe inventories. Instead of assuming an arbitrary level for the initial inventory level, we perform the estimation on a subsample which begins, for each consumer i at the first purchase occasion following a purchase at higher than regular prices. When prices are higher than regular and the consumer purchases a positive amount of the product, Proposition 1 says that the end of the period level of inventories is equal to zero. Therefore, eliminating the observations before the first purchase at high prices, we can comfortably assume that initial stocks are equal to zero, avoiding the initial condition problem.

Another problem we have to deal with in the estimation is the fact that we only observe prices paid by households. This means that when a household does not purchase, we do not have information on the price she would have paid. Then, in periods when the household does not purchase, we consider as the price she would have paid the mean price paid that week by households living in the same region.

Finally, to control for potential seasonal effects, we include dummies indicating each one of the four seasons of the year in all regressions.

Results

Estimated ρ are all positive as predicted and significant at 5%. We do not report the estimated values because the absolute value of ρ has no special meaning. In this context, it is not actually the risk aversion but the concavity of the utility function for a certain product, and it depends not only on the product category but also on the unit of measurement considered for q_{it}.

Table 1.16 brings the mean estimated α_i per percentile. The estimated coefficients show the signs predicted by the model. Furthermore, except for a few exceptions (for less than 10% of the households), they are significant at least at 5%.

Table 1.16: Percentiles of Estimated α_i

Product	Obs	Percent5	Percent25	Median	Percent75	Percent95
Milk	534172	2.66e-23	1.74e-4	2.20	110.4	717.2
Coffee	560578	3.27e-12	0.002	0.804	11.9	83.4
Tuna	417879	1.23e-06	0.084	3.00	23.0	101.1
Pasta	711960	3.08e-21	0.001	3.09	60.6	252.2
Butter	604873	2.25e-11e-21	0.005	2.14	30.4	120.1
Yogurt	834354	3.00e-14	4.39e-4	0.605	24.7	179.7

Below, estimated $\alpha_i s$ are used to simulate the long run price elasticities derived from the model and to study whether household observable characteristics can explain differences in estimated parameters.

Price Responsiveness in the Dynamic and Static Models

To obtain a measure of the long-run and short run price elasticities, we plug the estimated $\alpha_i s$ and $\beta_i s$ into (1.17) and (1.16). The $\beta_i s$ are estimated by OLS using (1.15).

Table 1.17 below shows elasticities for each product when considering estimated parameters significant at 5%. In both tables, the first column shows the average long run price elasticity, while the second column shows the short run price elasticity calculated using the estimated $\beta_i s$. The third column presents an alternative measure of the price-elasticities: the estimated coefficient of the regression of the log of the individual quantity purchased on the log of the price, denoted ϵ^R. We have decided to include this last column for illustrative purposes. However, we believe that to have an idea of the difference between the short and long run measures of price elasticities, it is best to compare the first two columns of Table 1.17. These two columns show price elasticities which yield from nested models (one is the static version of the other).

Table 1.17: Estimated Long and Short Run Price Elasticities per Product Category

Products	Average Price Elasticities		
	ϵ^{LR}	ϵ^{SR}	ϵ^{R}
Milk	-0.058	-0.122	-0.737
Coffee	-0.090	-0.190	-0.506
Tuna	-0.090	-0.181	-0.409
Pasta	-0.065	-0.148	-0.504
Butter	-0.084	-0.161	-0.413
Yogurt	-0.093	-0.188	-0.609

Note: Estimated parameter used to calculate elasticities yields from inventory model $\left(\epsilon^{LR}\right)$, static version of demand model with inventories $\left(\epsilon^{SR}\right)$, and the coefficient of the regression of log of quantity purchased on log of price $\left(\epsilon^{R}\right)$. Utility Specification: $u\left(c_{it}\right)=-\left(1/\rho\right)\exp\left[-\rho c_{it}\right].$

The short run measures are consistently higher than their long run counterparts. The upper bias varies from 80% up to more than 100% depending on the product. The difference in measures is a result of the short run elasticities capturing not only consumption responses to price variation, but also inventory responses.

Interestingly, the difference between the long and short run price elasticities is lowest in the case of yogurt and butter. Yogurt is less storable than other products (yogurt cannot be stored for a long time), and is expensive to store (needs a refrigerator). Butter needs to be frozen to be stored thus requiring the household to have a freezer and increasing storage costs. We expect therefore inventories to be less relevant in the case of those two product categories and this seems indeed to be the case.

Consumers Heterogeneity

We regress the estimated $\alpha_i s$ on a number of household characteristics in order to assess if observables explain at least partially individual differences in the theoretical model parameter. Notice that a higher α_i for a certain product indicates that the household is willing to spend a lower proportion of its income on that product. Therefore, differences in the coefficients across products shed light on relative preferences or taste over products.

The variables included in the regression are: dummy variables indicating whether the household has a cellar, lives in a house or an apartment, has a car, includes a child of 6 years of age or less, includes a child of 16 years of age or less. Furthermore, we control for region of residence, household size, age, gender and education of the person of reference, and household income. The tables with all the estimated coefficients is in the Appendix (Table 1.22 - 1.24). Very few coefficients are significant indicating that unobservables play an important role in explaining the marginal utility of income.

1.6 Robustness Check

To check the robustness of the model results, we consider an alternative procedure for estimating model parameters. Furthermore, we estimate the marginal utility of revenue and calculate price elasticities under different price expectation assumptions.

1.6.1 Alternative Estimation Procedure

Suppose a certain consumer purchases today at a price higher than regular. Proposition 1 says that her end of the period inventory level will be equal to zero and she will purchase again next period since her beginning of next period level of inventory is going to be zero. If next period, the price is still higher than the regular price, she will again purchase only enough to cover consumption, choosing to hold no stocks. This means that next period, her consumption is going to be equal to the quantity purchased, since she had nothing stored at home and she will not store anything either. More formally, let T_i' be the set of consecutive periods such that $p_{(t-1)'} > p_r$ and $p_{t'} > p_r$ and $q_{i(t-1)'} > 0$ and $q_{it'} > 0$. Then, for all $(t-1)'$ and t' belonging to T_i', $y_{i(t-1)'} = y_{it'} = 0$, and therefore, $q_{it'} = c_{it'}$. This observation suggests a rather simple alternative for estimating the parameters of the model which considers only the subset of observation in T'. The equation to be estimated is:

$$q_{it'}^{*} = h\left(\alpha_i p_{t'}\right) - v_{it'} \tag{1.19}$$

When the utility function is a CARA, as considered before, (1.19) becomes:

$$q_{it'}^{*} = -\frac{1}{\rho}\ln\alpha_i - \frac{1}{\rho}\ln p_{t'} - v_{it'} \tag{1.20}$$

We estimate α_i and ρ in (1.20) and calculate price elasticities as before. We then compare the long and short run measures of price elasticity, where the short run measure is the price elasticity of the static model. Results do not change much. However, they are less precise because standard errors are bigger since here we use less observations to estimate the model parameters.

1.6.2 Alternative Price Expectation Hypothesis

The assumption that consumers always expect prices to return to its regular level may be too strong. Here, we re-estimate the parameters of the model and calculate price-elasticities under an alternative price expectation hypothesis. We start by estimating two different Markov price processes. We then assume consumers have rational expectations and that they expect prices to follow the estimated price processes.

The first price process we consider is[8]:

$$\text{Process 1: } p_{it} = (a_h + b_h p_{it-1})h + (a_l + b_l p_{it-1})l + (a_r + b_r p_{it-1})r \tag{1.21}$$

where h, l, and r indicate that p_{t-1} is either higher, lower or equal to the regular price, respectively. The second process is a VAR(1):

$$\text{Process 2: } p_{it} = a + b p_{it-1} \tag{1.22}$$

The problem is we do not observe prices, only prices paid by the households. We could have defined p_t and p_{t-1} to be the average price paid at periods t and $t-1$. However, we believe this would fail to capture important regional and per brand price variations. We thus preferred to consider the prices paid by each household, thus the i index on the price variables in (1.21) and (1.22). For periods where household i did not purchase (hence we do not observe the price paid by i), we consider the price paid by i to be equal to the average price paid at the same period by households that live in the same region as i.

Table 1.18 brings the long run price elasticities implied by the inventory model when we consider Process 1 and Process 2 (ϵ^{LR1} and ϵ^{LR2}). Only estimated parameters which are significant at 5% are used (for all product categories, this respresents more than 90% of the estimated parameters). The third column brings the price elasticities yielded by the static model of demand. Those are the same measures seen in the second column of Table 1.17 since the static estimates are not affected by price expectations. We re-include them here to facilitate comparison with the long run measures.

Note: Estimated parameters used to calculate price elasticities yielded from inventory model under price process 1 $\left(\epsilon^{LR1}\right)$ and price process 2 $\left(\epsilon^{LR2}\right)$, and the static version of the model $\left(\epsilon^{SR}\right)$.

Under both price expectation hypothesis, the results of the model are maintained. Even

[8] Higher order processes do not alter results in a relevant manner.

Table 1.18: Estimated Long and Short Run Price Elasticities under Alternative Price Expectation Assumptions

Products	Average Price Elasticities		
	ϵ^{LR1}	ϵ^{LR2}	ϵ^{S}
Milk	-0.073	-0.089	-0.122
Coffee	-0.159	-0.142	-0.190
Tuna	-0.127	-0.148	-0.188
Pasta	-0.126	-0.109	-0.148
Butter	-0.148	-0.129	-0.161
Yogurt	-0.116	-0.114	-0.142

though the upperbias of the static measures is smaller than before, it is still very significative, with the difference between the short and long run price elasticities varying from 9% to 42% under the first price process, and from 16% to 30% under the second price process.

1.7 Conclusion

Ignoring dynamics in the demand behavior of consumers may lead to biased estimates of the long run demand price elasticities. We propose a model of demand where consumers stockpile and prices are random. An assumption on the level of consumption at periods without purchases enables identification of the long run price elasticity without having to solve the dynamic program We also derive and test implications of the model.

The empirical analysis is performed using a comprehensive dataset on household food products purchases. We estimate individual specific marginal utilities of income from the purchase probability equations yielded by the model. The estimates are then used to simulate the long run demand price elasticities. We find that price elasticities resulting from a static demand model significantly overestimate the long run price elasticities. Finally, we show that results are robust to different price expectation assumptions.

1.8 Appendix

Short Run Price Elasticities in the Inventory Model

Individual demand at period t (the short run demand) is equal to:

$$\begin{aligned}
D_{it}^{ST}(p_t) &= q_{it}\Pr(q_{it}>0 \mid p_t) \\
&= q_{it}\Pr(q_{it}>0 \mid p_t>p_r)\Pr(p_t>p_r)+q_{it}\Pr(q_{it}>0 \mid p_t\leq p_r)\Pr(p_t\leq p_r) \\
&= (c_{it}-x_{it})\Pr(q_{it}>0 \mid p_t>p_r)\Pr(p_t>p_r)+ \\
&\quad (y_{it}+c_{it}-x_{it})\Pr(q_{it}>0 \mid p_t\leq p_r)\Pr(p_t\leq p_r)
\end{aligned}$$

where we use the fact that $p_t > p_r$ and $p_t \leq p_r$ are complementary events and apply Bayes' Theorem. We also use the law of motion of inventories (equation 1.1) to write $q_{it} = y_{it}+c_{it}-x_{it}$, as well as Proposition 1 which says that $y_{it} = 0$ when prices are higher than regular prices.

We assume v_{it} is normally distributed. Moreover, the first order conditions for the dynamic model imply that in periods with purchases, $c_{it} = h(\alpha_i p_t) - v_{it}$ and, $y_{it} = \frac{\alpha_i(p_r-p_t)}{2\phi_i} - \frac{E_t\Psi_{it+1}}{2\phi_i}$. Therefore, we can write:

$$\begin{aligned}
\Pr(q_{it}>0|p_t>p_r) &= \Pr(c_{it}-x_{it}>0 \mid p_t>p_r) \\
&= \Pr(h(\alpha_i p_t)-v_{it}-x_{it}>0 \mid p_t>p_r) \\
&= \Pr(v_{it}<h(\alpha_i p_t)-x_{it} \mid p_t>p_r) \\
&= \Phi(V_{it}|p_t>p_r)
\end{aligned}$$

and

$$\begin{aligned}
\Pr(q_{it}>0|p\leq p_r) &= \Pr(y_{it}+c_{it}-x_{it}>0 \mid p\leq p_r) \\
&= \Pr\left(\frac{\alpha_i(p_r-p_t)}{2\phi_i}-\frac{E_t\Psi_{it+1}}{2\phi_i}+h(\alpha_i p_t)-v_{it}-x_{it}>0 \mid p\leq p_r\right) \\
&= \Pr\left(v_{it}<\frac{\alpha_i(p_r-p_t)}{2\phi_i}-\frac{E_t\Psi_{it+1}}{2\phi_i}+h(\alpha_i p_t)-x_{it}>0 \mid p\leq p_r\right) \\
&= \Phi(U_{it} \mid p_t\leq p_r)
\end{aligned}$$

where Φ is the Normal cumulative distribution function, and V_{it} and U_{it} are, respectively:

$$V_{it} = h\left(\alpha_i p_t\right) - x_{it}$$

and

$$U_{it} = \frac{\alpha_i\left(p_r - p_t\right)}{2\phi_i} - \frac{E_t \Psi_{it+1}}{2\phi_i} + h\left(\alpha_i p_t\right) - x_{it}$$

Substituting V_{it}, U_{it}, $\Phi\left(V_{it} \mid p_t > p_r\right)$ and $\Phi\left(U_{it} \mid p_t \leq p_r\right)$ in (**??**) we get:

$$D_{it}^{ST}\left(p_t\right) = \left(V_{it} - v_{it}\right)\Phi\left(V_{it} \mid p_t > p_r\right)\Pr\left(p_t > p_r\right) + \left(U_{it} - v_{it}\right)\Phi\left(U_{it} \mid p_t \leq p_r\right)\Pr\left(p_t \leq p_r\right)$$

Hence, the short run elasticity of demand (or the price elasticity of purchases) in the inventory model is equal to:

$$\begin{aligned}
\epsilon_{SR} &= \left[\begin{array}{c} \frac{dV_{it}}{dp_t}\Phi\left(V_{it}|p_t > p_r\right)\Pr\left(p_t > p_r\right) + \left(V_{it} - v_{it}\right)\Phi'\left(V_{it}|p_t > p_r\right)\frac{dV_{it}}{dp_t}\Pr\left(p_t > p_r\right) \\ + \left(V_{it} - v_{it}\right)\Phi\left(V_{it}|p_t > p_r\right)\frac{d\Pr\left(p_t > p_r\right)}{dp_t} \end{array}\right]\frac{p_t}{V_{it} - v_{it}} \\
&+ \left[\begin{array}{c} \frac{dU_{it}}{dp_t}\Phi\left(U_{it}|p_t \leq p_r\right)\Pr\left(p_t \leq p_r\right) + \left(U_{it} - v_{it}\right)\Phi'\left(U_{it}|p_t \leq p_r\right)\frac{dU_{it}}{dp_t}\Pr\left(p_t \leq p_r\right) \\ + \left(U_{it} - v_{it}\right)\Phi\left(U_{it}|p_t \leq p_r\right)\frac{d\Pr\left(p_t \leq p_r\right)}{dp_t} \end{array}\right]\frac{p_t}{V_{it} - v_{it}} \\
&= \frac{p_t}{V_{it} - v_{it}}\frac{dV_{it}}{dp_t}\Pr\left(p_t > p_r\right)\left[\Phi\left(V_{it}|p_t > p_r\right) + \left(V_{it} - v_{it}\right)\Phi'\left(V_{it}|p_t > p_r\right)\right] \\
&\quad + p_t\Phi\left(V_{it}|p_t > p_r\right)\frac{d\Pr\left(p_t > p_r\right)}{dp_t} \\
&\quad + \frac{p_t}{V_{it} - v_{it}}\frac{dU_{it}}{dp_t}\Pr\left(p_t \leq p_r\right)\left[\Phi\left(U_{it}|p_t \leq p_r\right) + \left(U_{it} - v_{it}\right)\Phi'\left(U_{it}|p_t \leq p_r\right)\right] \\
&\quad + p_t\Phi\left(U_{it}|p_t \leq p_r\right)\frac{d\Pr\left(p_t \leq p_r\right)}{dp_t} \\
&= \frac{p_t}{V_{it} - v_{it}}\frac{dV_{it}}{dp_t}\Pr\left(p_t > p_r\right)\left[\Phi\left(V_{it}|p_t > p_r\right) + \left(V_{it} - v_{it}\right)\Phi'\left(V_{it}|p_t > p_r\right)\right] \\
&\quad + \frac{p_t}{V_{it} - v_{it}}\frac{dU_{it}}{dp_t}\Pr\left(p_t \leq p_r\right)\left[\Phi\left(U_{it}|p_t \leq p_r\right) + \left(U_{it} - v_{it}\right)\Phi'\left(U_{it}|p_t \leq p_r\right)\right] \\
&\quad + p_t\left[\Phi\left(V_{it}|p_t > p_r\right) - \Phi\left(U_{it}|p_t \leq p_r\right)\right]\frac{d\Pr\left(p_t > p_r\right)}{dp_t} \\
&= -\frac{1}{\alpha_i p_t\left(V_{it} - v_{it}\right)}\Pr\left(p_t > p_r\right)\left[\Phi\left(V_{it}|p_t > p_r\right) + \left(V_{it} - v_{it}\right)\Phi'\left(V_{it}|p_t > p_r\right)\right] \\
&\quad - \frac{p_t}{V_{it} - v_{it}}\left(\frac{\alpha_i}{\theta_i} + \frac{1}{\alpha_i p_t^2}\right)\Pr\left(p_t \leq p_r\right)\left[\Phi\left(U_{it}|p_t \leq p_r\right) + \left(U_{it} - v_{it}\right)\Phi'\left(U_{it}|p_t \leq p_r\right)\right] \\
&\quad + p_t\left[\Phi\left(V_{it}|p_t > p_r\right) - \Phi\left(U_{it}|p_t \leq p_r\right)\right]\frac{d\Pr\left(p_t > p_r\right)}{dp_t}
\end{aligned}$$

Table 1.19: Regular Price on Household Characteristics (I)

Charac	Products					
	Milk	Coffee	Butter	Yogurt	Pasta	Tuna
Extra Room	0.04	0.103	0.035	0.037	0.037	0.048
	(7.66)**	(8.28)**	(17.62)**	(21.89)**	(21.89)**	(3.94)**
House	-0.011	0.012	0.001	-0.016	-0.016	0.012
	(20.95)**	(0.93)	(0.52)	(10.50)**	(10.50)**	(1.08)
Car	0.002	0.137	0.008	0.031	0.031	0.233
	(1.73)	(5.91)**	(2.28)*	(9.75)**	(9.75)**	(11.06)**
Fam Size=2	-0.036	-0.582	-0.137	-0.100	-0.100	-0.521
	(39.00)**	(26.88)**	(39.07)**	(30.26)**	(30.26)**	(25.38)**
Fam Size=3	-0.064	-0.696	-0.253	-0.152	-0.152	-0.647
	(60.05)**	(28.27)**	(62.87)**	(42.06)**	(42.06)**	(28.27)**
Fam Size=4	-0.094	-0.840	-0.327	-0.232	-0.232	-0.822
	(83.13)**	(32.30)**	(78.18)**	(62.35)**	(62.35)**	(34.41)**
Fam Size=5	-0.109	-1.063	-0.435	-0.317	-0.317	-1.125
	(89.40)**	(36.96)**	(94.91)**	(80.19)**	(80.19)**	(43.39)**
Fam Size=6	-0.127	-1.453	-0.491	-0.380	-0.380	-1.332
	(79.25)**	(37.10)**	(80.44)**	(78.35)**	(78.35)**	(39.10)**
Fam Size=7	-0.126	-1.933	-0.550	-0.507	-0.507	-1.660
	(49.85)**	(28.55)**	(49.08)**	(68.84)**	(68.84)**	(30.46)**
Fam Size=8	-0.146	-1.108	-0.666	-0.390	-0.390	-1.590
	(31.45)**	(9.68)**	(33.08)**	(33.42)**	(33.42)**	(13.02)**
Fam Size=9	-0.191	-2.878	-0.742	-0.758	-0.758	-2.308
	(29.75)**	(16.95)**	(37.26)**	(26.11)**	(26.11)**	(13.40)**
Inc 2	-0.048	-0.118	-0.285	-0.531	-0.531	-0.137
	(3.31)**	(0.42)	(4.01)**	(11.51)**	(11.51)**	(0.43)
Inc 3	0.007	0.530	-0.236	-0.314	-0.314	0.280
	(0.52)	(1.94)	(3.38)**	(7.02)**	(7.02)**	(0.89)
Inc 4	-0.000	0.542	-0.298	-0.400	-0.400	0.641
	(0.01)	((2.00)**	(4.27)**	(9.00)**	(9.00)**	(2.07)*
Inc 5	0.011	0.448	-0.268	-0.393	-0.393	0.502
	(0.76)	(1.66)	(3.86)**	(8.88)**	(8.88)**	(1.62)

Note: Absolute values of t statistics in parentheses, (i) * significant at 5%, ** significant at 10%.

Note: Absolute values of t statistics in parentheses, (i) * significant at 5%, ** significant at 10%.

Table 1.20: (Cont) Regular Price on Household Characteristics (II)

Charac	Products					
	Milk	Coffee	Butter	Yogurt	Pasta	Tuna
Inc 6	0.013	0.640	-0.187	-0.306	-0.306	0.655
	(0.89)	(2.37)*	(2.68)**	(6.91)**	(6.91)**	(2.12)*
Inc 7	0.031	0.952	-0.259	-0.324	-0.324	0.655
	(2.21)*	(3.53)**	(3.73)**	(7.34)**	(7.34)**	(2.12)*
Inc 8	0.032	0.909	-0.189	-0.325	-0.325	0.846
	(2.25)**	(3.37)**	(2.71)**	(7.36)**	(7.36)**	(2.73)**
Inc 9	0.030	0.951	-0.194	-0.297	-0.297	0.657
	(2.12)*	(3.52)**	(2.80)**	(6.73)**	(6.73)**	(2.12)*
Inc 10	0.037	1.024	-0.146	-0.258	-0.258	0.904
	(2.59)**	(3.80)**	(2.10)*	(5.85)**	(5.85)**	(2.92)**
Inc 11	0.052	1.264	-0.105	-0.227	-0.227	0.991
	(3.66)**	(4.69)**	(1.51)	(5.13)*	(5.13)**	(3.21)**
Inc 12	0.052	1.321	-0.087	-0.227	-0.277	1.012
	(3.70)**	(4.89)**	(1.25)	(5.13)**	(5.13)**	(3.27)**
Inc 13	0.059	1.471	0.000	-0.177	-0.177	1.372
	(4.19)**	(5.45)**	(0.00)	(4.00)**	(4.00)**	(4.44)**
Inc 14	0.066	1.587	-0.031	-0.152	-0.152	1.494
	(4.70)**	(5.87)**	(0.44)	(3.43)**	(3.43)**	(4.83)**
Inc 15	0.078	1.845	0.021	-0.085	-0.085	1.579
	(5.52)**	(6.81)**	(0.31)	(1.91)	(1.91)	(5.10)**
Inc 16	0.087	1.392	0.085	-0.130	-0.130	1.491
	(6.09)**	(5.09)**	(1.21)	(2.92)**	(2.92)**	(4.78)**
Inc 17	0.101	1.837	0.036	-0.049	-0.049	2.024
	(7.06)**	(6.70)**	(0.51)	(1.10)	(1.10)	(6.47)**
Inc 18	0.327	2.316	0.349	0.173	0.173	2.767
	(21.90)**	(8.01)**	(4.92)**	(6.25)**	(3.73)**	(8.54)**
Man	-0.002	-0.374	-0.025	-0.068	-0.086	-0.270
	(1.30)	(11.16)**	(4.63)**	(16.56)**	(17.85)**	(9.16)**
Child<16	-0.010	-0.064	-0.041	0.001	-0.025	-0.063
	(13.56)**	(3.84)**	(15.26)**	(0.52)	(12.33)**	(4.33)**
Child<6	0.040	0.066	0.018	0.000	0.014	0.009
	(53.93)**	(3.42)**	(6.10)**	(0.04)	(6.46)**	(0.55)

Table 1.21: (Cont) Regular Price on Household Characteristics (III)

Charac	Products					
	Milk	Coffee	Butter	Yogurt	Pasta	Tuna
Age 30	0.007	0.120	0.034	0.030	0.063	0.108
	(6.94)**	(3.97)**	(7.28)**	(9.56)**	(19.83)**	(4.77)**
Age 40	0.024	0.313	0.076	0.063	0.105	0.374
	(21.00)**	(9.89)**	(15.53)**	(18.83)**	(30.78)**	(15.54)**
Age 60	0.045	0.727	0.147	0.047	0.159	0.996
	(35.66)**	(21.98)**	(28.69)**	(13.15)**	(42.32)**	(38.16)**
Educ 1	0.029	0.444	0.095	0.049	0.137	0.307
	(31.12)**	(20.69)**	(27.43)**	(20.07)**	(49.87)**	(15.66)**
Educ 2	0.014	0.474	0.076	0.039	0.053	0.173
	(19.58)**	(29.78)**	(29.23)**	(19.84)**	(25.48)**	(11.80)**
Educ 3	0.010	0.125	0.046	0.023	0.022	0.009
	(17.19)**	(9.52)**	(21.33)**	(13.69)**	(12.73)**	(0.76)
Region Dummies	Yes					
Observations	431405	231383	327507	645906	329878	124732
R-squared	0.11	0.05	0.13	0.07	0.18	0.13

Table 1.22: Alpha on Household Characteristics I

Charac	Products					
	Milk	Coffee	Butter	Yogurt	Pasta	Tuna
Extra	5.04e+3	7.52e+16	-2.87e+17	-3.43e+21	-6.04e+21	-2.69e+14
	(1.91)	(1.50)	(0.31)	(0.03)	(0.84)	(0.66)
House	1.29e+3	-1.28e+17	-1.52e+18	-1.48e+23	4.85e+21	-5.07e+14
	(0.51)	(2.62)**	(1.58)	(1.46)	(0.70)	(1.27)
Car	-1.36e+3	3.08e+16	4.76e+17	1.96e+23	-1.87e+20	5.29e+14
	(0.27)	(0.32)	(0.25)	(1.05)	(0.01)	(0.69)
Fam Size=2	-8.27e+2	7.34e+16	4.69e+17	-3.99e+23	8.25e+21	-6.65e+14
	(0.17)	(0.76)	(0.28)	(2.46)*	(0.73)	(0.99)
Fam Size=3	-8.18E+3	3.16e+16	2.78e+18	-4.91e+23	-8.38e+21	-9.04e+14
	(0.15)	(0.29)	(1.47)	(2.56)*	(0.63)	(1.15)
Fam Size=4	1.86e+3	3.44e+16	1.09e+18	-5.04e+23	-1.01e+22	-8.98e+14
	(0.33)	(0.31)	(0.54)	(2.47)*	(0.71)	(1.08)
Fam Size=5	-1.81e+3	4.48e+16	1.32e+18	-5.12e+23	-1.14e+22	-8.62e+14
	(0.29)	(0.37)	(0.60)	(2.24)*	(0.72)	(0.94)
Fam Size=6	-332	5.13e+16	1.27e+18	-5.05e+23	-1.2e+22	-9.20e+14
	(0.04)	(0.34)	(0.43)	(1.62)	(0.55)	(0.75)
Fam Size=7	-871	6.83e+16	1.726e+18	-5.14e+23	-1.11e+22	-1.03e+15
	(0.07)	(0.28)	(0.33)	(0.98)	(0.32)	(0.48)
Fam Size=8	4.5e+3	2.62e+16	8.17e+17	-6.40e+23	-1.236e+22	-1.27e+15
	(0.23)	(0.06)	(0.11)	(0.69)	(0.21)	(0.34)
Fam Size=9	1.9e+3	7.33e+15	1.56e+18	-6.92e+23	-1.43e+22	-5.21e+14
	(0.06)	(0.01)	(0.11)	(0.43)	(0.12)	(0.08)
Region=2	329	5.06e+16	4.27e+18	7.06e+22	-1.75e+21	6.20e+14
	(0.07)	(0.60)	(2.56)*	(0.40)	(0.14)	(0.88)
Region=3	26.6	7.81e+16	7.63e+17	5.163+23	-1.38e+21	6.88e+14
	(0.01)	(0.88)	(0.44)	(2.80)**	(0.11)	(0.94)
Region=4	232	1.62e+17	6.12e+17	9.48e+22	1.13e+22	7.08e+14
	(0.06)	(2.13)*	(0.43)	(0.62)	(1.08)	(1.18)
Region=5	109	5.89e+16	5.50e+17	7.50e+22	-1.47e+21	2.39e+15
	(0.02)	(0.64)	(0.31)	(0.40)	(0.12)	(3.27)**
Region=6	6.6e+3	3.86e+16	3.32e+17	3.52e+22	-1.43e+21	4.20e+14
	(1.68)	(0.49)	(0.22)	(0.23)	(0.13)	(0.68)
Region=7	992	2.97e+16	2.80e+17	1.83e+22	-1.88e+21	3.58e+14
	(0.23)	(0.35)	(0.17)	(0.11)	(0.16)	(0.53)
Region=8	669	5.74e+16	3.45e+17	5.60e+22	-2.93e+21	6.00e+14
	(0.14)	(0.64)	(0.20)	(0.31)	(0.23)	(0.80)

Table 1.23: (cont) Alpha on Household Characteristics II

Charac	Products					
	Milk	Coffee	Butter	Yogurt	Pasta	Tuna
Man	-77.3	-1.12e+16	-3.06e+17	-3.83e+23	-6.69e+21	4.09e+15
	(0.01)	(0.08)	(0.13)	(1.68)	(0.40)	(4.13)**
Child⩽16	2.6e+3	3.77e+12	-1.45e+18	-7.47e+21	-1.76e+21	4.33e+13
	(0.81)	(0.00)	(1.13)	(0.05)	(0.18)	(0.08)
Child⩽6	-7.1e+33	-4.78e+15	9.65e+16	8.55e+22	4.24e+21	1.60e+14
	(2.01)*	(0.07)	(0.07)	(0.59)	(0.42)	(0.27)
age30	1.2e+3	2.12e+16	9.38e+17	8.45e+22	2.99e+21	5.75e+13
	(0.23)	(0.20)	(0.39)	(0.42)	(0.20)	(0.06)
age40	-5.0e+3	3.30e+16	1.27e+18	2.34e+23	1.00e+22	5.51e+14
	(0.88)	(0.29)	(0.51)	(1.11)	(0.64)	(0.58)
age60	-4.5e+3	1.20e+17	8.39e+17	5.58e+22	-5.02e+21	1.21e+14
	(0.73)	(0.99)	(0.32)	(0.25)	(0.30)	(0.12)
educ1	-2.7e+3	3.03e+16	8.61e+17	1.49e+22	4.22e+21	8.28e+14
	(0.62)	(0.35)	(0.53)	(0.09)	(0.35)	(1.20)
educ2	1.4e+3	1.23e+17	6.78e+17	2.06e+23	3.00e+21	-2.97e+14
	(0.44)	(1.90)	(0.54)	(1.52)	(0.32)	(0.56)
educ3	-1.5e+3	2.86e+16	1.16e+18	3.27e+22	8.65e+21	-1.56e+14
	(0.54)	(0.53)	(1.12)	(0.28)	(1.10)	(0.35)
Inc2	-3.4e+3	7.95e+16	-7.70e+17	4.65e+22	-4.13e+21	-9.18e+14
	(0.06)	(0.06)	(0.03)	(0.02)	(0.03)	(0.08)
Inc3	-1.8e+3	7.36e+16	-1.44e+18	6.40e+22	-3.63e+21	-7.05e+14
	(0.03)	(0.05)	(0.06)	(0.03)	(0.03)	(0.06)
Inc4	-2.0e+3	7.74e+16	-1.50e+18	1.53e+23	-4.10e+21	-4.81e+14
	(0.04)	(0.06)	(0.06)	(0.08)	(0.04)	(0.04)
Inc5	-338	6.48e+16	-1.57e+18	1.25e+23	-5.70e+21	-2.47e+14
	(0.01)	(0.05)	(0.07)	(0.07)	(0.05)	(0.02)
Inc6	-1,5e+3	5.93e+16	-1.65e+18	1.10e+23	-3.86e+21	-1.28e+14
	(0.03)	(0.04)	(0.07)	(0.06)	(0.03)	(0.01)
Inc7	-1.4e+3	6.6e+16	-1.85e+18	1.63e+23	-4.47e+21	-3.79e+14
	(0.03)	(0.05)	(0.08)	(0.09)	(0.04)	(0.03)

Table 1.24: (cont) Alpha on Household Characteristics III

Charac	Products					
	Milk	Coffee	Butter	Yogurt	Pasta	Tuna
Inc8	-1.8e+3	7.99e+16	-1.92e+18	9.07e+23	-4.65e+21	-7.23e+13
	(0.03)	(0.06)	(0.08)	(0.48)	(0.04)	(0.01)
Inc9	-2.0e+3	7.62e+16	-1.94e+18	1.70e+23	-4.17e+21	-2.94e+12
	(0.04)	(0.06)	(0.08)	(0.09)	(0.04)	(0.00)
Inc10	-2.52e+3	6.93e+16	-1.8e+15	2.06e+23	1.12e+22	1.10e+14
	(0.04)	(0.05)	(0.00)	(0.11)	(0.10)	(0.01)
Inc11	-2.8e+3	6.97e+16	-2.18e+18	2.07e+23	-4.37e+21	1.67e+14
	(0.05)	(0.05)	(0.09)	(0.11)	(0.04)	(0.02)
Inc12	-2.78e+3	7.69e+16	-2.17e+18	2.16e+23	-4.17e+21	1.92e+14
	(0.05)	(0.06)	(0.09)	(0.12)	(0.04)	(0.02)
Inc13	-2.54e+3	2.98e+17	-2.09e+18	2.13e+23	-4.21e+21	2.42e+14
	(0.04)	(0.22)	(0.09)	(0.11)	(0.04)	(0.02)
Inc14	9.38e+3	5.73e+16	-2.07e+18	2.00e+23	-2.62e+21	2.14e+15
	(0.17)	(0.04)	(0.09)	(0.11)	(0.02)	(0.20)
Inc15	-3.09e+3	7.28e+16	-2.15e+18	2.32e+23	-4.22e+21	2.91e+14
	(0.05)	(0.05)	(0.09)	(0.12)	(0.04)	(0.03)
Inc16	-3.13e+3	5.37e+16	-2.15e+18	2.17e+23	-3.53e+21	1.04e+14
	(0.05)	(0.04)	(0.09)	(0.12)	(0.03)	(0.01)
Inc17	-2.80e+3	6.89e+16	-2.64e+18	2.11e+23	-1.76e+21	1.52e+14
	(0.05)	(0.05)	(0.11)	(0.11)	(0.01)	(0.01)
Inc18	-4.55e+3	9.97e+16	-1.95e+18	2.34e+23	-7.14e+20	3.05e+14
	(0.07)	(0.07)	(0.08)	(0.12)	(0.01)	(0.03)
Const	4.46e+3	-2.32e+17	-4.10e+17	-1.70e+23	-3.48e+21	-5.28e+14
	(0.08)	(0.17)	(0.02)	(0.09)	(0.03)	(0.05)
Obs	3083	4569	3716	6110	4396	3086
R2	0.01	0.01	0.01	0.01	0.01	0.00

Chapter 2

Consumers' Quality Choices during Demand Peaks

2.1 Introduction

Common sense preaches that when demand increases, prices should increase as well. Lately however, a number of studies brings evidence that for certain categories of products, demand peaks are actually periods of lower, not higher, prices. The natural question is of course why. What drives such a counter intuitive result?

The literature presents at least three different explanations for the phenomena of decreasing prices during a demand peak. First, Warner and Barsky (1995) argue that periods of high aggregate demand are also periods where consumers are willing to invest more on information and transportation to find the lowest price. In other words, since consumers will usually have a longer list of items to purchase during Christmas, for instance, they will be willing to pay the extra search cost for the best price. They are then more vigilant and better informed and retailers perceive their demand to be more elastic. The optimal markups are thus lower, driving prices down. Notice that Warner and Barsky's argument works only for periods of high aggregate demand, not product category specific peaks.

Second, Chevalier, Kashyap and Rossi (2003) find evidence in favour of the loss-leader model of advertising as exemplified by Lal and Matutes (1994). The idea is that high demand items are advertised at sales prices to attract customers to the store (where all unadvertised products are sold at reservation prices). Hence, a product-specific shift in demand could trigger advertising, and thus price reductions for that product.

Finally, an alternative explanation, proposed by Nevo and Hatzitaskos (2006), is based on a change in brand-level demand. The idea is that price sensitivity may be higher during demand peaks due to a change in the brand-level preferences of consumers.

In this paper, we contribute to the emerging literature on the effect of demand peaks on prices by formalizing Nevo and Hatzitaskos (2006)'s argument of brand-level shift on demand. We derive testable implications of the model, which are tested using micro data on ice-cream purchases. Furthermore, we structurally estimate demand and calculate cross-price elasticities, showing they are higher during demand peaks, which is a direct evidence against Chevalier et al.'s loss-leader model.

The demand for ice-cream increases significantly during the summer, while the average price paid goes down. To study the quality choice shift, we consider quality as the only product differentiation dimension and we restrict the number of quality levels to two. The quality indicator considered is whether the brand of the product is a national (high quality) or a store (low quality) brand. Although we do not want to start a debate on the actual taste differentials, we claim that in France there is widespread belief that store brand products are of lower quality than national brand products. The perceived quality difference can also be inferred from prices: national brands are on average a lot more expensive than store brands. To check whether results are dependent on the restrictive product definition, tests are also performed considering other dimensions of differentiation (store, brand and product characteristics).

Tests are performed using home-scan data. Household level information on store visits, purchases, and prices paid were collected during three years (1999, 2000, 2001) from a nationally representative survey. Data on household characteristics, including characteristics of the home and of the individuals composing the household, and product and store characteristics were also collected.

We find evidence that during periods of exogenous increase in demand, consumers are more price elastic. As a results, they tend to shift product choice towards cheaper products. The higher elasticity and the shift in demand drive prices down. The decrease in prices is however less important than the decrease measured by average aggregate prices paid. Since demand migrates towards lower cheaper products, ignoring product differentiation and looking only at average category prices leads to an overstatement of the importance of the price decrease. Results point out the importance of considering product differentiation.

The paper is organized as follows. In the next section, we briefly review the literature on price responses to demand peaks. The theoretical model and its implications are presented in Section 3, where we also discuss identification issues and tests of alternative explanations to the decrease in prices during demand peaks. Section 4 describes the data and presents some descriptive statistics of interest. The empirical analysis is in Section 5: we test implications of the model, estimate demand and price-elasticities and also perform a robustness check of the results by considering a less restrictive definition of product. The last section concludes the study.

2.2 Literature Review

Warner and Barsky (1995) use data on daily observed prices for consumer good items in a total of seventeen outlets from November 1987 to February 1988. They report that for the eight goods considered, prices fall as the weekend approaches, reaching its weekly minimum on Fridays and Saturdays and rising to their peak on Mondays. Prices fall also in December, although they reach their lowest in January. The explanation given by the authors for this pricing pattern is based on economies of scale in search costs. They argue that on periods characterized by exogenously high intensity of shopping activity, such as weekends and Christmas, the search for the best price takes place more efficiently. With better informed consumers, the demand perceived by retailers is more elastic, driving price cost margins and retail prices down.

Chevalier, Kashyap and Rossi (2003), henceforth CKR, test between three theories of imperfect competition which generate countercyclical prices using weekly data from a supermarket chain in Chicago. The data set includes not only information on prices and quantities, but also a measure of wholesale prices.

The first of the theories tested is that of economies of scale in search costs, such as Warner and Barsky (1995)'s. The second one is based on Rotemberg and Saloner (1986) where countercyclical markups result from changes in the ability of firms to tacitly collude during periods of high demand, when the gains from charging a price below the collusion price are higher. The last class of models are the loss-leader models of advertising (Lal & Matutes, 1994). In this setting, retailers compete for customers via advertised prices. It is efficient for retailers to advertise and discount items in high relative demand and therefore attract customers to the store. Under relatively high search costs, customers will purchase the loss-leader product, as well as the rest of the good basket at the same store.

The two first models predict lower prices during periods of aggregate peak demand, but not during periods of product-specific high demand. The loss-leader model however predicts price decreases during product specific demand shifts as well. The main strategy used by CKR consists in separating periods of high overall demand from periods of idiosyncratic peak demand. The authors find evidence that prices decline even during product specific demand peaks.

Finally, CKR estimate category-level demand, finding no evidence of higher sensitivity during demand peaks, which is direct evidence against Warner and Barsky's argument.

All tests are performed either at category level or using an aggregate of highly price-correlated products. In all cases, variable-price indexes are constructed, where the price weights are the market shares of each product at each period. Since there is no product differentiation, brand or quality shifts are not identifiable.

Nevo and Hatzitaskos (2006)'s explanation for price falls during demand peaks is based on changes in brand-level demand, which could be of two types. First, price sensitivity may be higher during periods of peak demand leading to lower equilibrium prices (because the mix of consumers might be changing or a given consumer might be more price sensitive because the product is used differently during a period of high demand). Second, brand preferences within a product category might change (because during peak demand the product is used differently), leading consumers to migrate towards low-quality products .

The main difference between the analysis in Nevo and Hatzitaskos and in CKR is that the latter advocates the importance of incorporating retailer behavior, but does so at the cost of considering product as (essentially) homogeneous. Nevo and Hatzitaskos, on the other hand, place more emphasis on product differentiation. Their empirical study begins by re-examining CKR's data. They repeat CKR's analysis using a fixed-weights price-index, instead of their variable-weight price index. The variable price index used in CKR (prices of individual items are averaged proportional to quantity sold) can be potentially misleading if preferences for brands shift towards cheaper products. Actually, even with no change in prices, a variable price index might change due to composition effect. The fixed-weight index, on the other hand, will not be affected by a change in demand composition and will, therefore, remain unaltered if a change in composition is all that is happening. Indeed, for the products identified by CKR as loss-leaders, a fixed-weights price index displays much smaller price declines (in some cases none at all).

Next, they focus the analysis on tuna, which faces a demand peak during Lent. They find that much of the increase in the quantity sold is due to two relatively cheap products that nearly triple their market share. Interestingly, these two products do not reduce their prices. Results of estimation of item-level demand for tuna shows that consumers are more price-sensitive during Lent, when brand preferences seem to change. Furthermore, they find no evidence that advertising is more effective during Lent, which is a testable implication of the loss-leader theory.

This paper models the effect of an exogenous seasonal shock on individual choice of product quality. We show that for high positive realizations of the shock, consumers shift their quality choice towards less expensive products, which have lower perceived quality. We find empirical evidence that supports Nevo and Hatzitaskos's demand movements explanation against CKR's loss-leader theory.

2.3 Demand Peaks

In this section we develop a model of individual consumer demand for a certain good that comes in two different quality levels. An exogenous demand peak is modeled as a high realization of a seasonal shock to utility[1]. During demand peaks, the individual demand for each product, conditional on the choice of that product, increases. However, given prices, consumers tend to shift choice towards lower quality products, which are cheaper. Testable implications of the model are derived.

The description of the model of demand is followed by a discussion on identification of the model testable implications. Supply movements or change on the mix of consumers in the market could generate some of the same implications derived from the model of quality choice. It is essential therefore to make sure that what is driving the results is indeed a shift in the behavior of consumers. To do that, we derive implications of alternative explanations which will be later tested.

2.3.1 The Model

Given a certain product category, the problem of the consumer is:

$$\max_{q_{ijt}} u\left(q_{ijt} - v_t\right) + \beta_i Q_j - \alpha p_{jt} q_{ijt} \tag{2.1}$$

where $v_t > 0$ is an exogenous seasonal shock to utility, q_{ijt} is the quantity purchased of product j at period t by consumer i, Q_j is the quality of product j, and p_{jt} is the price of product j at period t. The parameter α represents the marginal utility of income, while β_i is a consumer-specific taste parameter. We can write β_i as $\beta_i = \bar{\beta} + \xi_i$, where $\bar{\beta}$ is the mean taste for quality across the population of consumers, and ξ_i is consumer i's taste variation around the mean, which has distribution F across the population of size N. The utility function u is increasing

[1] The model is similar to Gabszewicz and Thisse (1979, 1980) and Shaked and Sutton (1982, 1983), except that here demand is elastic and we include a seasonal shock to utility.

and concave. For simplicity, we assume[2] a functional form for u. Specifically, we assume that $u(q_{ijt} - v_t) = \ln(q_{ijt} - v_t)$. The first order condition then implies that q_{ijt} that maximizes utility is:

$$q_{ijt} = \frac{1}{\alpha p_{jt}} + v_t \tag{2.2}$$

Therefore, all else equal, higher realizations of the seasonal shock to utility, v_t, increase purchased quantity conditional on product choice

There are two products, H and L, that differ in quality. Product H is the high-quality product, and product L, the low quality product. The quality of product H is set equal to 1 ($Q_H = 1$), and the quality of product L, equal to zero ($Q_L = 0$). The two products are therefore vertically differentiated, i.e., at the same prices, all consumers prefer product H to product L, independently of the parameter β. However, in general, prices are not the same and consumers will compare indirect utilities in order to choose which product to purchase. The indirect utility of H and L are, respectively:

$$\begin{aligned} V_t^H &= \ln\left(\frac{1}{\alpha p_{Ht}}\right) - \alpha p_{Ht}\left(\frac{1}{\alpha p_{Ht}} + v_t\right) + \beta_i \\ &= \ln\left(\frac{1}{\alpha p_{Ht}}\right) - \alpha p_{Ht}\left(\frac{1}{\alpha p_{Ht}} + v_t\right) + \bar{\beta} + \xi_i \end{aligned} \tag{2.3}$$

and

$$V_t^L = \ln\left(\frac{1}{\alpha p_{Lt}}\right) - \alpha p_{Lt}\left(\frac{1}{\alpha p_{Lt}} + v_t\right) \tag{2.4}$$

Consumers will choose to purchase product L as long as their indirect utility of consuming L is greater than their indirect utility of consuming H ($V_t^L \geqslant V_t^H$). The probability that consumer i chooses product L at t is therefore :

$$P(y_{itL} = 1) = P\left(\xi_i \leq -\bar{\beta} + \ln\left(\frac{p_{Ht}}{p_{Lt}}\right) + \alpha v_t\left(p_{Ht} - p_{Lt}\right)\right) \tag{2.5}$$

[2] We also consider a family of CES utility functions with the relative risk aversion greater than 1, that is $u(q_{ijt} - v_t) = \frac{(q_{ijt} - v_t)^{1-\gamma}}{1-\gamma}$ with $\gamma > 1$. We estimate the structural demand parameters under alternative γ values. Results are robust to alternative utility specifications. Furthermore, we compare log-likelihoods under different values for γ. The log-likelihood is maximized for $\gamma = 1.25$ but it is very close to the second highest log likelihood wich is obtained for $\gamma \to 1$, in which case $u(q_{ijt} - v_t) = \ln(q_{ijt} - v_t)$. See Appendix for more details.

We assume that at each period, the total demand for both quality levels is positive[3], that is $0 < P(y_{itL} = 1) < 1$ and $p_{Ht} > p_{Lt}$. Furthermore, we do not model the decision to purchase, implying that the analysis of consumer behavior is made conditional on purchases, as it is usual in the literature. Instead of interpreting the model as describing product choice conditional on purchases, we can assume that conditional on having chosen product j, $q_{ijt} \in [\delta, +\infty]$ for every t and i, where δ is a positive number very close to zero. We have decided not to model participation into the market because we believe that it does not add significantly to the understanding of the phenomena under study (main results remain basically the same), while it does have the cost of adding complexity to the model.

Let $\tilde{\beta}_t \equiv -\bar{\beta} + \ln\left(\frac{p_{Ht}}{p_{Lt}}\right) + \alpha v_t\left(p_{Ht} - p_{Lt}\right)$ and $P_{it} = P(y_{itL} = 1)$. From (2.5), and assuming ξ_i has a logistic distribution, we can derive the cross-price elasticity ε_H^L of the probability of choosing the low quality product:

$$\begin{aligned}
\varepsilon_H^L &= \frac{dP_{it}}{dp_{Ht}}\frac{p_{Ht}}{P_{it}} \\
&= \frac{e^{\tilde{\beta}_t}}{\left(1+e^{\tilde{\beta}_t}\right)^2}\frac{d\tilde{\beta}_t}{dp_{Ht}}\frac{p_{Ht}}{P_{it}} \\
&= \frac{P_{it}}{1+e^{\tilde{\beta}_t}}\frac{d\tilde{\beta}_t}{dp_{Ht}}\frac{p_{Ht}}{P_{it}} \\
&= \frac{p_{Ht}}{1+e^{\tilde{\beta}_t}}\frac{d\tilde{\beta}_t}{dp_{Ht}} \\
&= \frac{(1+\alpha v_t p_{Ht})}{1+e^{\tilde{\beta}_t}} \qquad (2.6)
\end{aligned}$$

The cross-price elasticity derived above increases at high realizations of the utility shock v_t (see proof in the Appendix). That is, at periods of demand peak, consumers will pay relatively more attention to prices than to quality level and smaller variations of price will be sufficient to trigger a shift in quality choice.

Finally, what happens to market shares (in volume) at periods of high exogenous demand?

[3] This assumption describes well the market we consider in the empirical session. It can be easily verified in the data.

Aggregate demand for each quality level is given by:

$$D_{Lt} = P_{it} q_{itL}$$

and

$$D_{Ht} = (1 - P_{it})\, q_{itH}$$

The volume market share of the low quality product is then equal to:

$$S_{Lt} = \frac{D_{Lt}}{D_{Lt} + D_{Ht}} = \frac{P_{it} q_{itL}}{P_{it} q_{itL} + (1 - P_{it})\, q_{itH}} \tag{2.7}$$

In the Appendix we show that, given prices, $\frac{dS_{Bt}}{dv_t} \geqslant 0$, i.e., at periods of high demand, consumers will tend to shift their choice towards lower quality products.

In summary, the model implies that during periods of exogenous demand peaks:

Implication 1 Consumers are more cross-price elastic, caring relatively more about prices than quality;

Implication 2 Consumers shift their choice towards the less expensive product which has lower quality, increasing the market share of the low quality product (and decreasing the market share of the high quality product).

Implication 2 affects average category prices. To see that, consider two ways of measuring the category price level: a variable weight price index and a fixed weight price index, where the weight is the market share of each product. For the variable weight index, current prices are weighted by the current market share, whereas for the fixed weight index, current prices are weighted by the average market share during the whole data time span. Then, the variable weight index would show a decrease in prices during a demand peak, even if all that was happening was a demand shift towards cheaper products. On the other hand, by using the fixed weight index it is as if we were forcing consumers not to change their quality-level choice. The measure obtained in this way will thus yield a better measure of the price movements, since it will clean price movements from demand shifts.

To make clearer the difference between the two price indexes, consider these two examples. First, imagine that practised prices are constant but consumers are shifting towards less (more) expensive products. In this case, the variable price index would capture a decrease (an increase) in prices because the market share of less (more) expensive products is going up. The fixed weights index, however, would not change, since the market shares are kept constant. Now, imagine that practised prices are going down and there is no quality or brand-level demand shift. In this case, the fixed and variable weight index would coincide. Finally, consider the situation where, in a given period, practised prices decrease and there is a demand shift towards cheaper products. The variable weight index would decrease both because prices decrease and because the market share (the weight) of cheaper products decreases. The variable weight price index, therefore, will overstate the price decrease. What about the fixed price index? Since market shares are held fixed, the fixed weight index will decrease solely because prices are lower. Therefore, given prices and *Implication 2*, we have:

Implication 3 Category prices measured by a variable price index vary more than prices as measured by a fixed price index.

If firms behavior and consumer mix did not change during demand peak, the implications of the model of quality-level demand shift could be directly tested for a product category which experiences an exogenous demand peak and for which we can identify two distinct quality levels. However, assuming that firms do not react to the demand increase or that there are no new consumers entering the market is too strong. Below, we discuss which supply reactions could jeopardize identification of the model results and how to test whether they are effective. We also discuss the effects of new consumers entering the market.

2.3.2 Identification Issues

In this section, we discuss what could be happening in the supply side or in the consumer mix that would generate the same testable implications of the quality-shift demand model. We also derive ways to test if the alternative supply side story is accepted or rejected by the data.

Supply

We have described the behavior of consumers during a seasonal demand peak *taking prices as given*. It is out of the scope of this paper to model contracts between producers and retailers and their behavior[4]. However, we cannot ignore that an annual period of exogenous increase in demand, as is the case of summer for the demand for ice-cream in our empirical application, is anticipated by firms and they will certainly react to it.

In the data, we observe prices and demand choices made in equilibrium, i.e., after both consumers and firms reacted to the demand increase. For certain ranges of price variation, the above results on the behavior of consumers may be invalid. Therefore, we need to be careful when testing the implications of the model. Clearly, we should reject the quality choice demand model if we reject the model implications: either price variation during the demand peak falls into the range that invalidates our results or our description of consumer behavior is wrong. However, not rejecting the model implications is not sufficient for not rejecting the model. We have to make sure the empirical results are not being driven by the behavior of firms but by the quality-choice shift in demand.

We are specially concerned with two potential changes in the behavior of firms: entry and product advertisement to attract customers to the store, as formalized in loss-leader models.

In some cases, periods of seasonal exogenous demand peak may be periods when it is optimal for firms to introduce new products. Take the example of ice-cream during summer. The increase in demand is an annual event and it lasts for a relatively long time. Aware of the changes in demand behavior at this period, firms may decide to commercialize a set of products that are not in the market the rest of the year[5]. The introduction of these demand-peak exclusive products may represent a source of price variation because it may increase competition among

[4] We could have, for instance, considered a simple Cournot-Nash duopoly model. But even in this simple setting, comparative statistics are an algebraic nightmare. More importantly, the benefits of modeling the supply side in this simple way are low because the model does not address entry nor loss-leader strategies, which are our main concern as explained below.

[5] Notice that in our model, the only differentiation dimension is quality and we restrict the number of quality levels to two. Therefore, in the consumers' point of view, entry does not exist since any new product will be classified under product A (high quality) or B (low quality).

close substitutes or simply because new products are cheaper than existing ones, driving average category prices down.

In loss-leader models, retailers advertise products at low prices to attract customers to the store. Once at the store, consumers find out about unadvertised prices, which are set equal to reservation prices. Given sufficiently high transportation costs, consumers purchase the advertised product and the rest of the basket (at reservation prices) at the same store. CKR show that in the particular Lal and Matutes (1994)'s version of the model, non-negativity constraints imply that retailers choose to advertise those products that have an idiosyncratic demand peak at the period. Lal and Matutes and CKR do not consider product differentiation. However, we can imagine that the same logic could be applied to a differentiated product market and that the low quality product could be chosen for advertisement. In this scenario, the price of the low quality product would go down, and the market share of the low quality product would go up due to firm's behavior and not consumers' quality shift during demand peaks.

Entry and the loss-leader model can thus lead firms to set lower prices during the demand peak. Assume that the price p_L of the low quality product decreases more than the price p_H of the high quality product. Then, in equilibrium, the relative demand for the low quality product should increase solely as a response to price variations. That is, even if the demand shock does not alter consumer choices, an increase in the market share of the low quality product could be observed as a response to firms price setting behavior. Fortunately, there is a straightforward way to test for this possibility. If the effect of the demand peak on the price of the low quality product is not stronger than on the price of the high quality product, we can safely conclude that the increase in the market share of the low quality product does not result from changes in the supply side of the market, but originate on the demand side. Additionally, if consumers are more price-elastic during demand peaks, then there is direct evidence against the supply side based story.

Consumer Mix

Assume that there is a group of consumers who only enter the market during the demand peak and have higher cross-price elasticities than the year-round consumers. Then observing higher price-elasticities during the summer does not necessarily mean that consumers are *behaving* differently but that they *are* different. We compare price-elasticities during the demand peak for both groups of consumers to check whether one group is more responsive to prices than the other (see the Empirical Section where we find evidence that there is not significant differences in the way both groups react to prices). Furthermore, to avoid endogeneity bias in the demand estimation, we include a "only-summer consumers" fixed effect in estimated regressions. We also let the price coefficient vary between only summer and year-round consumers in the demand equation.

2.4 Data and Descriptive Statistics

The database is a representative survey of households distributed across all regions of France. We have information on three years: 1999, 2000, and 2001. Each household was given a scanner with which it should register every food product purchased. For each product purchased, we have information on its brand and characteristics, including price and pack size, label, the date of the purchase and the brand of the retailer where it was purchased. We also have comprehensive information on household demographics.

The product category we study is ice-cream, which presents a demand peak during the summer. We consider the sample of households that purchased ice-cream at least once during the three years. Table 1 displays some basic descriptive statistics for the whole sample and the Summer and Not Summer subsamples. Those statistics are total purchased quantity, average daily quantity, mean individual quantity purchased per purchase occasion, the average category price, and the number of observations, where one observation is one purchase made by the household. Quantities are measured in liters and prices are measured in euros per liter.

The statistics on average daily purchases make clear that demand increases during the summer. The increase in ice-cream purchases, however, does not seem to pressure up prices. Actually, when uncontrolled for as in Table 1, average prices are lower during the Summer.

Table 1: Purchased Quantities and Prices on Ice-Cream Category

	Whole Time Span	**Summer**	**Not Summer**
Total Purchased Quantity (l)	54154.88	21657.65	32497.23
Daily Average Quantity (l)	50.004	79.624	40.070
Mean Individual Quantity per Purchase (l)	1.130	1.155	1.114
Average Price (€/l)	3.104	2.974	3.233
Number of Obs	47926	18749	29177

Part of the increase in demand seems to come from an increase in the purchased quantity per purchase occasion. It could also come from more people purchasing ice-cream during the Summer. Indeed, as can be seen in Table 2, around 12% of the households in the sample ("only summer" households) only purchase ice-cream during the Summer[6].

Table 2: Number of "Only Summer" and "All year" households on Ice-Cream Category

	Freq	**Percent**
all year hh	4381	87.11
only summer hh	648	12.89
Total	5029	100.00

Although ice-cream comes in different brands and flavors, we concentrate in only one differentiation dimension: whether the product is a national or a store brand. We call this dimension "quality level". The store brands represent the low quality level and the national brands, the high quality level. Average prices during summer and during the rest of the year for national and store brands are in Table 3. National brands are significantly more expensive than store brands in both sub-periods.

Table 3: Average Price per Quality Level

	Average Price (€/l)	
	Not Summer	**Summer**
National Brand	3.736	3.400
Store Brand	2.307	2.241

Further evidence on the price differential between store and national brands can be inferred from Graphs 1 to 4 in the Appendix, which show average weekly prices of the four most popular ice-cream flavors (vanilla, coffee, chocolate, and rum raisin), sold at the five most popular stores.

[6] In the empiriral analysis, to avoid endogeneity caused by unobserved heterogeneity, we control for households only purchasing during the summer. We also check if and which observable characteristics explain selection into the "only summer" household group.

The darker spots are the average prices of store brands, while the lighter colored spots represent the consistently higher prices of the national brands.

Given our product definition which includes only the quality dimension, entry is not an issue. New products sold only during summer will fall either in the low quality product or in the high quality product, depending on whether it is of a national or store brand. Certainly, this may be an over-simplified manner of treating product differentiation. We claim it is enough to identify consumers' quality choice behavior. Anyway, in the next section we do a robustness check where we consider a more precise definition of product, which includes store, brand and product characteristics (cream, sorbet or yogurt). With product thus defined, entry may exist. Unfortunately, our data is at the individual level therefore we do not have information on which products were offered but only on which products were purchased. Hence we cannot know for sure if a certain product is offered only during the summer. We then consider as "only summer products" those products that were purchased only during summer (having in mind that at least part of them could be on shelves during the whole year but not chosen by sampled households). In Table 4 below we have how frequently households purchased "only summer" products, the percentage of total purchases represented by only summer products, as well as average prices.

Table 4: Products only purchased during the summer

			Average Prices	
	Nb of Purchases	**Percentage**	**Summer**	**Not Summer**
only summer prod=0	1370	88.56	3.510	3.665
Store Brand	*266*	*19.42*	*2.437*	*2.499*
National Brand	*1104*	*80.58*	*3.781*	*3.937*
only summer prod=1	177	11.44	3.561	
Store Brand	*29*	*16.38*	*2.502*	
National Brand	*148*	*83.62*	*3.769*	
Total	1547	100.00		

Note: Here, product is defined as a combination of store where it was purchased, brand and "type" (cream, sorbet, or yogurt); only summer=1 stands for products that were purchased only during the summer.

Average prices of entrants are higher than average prices of incumbents. This means that average price movements during summer cannot be explained uniquely by entry of products whose prices are pushing the average down, although it may be the case that entry is affecting competition and price setting behavior of firms.

2.5 Empirical Analysis

In this section, we test the model implications and estimate the demand. Structural estimates of demand parameters are used to compute price-elasticities. We also check if the alternative hypothesis discussed in the section on identification issues can be discarded. Finally, we repeat the main tests using a different and less restrictive definition of product, confirming that our results are robust.

2.5.1 Reduced-Form Tests

Before testing the model Implications, we have to verify that consumer quality-choices can be identified. As discussed in Section 2, if the price of the low quality product decreases more than the price of the high quality product, then the migration towards lower quality products may be just a demand response to the change in the way firms set prices. We can directly test the effect of summer on the prices of the high and low quality product by running a regression of average daily prices per quality level on the interaction of a low quality product dummy with the summer dummy. Estimated coefficients are in Table 5. The coefficient of the interaction of store brand and summer is not significant, indicating that both quality levels experience similar price variation during the demand peak, and that we should not worry about non-identification of consumer quality shifts. Notice also that the coefficients of the store brand and summer dummies are negative and significant, confirming that store brands are indeed less expensive than national brands and that summer is a period of lower prices.

Table 5: Effect of Summer on Prices per Quality Level

Independent Variable	**Coefficients**
store brand	-1.302
	(8.68)**
summer	-0.293
	(2.44)*
store brand X summer	0.221
	(1.30)
store brand X weekend	-0.107
	(0.72)
store brand X Xmas	0.065
	(0.23)
store brand X year2000	-0.197
	(1.10)
store brand X year 2001	-0.029
	(0.16)
seasonal dummies	Yes
constant	Yes
Obs	2105
R^2	0.15

Note: Absolute value of t statistics in parentheses; + significant at 10%; * significant at 5%; ** significant at 1%. Dependent variable is daily average price of store and national labels. One observation is one day. Seasonal dummies are dummies indicating season of the year, year, Christmas period, and weekends.

We now test implications 2 and 3 of the model. *Implication 2* says that during demand peaks, the market share of the low quality product increases while the market share of the high quality product decreases. Table 6 show results from the regression of the daily volume market shares of the high quality products on seasonal dummies. As predicted by the model, the coefficient of summer is negative and significant.

Table 6: Effect of Summer on Volume Market Shares

	Market Share (in volume) of National Brands
summer	-0.035
	(3.42)**
weekend	0.009
	(0.99)
Christmas	0.019
	(1.15)
year=2000	0.034
	(3.15)**
year=2001	0.120
	(11.24)**
constant	0.596
	(6.64)**
Obs	1070
R^2	0.12

Note: Absolute value of t statistics in parentheses; + significant at 10%; * significant at 5%; significant at 1%. Dependent variable is market share (in volume) of (1) store brand and (2) national brand. Year dummy for 1999 is base-category. One observation is one day.

To test for *Implication 3,* which says that prices vary more if measured by a variable price index than by a fixed price index, we regress price indexes on seasonal dummies. Estimated coefficients for the variable and fixed price index regressions are in the first and second column, respectively, of Table 7 (the period horizon for the indexes is one day).

Table 7: Effect of Summer on Price Indexes

	Variable PI	**Fixed PI**
summer	-0.210	-0.136
	(2.30)*	(1.30)
weekend	0.101	-0.040
	(1.28)	(0.44)
Christmas	0.128	0.100
	(0.86)	(0.58)
year=2000	0.258	0.187
	(2.70)**	(1.71)
year=2001	0.208	0.023
	(2.17)*	(0.21)
Constant	3.030	3.093
	(37.71)**	(33.59)**
Obs	1083	1083
R^2	0.02	0.01

Note: Absolute value of t statistics in parentheses; + significant at 10%; * significant at 5%; significant at 1%. Dependent variable is daily prices as measured by (1) variable price index, (2) fixed price index. Year dummy for 1999 is base-category. One observation is one day. Product definition: whether it is a national brand or a store brand.

The effect of summer on the variable price index is negative and significant at 1%. It is also stronger in magnitude and more significant than the summer effect on the fixed price index. Results thus confirm model implications, indicating that prices remain fairly constant during summer (fixed price variation is not significant) but the weight of less expensive products increases.

2.5.2 Structural Demand and Price-Elasticities

We fully estimate structural demand using equation (2.5) and assuming the taste-for-quality parameter β is logistically distributed[7]. Estimated parameters are then plugged in (2.6) for the calculation of cross-price elasticities.

To account for the seasonal demand shock, we interact $(p_{Ht} - p_{Lt})$ with a summer dummy variable. However, if we only included the interaction of prices with the summer dummy, we would be forcing the demand shock to be equal to zero at all other periods, which is too strong. Therefore, we include other seasons dummies covering all periods of the time span. The coefficient of the price difference interacted with the summer dummy will give us the magnitude of the shock to utility. But notice that the shock v_t is identifiable up to the constant α, the marginal utility of income (fortunately, in the price elasticity formula, the shock also appears multiplied by α).

To allow for heterogeneity in price-responsiveness, we also interact $(p_{Ht} - p_{Lt})$ with household characteristics. This means that the marginal utility of income is allowed to vary across observably different consumers.

Note that we only have data on prices paid by consumers. Therefore, if the consumer chose to purchase the low (high) quality product, we have information on how much he paid for it, but not on the price of the high (low) quality product which she did not purchase. Hence, we

[7] We repeated the exercise assuming that β was normally distributed, using thus a probit model to estimate demand parameters. Qualitatively, results do not change. Coefficient estimates and implied price-elasticities for the probit model are presented in Tables A1 to A3 in the Appendix.

construct p_{At} as the average price paid by other consumers for the high quality product at day t.

Table 8 shows results from the individual demand estimation. We include household characteristics, namely size of the household, presence of a child younger than six years of age, presence of a child younger than 16 years of age, and age and level of education of the household head. We also include product characteristics (store, brand, and type) and seasonal dummies. The coefficient of $ln\left(p_{Ht}/p_{Lt}\right)$ is restricted to be equal to 1. In the first column, we have results from an estimation with restricted price-response heterogeneity, where the price difference coefficient is allowed to vary across consumers only with respect to the "only summer households" dummy. In column (2), on the other hand, the price difference coefficient is allowed to vary with respect to income, household size, and presence of a child of 16 or less years of age.

Table 8: Estimation of Demand Parameters

Independent Variables	$\mathbf{P(y_{iLt}=1)}$	
	(1)	(2)
$(p_{Ht}-p_{Lt})$	0.771	1.299
	(19.86)	(6.40)**
(summer = 1)*$(p_{Ht}-p_{Lt})$	0.507	0.511
	(9.92)**	(9.99)**
(spring =1)*$(p_{Ht}-p_{Lt})$	-0.771	-0.674
	(19.84)**	(16.81)**
(autumn = 1)*$(p_{Ht}-p_{Lt})$	-0.115	-0.110
	(2.13)*	(2.03)*
only summer hh	0.348	0.357
	(3.95)**	(4.05)**
(only summer hh = 1)*$(p_{Ht}-p_{Lt})$	-0.098	-0.110
	(1.03)	(1.15)
constant $\left(-\bar{\beta}\right)$	-1.691	-2.050
	(7.18)**	(7.22)**
$\ln\left(\frac{p_{Ht}}{p_{Lt}}\right)$	restricted to 1	
income*$(p_{Ht}-p_{Lt})$	no	yes
children below 16*$(p_{Ht}-p_{Lt})$	no	yes
hh size*$(p_{Ht}-p_{Lt})$		
hh characteristics	yes	yes
seasonal dummies	yes	yes
other product characteristics	yes	yes
Obs	47407	

Note: Absolute value of t statistics in parentheses; + significant at 10%; * significant at 5%; significant at 1%. Logit model (see Appendix for Probit results). Dependent variable is a discrete binomial variable equal to

1 if consumer chose to purchase the low quality product (store brand) and zero otherwise. One observation is one purchase made by one household. Price difference allowed to vary with: (1) whether the household only purchases during the summer, (2) whether the household only purchases during the summer and other household characteristics (income, household size, and dummy indicating presence of a child under 16 years of age). Controls for household characteristics are: dummy indicating whether there is a child under 6 years of age living in the household, dummy indicating whether there is a child under 16 years of age living in the household, region of residence, age of the household head, level of education of the household head, income, dummy indicating whether the household has a freezer, size of the household. Seasonal dummies indicate the four season. Product characteristics are store where product was purchased and type of ice-cream (creme, sorbet or yogurt).

The coefficient of the interaction of summer and price differences is positive and significant and stronger than that for other periods in both equations, confirming that summer is a period of positive demand shock for the demand of ice-cream. This result is clearly not driven by consumers entering the market during summer since the coefficient of the interaction between prices and the only summer households is not significant. Nevertheless, the coefficient of the only summer households fixed effect is significant, confirming the importance of controlling for unobservable heterogeneity. Finally, the estimate of the mean taste for quality is positive, bringing evidence that vertical differentiation describes well the type of product differentiation in this market.

In the Appendix, we show demand parameters estimation results for alternative specifications of the utility function and the distribution of the taste parameter. Main results are robust to alternative specifications. Specifically, we consider a CES utility function family where we vary the relative risk aversion coefficient γ, and compare log-likelihoods under these different values. The value of γ which maximizes the log-likelihood is 1.25. The log-likelihood thus obtained is very close to the second highest log-likelihood, which is obtained when γ converges to 1 and the utility function is of the logarithm form (as in Table 8). In what concerns the distribution of the taste-parameter, we consider it to be normally distributed and estimate parameters using a Probit model.

Let Z_i be the household characteristics with respect to which price sensitivity is allowed to vary. The average values of $\alpha(Z_i)v_t$ implied by the demand coefficient estimates are shown in Table 9. For each season, we present the average estimate yielding from the restricted heterogeneity model, where $\alpha(Z_i)$ is allowed to vary only in relation to the "only summer households" indicator, and from the generalized heterogeneity model, where other household characteristics are interacted with the price difference.

Table 9: Average Estimates of $\alpha(Z_i)v_t$

Season	**Obs**	**Mean**	**Std Dev**	**Min**	**Max**
Summer	51				
restricted heterogeneity		*1.230*	*0.0496*	*1.180*	*1.278*
general heterogeneity		*1.160*	*0.321*	*0*	*1.488*
Autumn	26				
restricted heterogeneity		*0.656*	.	.	.
general heterogeneity		*0.616*	*0.217*	*0*	*0.867*
Winter	26				
restricted heterogeneity		*0.771*	.	.	.
general heterogeneity		*0.717*	*0.243*	*0*	*0.977*
Spring	25				
restricted heterogeneity		*-0.000*	.	.	.
general heterogeneity		*0.098*	*0.124*	*-0.091*	*0.302*

Plugging estimated values for $\alpha(Z_i)v_t$ into (2.6) and considering prices of the low and high quality products at their seasonal average level, we get the estimated demand cross-price elasticities in Table 10. One observation is a "type" of household, where type is defined by income, presence of a child of 16 or less years of age, and whether the household purchases only during the summer.

Table 10: Seasonal Cross-price Elasticities

Season	**Obs**	**Mean**	**Std Dev**	**Min**	**Max**
Summer	51				
restricted heterogeneity		*2.605*	*0.612*	*1.783*	*5.450*
general heterogeneity		*2.507*	*0.840*	*0.503*	*5.581*
Autumn	26				
restricted heterogeneity		*1.852*	*0.277*	*1.419*	*2.514*
general heterogeneity		*1.790*	*0.529*	*0.529*	*2.725*
Winter	26				
restricted heterogeneity		*1.856*	*0.258*	*1.364*	*2.340*
general heterogeneity		*1.777*	*0.492*	*0.497*	*2.497*
Spring	25				
restricted heterogeneity		*0.605*	*0.063*	*0.475*	*0.733*
general heterogeneity		*0.781*	*0.251*	*0.446*	*1.219*

Note: One observation is one day. Prices used for calculation of cross-price elasticity are daily averages.

The cross-price elasticities are clearly higher during the demand peak, as predicted by the model. This result is a direct evidence against CKR's loss leader model and, in general, against models based on supply side behavior explanations for the price decrease during periods of exogenously high demand. Notice also that in all seasons except spring, cross-price elasticities are underestimated when do not allow for consumer heterogeneity.

2.5.3 Robustness Check

In this section, we consider a less restrictive definition of product, which includes store, brand and product characteristics. We test if the main results of the model, namely, the effect of summer on price indexes and the higher price-responsiveness during the demand peak, are still valid. Table 11 presents estimated coefficients in the regression of daily variable price index (second column), and fixed price index (last column) on seasonal dummies. Results remain basically the same. Summer coefficients on the variable price index regressions are negative as before. In the fixed price index equation though, the summer coefficient is positive. Interpretation does not change much however: the market shares of cheaper products increases during summer, pushing down the category average price, although firms may be setting higher prices.

Table 11: Effect of Summer on price indexes when considering a less restrictive definition of product

	Variable PI	**Fixed PI**
summer	-0.134	0.293
	(3.30)**	(10.37)**
weekend	0.166	-0.240
	(4.70)**	(9.85)**
Christmas	0.142	-0.089
	(2.14)*	(1.93)
year=2000	0.181	0.052
	(4.23)**	(1.75)
year=2001	0.206	0.036
	(4.82)**	(1.22)
Constant	2.765	0.763
	(77.11)**	(30.74)**
Obs	1083	
R^2	0.06	0.17

Note: Absolute value of t statistics in parentheses; + significant at 10%; * significant at 5%; significant at 1%. Dependent variable is daily prices as measured by (1) variable price index, (2) fixed price index. Year dummy for 1999 is base-category. One observation is one day. Product definition is combination of store where it was purchased, brand and ice-cream "type" (cream, sorbet, yogurt).

To verify if the price-responsiveness of consumers is higher during summer under the less restrictive product definition, we estimate a multinomial logit model of product choice where prices are interacted with the summer dummy. A negative and significant coefficient of the interaction of summer and prices indicate that price-responsiveness is indeed higher during the demand peak. We also interact prices and the dummy indicating whether the household only purchases ice-cream during the summer. For dimensionality reasons, we consider only the 26 products with highest market share. As before, the regression includes controls for household characteristics and product characteristics (whether the product is a store or national brand and type of the ice-cream).

Table 12: Price-Responsiveness with less restricted product definition

Explanatory Variables	**Mlogit prob of choosing product j**
price	-0.003
	(0.77)
summer	0.028
	(0.62)
summer X price	-0.036
	(2.70)**
only summer hh	0.031
	(0.24)
only summer hh X price	-0.056
	(1.42)
hh characteristics	yes
product characteristics	yes
constant	yes
Obs	47926

Note: Absolute value of t statistics in parentheses; + significant at 10%; * significant at 5%; significant at 1%. Dependent variable is probability of choosing product $j \in (1, ..., 27)$. Model is multinomial logit. Product definition is combination of store, brand and ice-cream "type" (cream, sorbet, yogurt). Household characteristics included are: whether household has a freezer, presence of a child of less than 6 years of age, presence of a child of less than 16 years of age, age and education of household head, household size, and income. Product characteristics are whether it is a store label and type.

Estimated coefficients in Table 12 confirm results found earlier. Namely, consumers are indeed more price sensitive during summer, as indicated by the significant coefficient of the interaction of prices and summer. Furthermore, households that only purchase during the summer seem to respond to price variations in the same way as other households: the coefficient for the interaction of only summer households and prices is not significant.

In sum, the results obtained in this section using a different and less restricted definition of product back the results found earlier, leading to the non rejection of the model of quality choice shift.

2.6 Conclusion

Recently, empirical evidence has shown that periods of high exogenous demand are frequently accompanied by lower prices. Explanations for the phenomena can be categorized in two main groups: one that defends that the decrease in prices is due mainly to changes in consumer behavior during demand peaks (higher price-elasticity), and another who argues that the supply side is in the origin of the price variation.

This paper contributes to the literature by formalizing consumer behavior in a model which shows that a positive exogenous shock to demand increases cross-price elasticities, shifting consumers quality choices towards cheaper lower quality products. The model implies that the market share of lower quality products increases during demand peaks. The increase in the weight of cheaper products decreases average category prices. We also reject tests of alternative explanations based on firm behavior

Implications of the model are tested using a comprehensive database on individual purchase choices of ice-cream, which have a peak during the summer. We estimate demand and use structural estimates of demand parameters to compute price elasticities. Results highly corroborate the predictions of the model.

2.7 Appendix

Proposition 1 *Given prices, the cross-price elasticity (2.6) of the probability of purchasing product L, the low-quality product, increases with the magnitude of the utility shock v_t. That is,* $\frac{d\varepsilon_H^L}{dv_t} \geqslant 0$.

Proof.

$$\begin{aligned}
\frac{d\varepsilon_H^L}{dv_t} &= p_{Ht}\frac{d}{dv_t}\left(\frac{d\tilde{\beta}_t}{dp_{Ht}}\right)(1-P_{it}) - p_{Ht}\frac{d\tilde{\beta}_t}{dp_{Ht}}\frac{dP_{it}}{dv_t} \\
&= p_{Ht}\frac{d}{dv_t}\left(\frac{d\tilde{\beta}_t}{dp_{Ht}}\right)(1-P_{it}) - p_{Ht}\frac{d\tilde{\beta}_t}{dp_{Ht}}\frac{d\tilde{\beta}_t}{dv_t}P_{it}(1-P_{it}) \\
&= p_{Ht}\left(1-P_{it}\right)\left[\frac{d}{dv_t}\left(\frac{d\tilde{\beta}_t}{dp_{Ht}}\right) - \frac{d\tilde{\beta}_t}{dp_{Ht}}\frac{d\tilde{\beta}_t}{dv_t}P_{it}\right]
\end{aligned}$$

The sign of $\frac{d\varepsilon_H^L}{dv_t}$ is determined by the sign of $w \equiv \frac{d}{dv_t}\left(\frac{d\tilde{\beta}_t}{dp_{Ht}}\right) - \frac{d\tilde{\beta}_t}{dp_{Ht}}\frac{d\tilde{\beta}_t}{dv_t}P_{it}$. We have:

$$\frac{d\tilde{\beta}_t}{dp_{Ht}} = \frac{1+\alpha v_t p_{Ht}}{p_{Ht}}$$

$$\frac{d}{dv_t}\left(\frac{d\tilde{\beta}_t}{dp_{Ht}}\right) = \alpha$$

and

$$\frac{d\tilde{\beta}_t}{dv_t} = \alpha\left(p_{Ht}-p_{Lt}\right)$$

Hence

$$\begin{aligned}
w &= \alpha - \alpha\frac{1+\alpha v_t p_{Ht}}{p_{Ht}}\left(p_{Ht}-p_{Lt}\right)P_{it} \\
&= \frac{\alpha}{p_{Ht}}\left[p_{Ht} - \left(1+\alpha v_t p_{Ht}\right)\left(p_{Ht}-p_{Lt}\right)P_{it}\right] \\
&= \frac{\alpha}{p_{Ht}}\left[p_{Ht} - \alpha v_t p_{Ht}\left(p_{Ht}-p_{Lt}\right)P_{it} - \left(p_{Ht}-p_{Lt}\right)P_{it}\right] \\
&= \frac{\alpha}{p_{Ht}}\left[p_{Ht}\left(1-\alpha v_t\left(p_{Ht}-p_{Lt}\right)P_{it}\right) - \left(p_{Ht}-p_{Lt}\right)P_{it}\right] \\
&\geqslant \frac{\alpha}{p_{Ht}}\left[p_{Ht}\ln\left(\frac{p_{Ht}}{p_{Lt}}\right) - \left(p_{Ht}-p_{Lt}\right)\right] \\
&\geqslant 0
\end{aligned}$$

where, the first inequality follows from the assumption that markets are covered, which implies that $1-\alpha v_t\left(p_{Ht}-p_{Lt}\right) \geqslant \ln\left(\frac{p_{Ht}}{p_{Lt}}\right)$ and the fact that $P_{it} \in (0,1)$. We now prove the second inequality (that says $w \geqslant 0$). Let $x \equiv 1-\frac{p_{Lt}}{p_{Ht}}$ (notice that $0 < x \leq 1$since $p_{Lt} \leq p_{Ht}$). With this change of variables, $p_{Ht}\ln\left(\frac{p_{Ht}}{p_{Lt}}\right)-\left(p_{Ht}-p_{Lt}\right) \geqslant 0$ becomes $\frac{1}{1-x} \geqslant e^{x}$ which is analogous to

$$1-x < e^{-x} \Leftrightarrow e^{-x}+x < 1$$

Now, let $h(x)=e^{-x}+x$. For $x=0$, we have $h(0)=1$. Since $h'(x)>0$ for $x>0$, we have that $h(x)>1$ for $x>0$. ■

Proposition 2 *Given prices, the market share in volume of the low-quality product L increases at periods of high realizations of the utility shock v_t, i.e.,* $\frac{dS_{Lt}}{dv_t} \geqslant 0$.

$$
\begin{aligned}
\frac{dS_{Lt}}{dv_t} &= \frac{\frac{dD_{Lt}}{dv_t}\left(D_{Lt}+D_{Ht}\right)-D_{Lt}\left(\frac{dD_{Lt}}{dv_t}+\frac{dD_{Ht}}{dv_t}\right)}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{D_{Ht}\frac{dD_{Lt}}{dv_t}-D_{Lt}\frac{dD_{Ht}}{dv_t}}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{(1-P_{it})q_{Ht}\left(\frac{dP}{dv_t}q_{Lt}+P_{it}\frac{dq_{Lt}}{dv_t}\right)-P_{it}q_{Lt}\left(-\frac{dP}{dv_t}q_{Ht}+(1-P_{it})\frac{dq_{Ht}}{dv_t}\right)}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{(1-P_{it})q_{Ht}\left[\frac{d\tilde{\beta}}{dv_t}P_{it}(1-P_{it})q_{Lt}+P_{it}\right]-P_{it}q_{Lt}\left[-\frac{d\tilde{\beta}}{dv_t}P_{it}(1-P_{it})q_{Ht}+(1-P_{it})\right]}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{q_{Ht}\left[\frac{d\tilde{\beta}}{dv_t}(1-P_{it})q_{Lt}+1\right]-q_{Lt}\left[-\frac{d\tilde{\beta}}{dv_t}P_{it}q_{Ht}+1\right]}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{\frac{d\tilde{\beta}}{dv_t}(1-P_{it})q_{Ht}q_{Lt}+q_{Ht}+\frac{d\tilde{\beta}}{dv_t}P_{it}q_{Ht}q_{Lt}-q_{Lt}}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{\frac{d\tilde{\beta}}{dv_t}q_{Ht}q_{Lt}-(q_{Lt}-q_{Ht})}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{\frac{\alpha(p_{Ht}-p_{Lt})}{\alpha^2 p_{Ht}p_{Lt}}(1+\alpha v_t p_{Ht})(1+\alpha v_t p_{Lt})-\left(\frac{1+\alpha v_t p_{Lt}}{\alpha p_{Ht}}-\frac{1+\alpha v_t p_{Ht}}{\alpha p_{Lt}}\right)}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{\frac{\alpha(p_{Ht}-p_{Lt})}{\alpha^2 p_{Ht}p_{Lt}}(1+\alpha v_t p_{Ht})(1+\alpha v_t p_{Lt})-\left(\frac{\alpha p_{Ht}+\alpha^2 v_t p_{Lt}\alpha p_{Ht}-\alpha p_{Lt}-\alpha^2 v_t\alpha p_{Lt}p_{Ht}}{\alpha^2 p_{Ht}p_{Lt}}\right)}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{\frac{\alpha(p_{Ht}-p_{Lt})}{\alpha^2 p_{Ht}p_{Lt}}(1+\alpha v_t p_{Ht})(1+\alpha v_t p_{Lt})-\left(\frac{\alpha p_{Ht}-\alpha p_{Lt}}{\alpha^2 p_{Ht}p_{Lt}}\right)}{\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{\alpha\left(p_{Ht}-p_{Lt}\right)\left[(1+\alpha v_t p_{Ht})(1+\alpha v_t p_{Lt})-1\right]}{\alpha^2 p_{Ht}p_{Lt}\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{\alpha\left(p_{Ht}-p_{Lt}\right)\left[1+\alpha v_t p_{Lt}+\alpha v_t p_{Ht}+\alpha^2 v_t^2 p_{Ht}p_{Lt}-1\right]}{\alpha^2 p_{Ht}p_{Lt}\left(D_{Lt}+D_{Ht}\right)^2} \\
&= \frac{\alpha\left(p_{Ht}-p_{Lt}\right)\left[\alpha v_t p_{Lt}+\alpha v_t p_{Ht}+\alpha^2 v_t^2 p_{Ht}p_{Lt}\right]}{\alpha^2 p_{Ht}p_{Lt}\left(D_{Lt}+D_{Ht}\right)^2} \\
&\geqslant 0
\end{aligned}
$$

Structural Demand Coefficient Estimates and Price-Elasticities for the Probit Model

To check if the main results related to demand estimation are dependent on distributional assumptions, we estimate demand now assuming that β is normally distributed. Results are in Table A1.

Table A1: Estimation of Demand Parameters for β_i Normal

Independent Variables	$\mathbf{P(y_{iLt}=1)}$	
	(1)	(2)
$(p_{Ht} - p_{Lt})$	0.218	0.473
	(12.98)**	(4.82)**
(summer = 1)*$(p_{Ht} - p_{Lt})$	0.191	0.193
	(8.65)**	(8.67)**
(spring =1)*$(p_{Ht} - p_{Lt})$	-0.238	-0.236
	(14.19)**	(13.79)**
(autumn = 1)*$(p_{Ht} - p_{Lt})$	-0.048	-0.049
	(2.05)*	(2.07)*
only summer hh	0.159	0.165
	(3.52)**	(3.62)**
(only summer hh = 1)*$(p_{Ht} - p_{Lt})$	0.008	0.001
	(0.17)	(0.02)
constant $\left(-\bar{\beta}\right)$	-0.853	-1.027
	(6.36)**	(6.69)**
$\ln\left(\frac{p_{Ht}}{p_{Lt}}\right)$	restricted to 1	
income*$(p_{Ht} - p_{Lt})$	no	yes
children below 16*$(p_{Ht} - p_{Lt})$	no	yes
hh size*$(p_{Ht} - p_{Lt})$	no	yes
hh characteristics	yes	
seasonal dummies	yes	
other product characteristics	yes	
Obs	47407	

Note: Absolute value of t statistics in parentheses; + significant at 10%; * significant at 5%; significant at 1%.

Probit model. Dependent variable is a discrete binomial variable equal to 1 if consumer chose to purchase the low quality product (store brand) and zero otherwise. One observation is one purchase made by one household. Price difference allowed to vary with: (1) whether the household only purchases during the summer, (2) whether the household only purchases during the summer and other household characteristics (income, household size, and dummy indicating presence of a child under 16 years of age). Controls for household characteristics are: dummy indicating whether there is a child under 6 years of age living in the household, dummy indicating whether there is a child under 16 years of age living in the household, region of residence, age of the household head, level of education of the household head, income, dummy indicating whether the household has a freezer, size of the household. Seasonal dummies indicate Product characteristics are store where product was purchased and type of ice-cream (creme, sorbet or yogurt).

As in the logit model, the coefficient of the interaction of summer and price differences is positive and significant and stronger than that for other periods in both equations. Furthermore, the interaction of prices and the only summer households is not significant.

Tables A2 and A3 below bring the average $\alpha_i v_t$ and the prices elasticities, respectively, for both the restricted and unrestricted heterogeneity model in the case of the probit model.

Table A2: Average Estimates of $\alpha_i v_t$ when β is Normal

Season	**Obs**	**Mean**	**Std Dev**	**Min**	**Max**
Summer	51				
restricted heterogeneity		*0.412*	*0.004*	*0.409*	*0.416*
general heterogeneity		*0.387*	*0.100*	*0*	*0.471*
Autumn	26				
restricted heterogeneity		*0.170*	.	.	.
general heterogeneity		*0.157*	*0.051*	*0*	*0.229*
Winter	26				
restricted heterogeneity		*0.218*	.	.	.
general heterogeneity		*0.202*	*0.063*	*0*	*0.278*
Spring	25				
restricted heterogeneity		*-0.020*	.	.	.
general heterogeneity		*-0.016*	*0.022*	*-0.060*	*0.042*

Table A3: Seasonal Cross-price Elasticities when β is Normal

Season	**Obs**	**Mean**	**Std Dev**	**Min**	**Max**
Summer	51				
restricted heterogeneity		1.271	0.259	0.941	2.470
general heterogeneity		1.234	0.312	0.504	2.433
Autumn	26				
restricted heterogeneity		0.938	0.120	0.750	1.207
general heterogeneity		0.917	0.164	0.523	1.240
Winter	26				
restricted heterogeneity		0.941	0.108	0.716	1.149
general heterogeneity		0.915	0.157	0.511	1.176
Spring	25				
restricted heterogeneity		0.567	0.057	0.450	0.683
general heterogeneity		0.575	0.070	0.449	0.720

Notice that the probit model, as the logit, yields price elasticities which are higher during the summer.

Structural Demand Coefficient Estimates and Price-Elasticities for the CES Utility Function Family

Let $u\left(q_{ijt} - v_t\right) = \frac{\left(q_{ijt}-v_t\right)^{1-\gamma}}{1-\gamma}$. In this case, the first order condition of the consumer problem is:

$$q_{ijt} = \left(\frac{1}{\alpha p_{jt}}\right)^{1/\gamma} + v_t$$

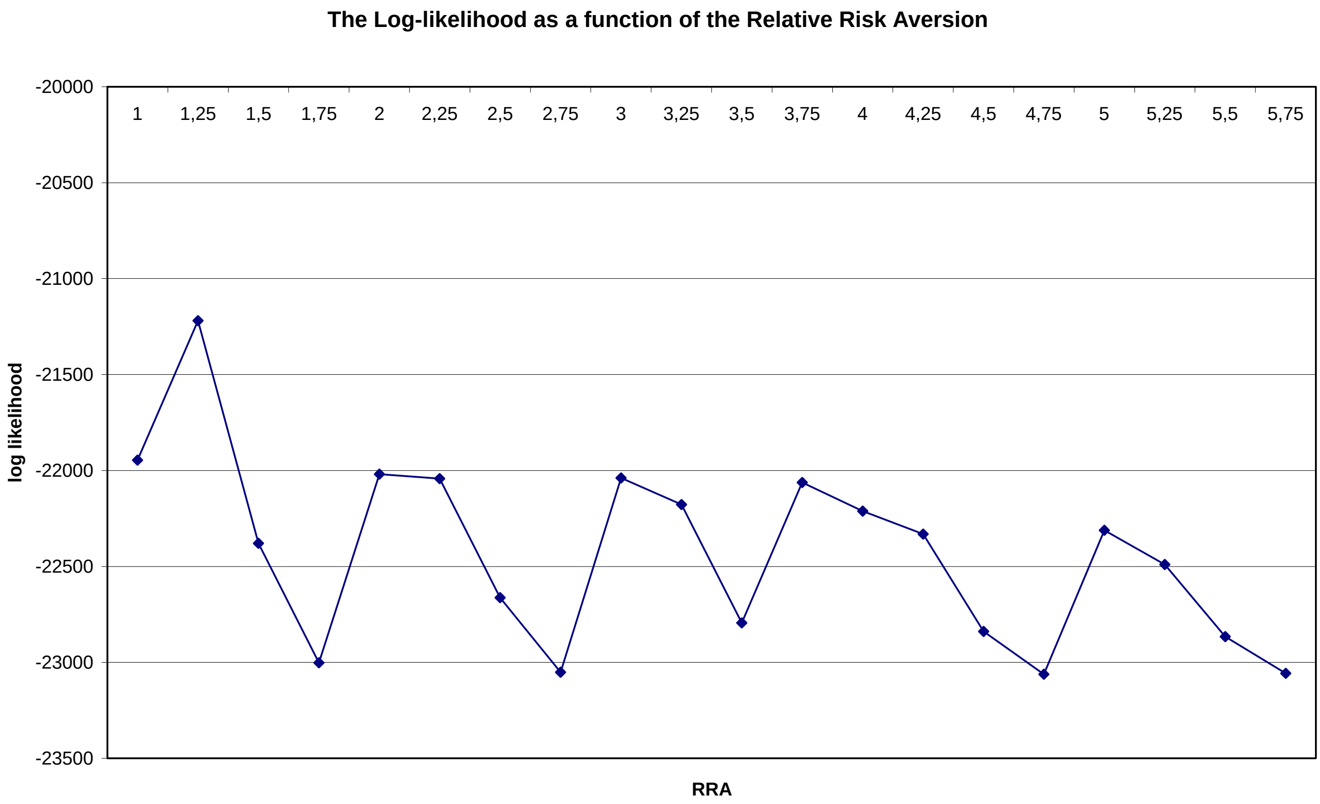


Figure 2.1: Log-likelihood in function of the Relative Risk Aversion

The indirect utility of consuming the high and the low quality products are, respectively:

$$V_{it}^{H} = \frac{1}{1-\gamma} \left(\alpha p_{Ht}\right)^{\frac{\gamma-1}{\gamma}} - \alpha v_t p_{Lt} + \bar{\beta} + \xi_i$$

$$V_{it}^{L} = \frac{1}{1-\gamma} \left(\alpha p_{Lt}\right)^{\frac{\gamma-1}{\gamma}} - \alpha v_t p_{Lt}$$

Therefore, the probability that consumer i will choose the low quality product at period t is:

$$P\left(y_{itL} = 1\right) = P\left(\xi_i \leq \frac{1}{1-\gamma} \alpha^{\frac{\gamma-1}{\gamma}} \left(p_{Ht}^{\frac{\gamma-1}{\gamma}} - p_{Lt}^{\frac{\gamma-1}{\gamma}}\right) + \alpha v_t \left(p_{Ht} - p_{Lt}\right) - \bar{\beta}\right)$$

If we assume ξ_i is logistically distributed and fix the value of γ, we can identify the parameters of the model. Furthermore, we can vary the γ value and choose which one maximizes the log-likelihood.

We let γ vary from 1.25 to 5.75 in intervals of 0.25. Figure 1 shows the log-likelihood in function of the different values of γ, including the case where γ tends to 1 and the utility function is of the logarithm form. The value that maximizes the log-likelihood is $\gamma = 1.25$, in which case the log-likelihood is equal to -21219.558. The second highest value of the log-likelihood is obtained when γ tends to 1 and the log-likelihood is equal to -21946.639, which is not very different from the highest value, indicating that assuming that the utility function is of the logarithm form is not a very strong assumption.

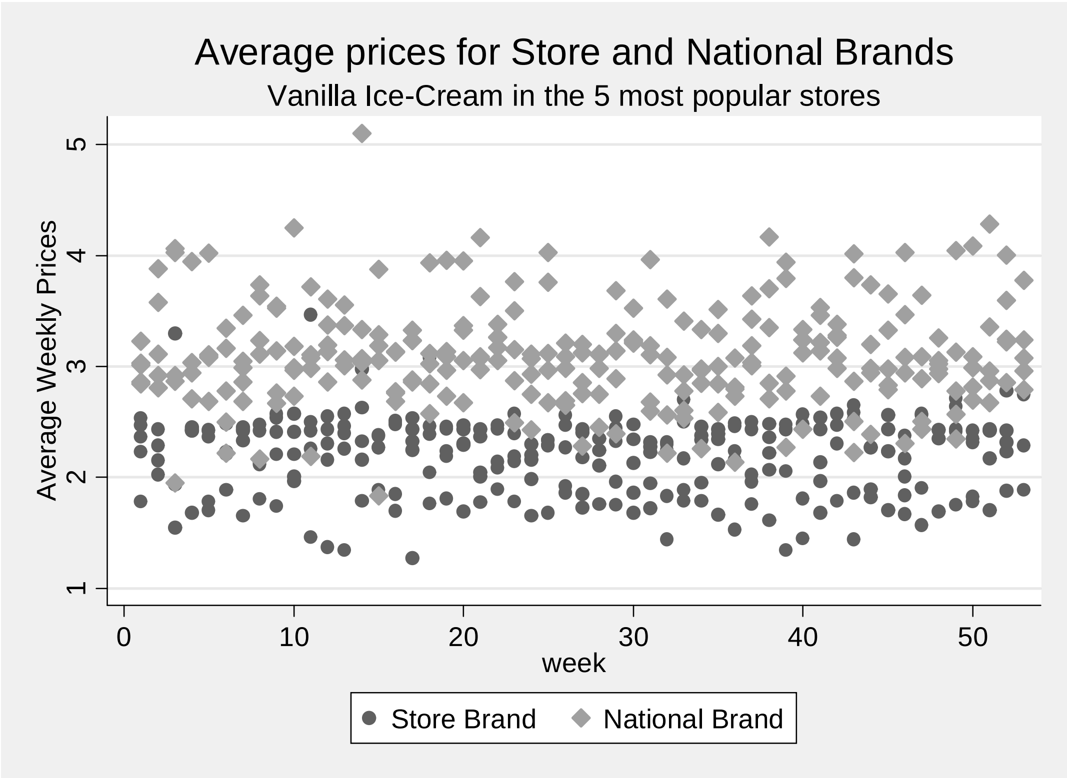
Results for the estimation of demand parameter when $\gamma = 1.25$ are shown in Table A4. Notice that there is no qualitative change in results with respect to the logarithm specification: coefficients have the expected sign, summer is a period of demand peak, and letting the price sensitivity change with the only summer households dummy does not alter estimation.

Table A4: Estimation of Demand Parameters for CES function with $\gamma = 1.25$

Independent Variables	**P($\mathbf{y}_{iLt}$= 1)**	
	(1)	(2)
$(p_{Ht}^{1.25} - p_{Lt}^{1.25})$	10.880	14.700
	(33.52)**	(43.5)**
$(p_{Ht} - p_{Lt})$	0.196	0.222
	(4.51)**	(1.09)
(summer = 1)*$(p_{Ht} - p_{Lt})$	0.405	0.404
	(8.43)**	(8.19)**
(spring =1)*$(p_{Ht} - p_{Lt})$	-0.286	-0.138
	(7.12)**	(3.35)**
(autumn = 1)*$(p_{Ht} - p_{Lt})$	0.094	-0.104
	(1.90)$^{+}$	(2.08)*
only summer hh	0.314	0.322
	(3.71)**	(3.72)**
(only summer hh = 1)*$(p_{Ht} - p_{Lt})$	-0.059	0.059
	(0.65)	(0.63)
constant $\left(-\bar{\beta}\right)$	-1.612	-1.897
	(6.69)**	(6.58)**
hh characteristics	yes	
income16*$(p_{Ht} - p_{Lt})$	no	yes
children below 16*$(p_{Ht} - p_{Lt})$	no	yes
hh size*$(p_{Ht} - p_{Lt})$	no	yes
seasonal dummies	yes	yes
other product characteristics	yes	yes
Obs	47407	47407

Note: Absolute value of t statistics in parentheses; + significant at 10%; * significant at 5%; significant at 1%. Logit model, utility specification: $u\left(q_{ijt} - v_t\right) = \frac{(q_{ijt}-v_t)^{1-\gamma}}{1-\gamma}$ with $\gamma = 1.25$. Dependent variable is a discrete binomial variable equal to 1 if consumer chose to purchase the low quality product (store brand) and zero otherwise. One observation is one purchase made by one household. Price difference allowed to vary with (1) whether the household only purchases during the summer, (2) whether the household only purchases during the summer and other household characteristics (income, household size, and dummy indicating presence of a child under 16 years of age). Controls for household characteristics are: dummy indicating whether there is a child under 6 years of age living in the household, dummy indicating whether there is a child under 16 years of age living in the household, region of residence, age of the household head, level of education of the household head, income, dummy indicating whether the household has a freezer, size of the household. Seasonal dummies indicate Product characteristics are store where product was purchased and type of ice-cream (creme, sorbet or yogurt).

Figure 2.2: Average Prices of Store and National Brands - Vanilla Ice-Cream in the 5 most popular stores



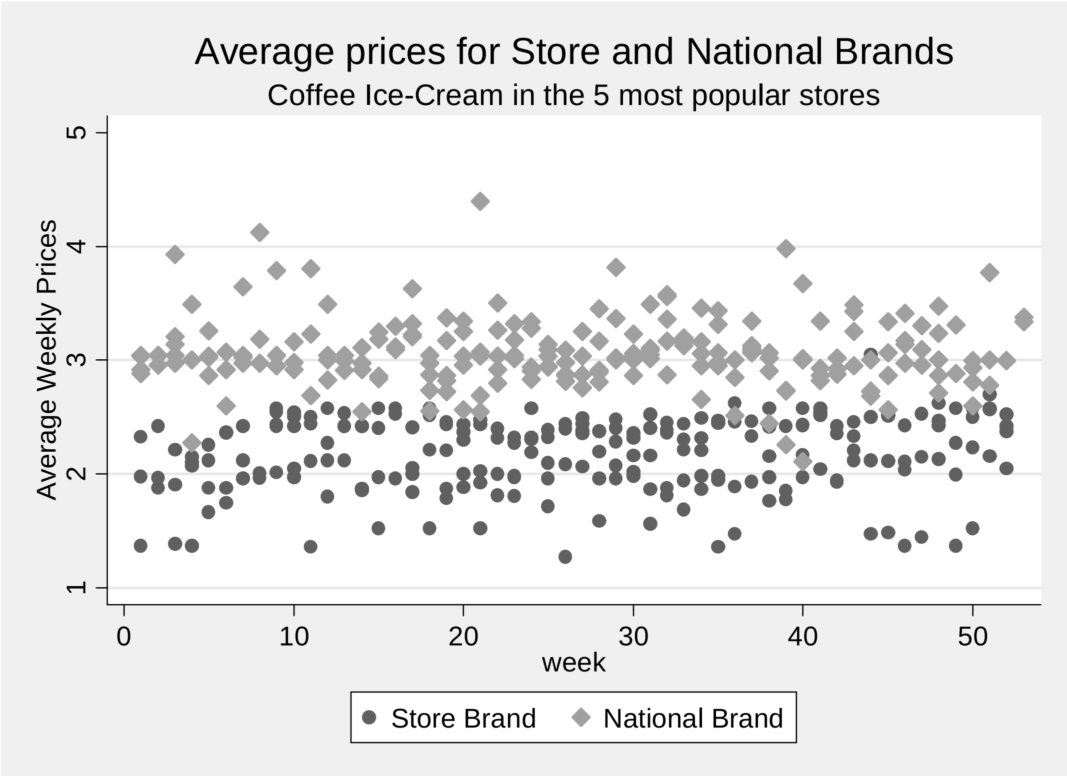


Figure 2.3: Average Prices for Store and National Brands - Coffee Ice-Cream in the 5 most popular stores

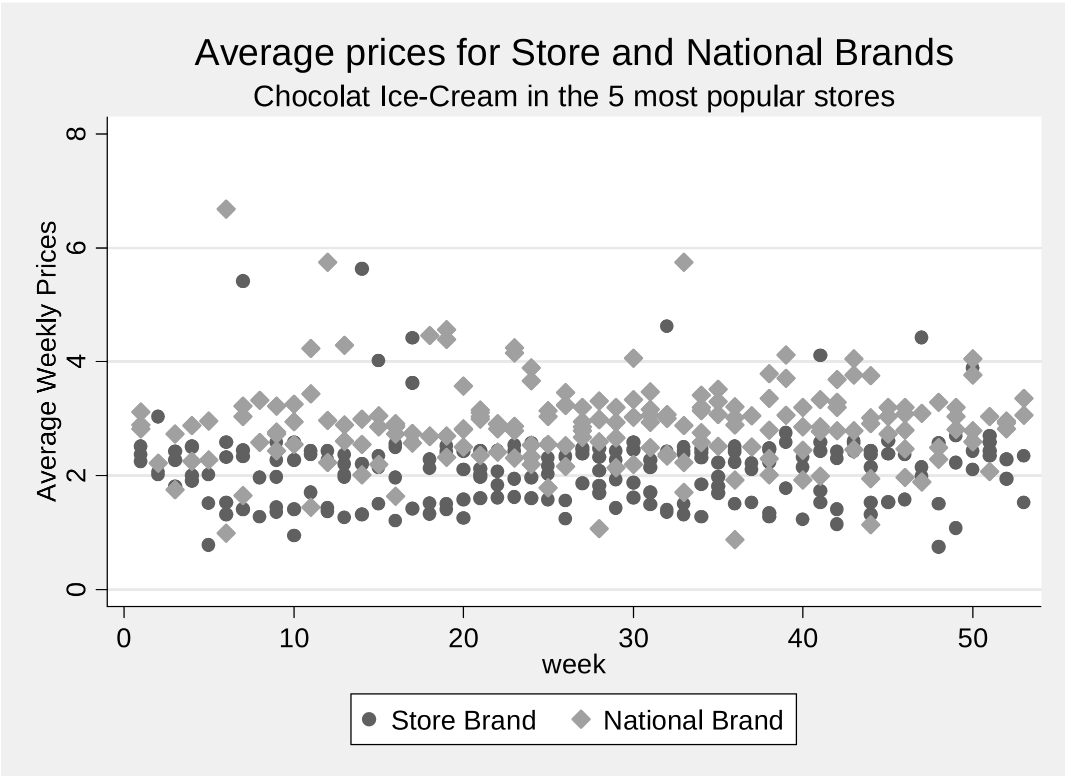


Figure 2.4: Average Prices for Store and National Brands - Chocolat Ice-Cream in the 5 most popular stores

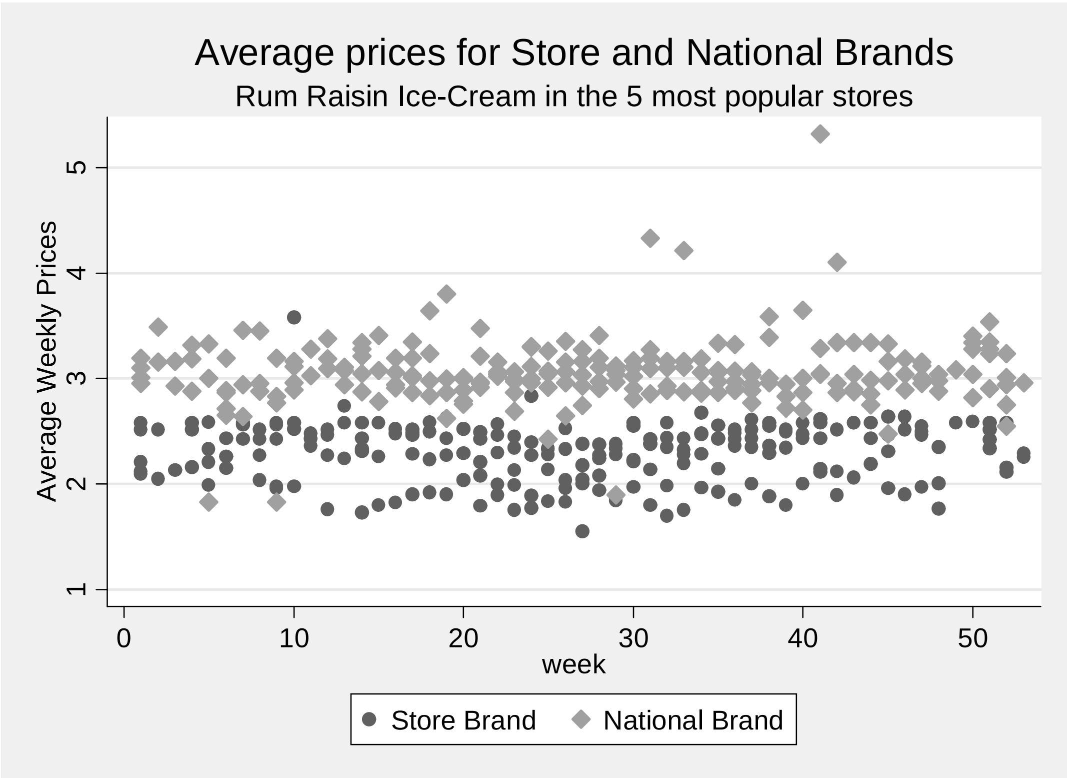


Figure 2.5: Average Prices of Store and National Brands - Rum Raisin Ice-Cream in the 5 most popular stores

Chapter 3

Price Dispersion and Search Costs: The Roles of Imperfect Information and Product Differentiation

3.1 Introduction

Since the seminal work of Stigler (1961), Economic theorists have developed models trying to explain price dispersion. Examples are Salop and Stiglitz (1982), Burdett and Judd (1983), Stahl (1989), among many others. The theory shows that when stores and consumers are identical and perfectly informed (or, equivalently, when search costs are null) and there is no capacity constraint, the unique Nash equilibrium is the Bertrand outcome, where prices are perfectly competitive. For price dispersion to arise there must be some heterogeneity between consumers or stores and positive search costs.

Although it is hard to contest the empirical observation that prices for otherwise homogeneous products differ across stores and over time, this does not necessarily mean that the Bertrand Equilibrium is not verified empirically. Price dispersion may be "illusory", a result of "hidden" price differentiation, in the words of Baylis and Perloff (2002). That is, the same product sold at different locations may actually be a differentiated product because store heterogeneity ends up reflected on the product itself.

It is therefore important to identify empirically the importance of the price differences that remain after controlling for observed and unobserved store heterogeneity, as well as market heterogeneity. Drawing a map of price dispersion for food products in France is the first goal of this paper.

The second goal is to identify the source of the price dispersion that cannot be accounted for by store and market characteristics. In particular, we test if there is evidence that search costs are driving price differentials. We also study the effect of the opportunity cost of time on prices paid by consumers.

Finally, we estimate the distribution of consumers' search costs. Recovering the distribution of search costs is important because the existence of search costs affects competition policy issues. For instance, in the presence of search costs, firm entry does not necessarily improve welfare. Stahl (1989) shows that an increase in the number of firms in the market may actually decrease welfare depending on how search costs are distributed among the population of consumers. Also, if search costs are important, firms may retain considerable market power even in seemingly competitive situations, which in some markets may justify price regulations See for example, Giulietti, Price and Waterson (2005), who study the UK energy market.

We consider a number of identical food products sold at different stores in France. Following Lach (2002), for each product considered, we regress the log of prices (expressed in differences from the period's average so that all difference is cross-sectional), pulled over time and store, on a chain-store fixed effect, a market fixed effect, a time-period effect, and a store size effect. The residual from such a regression can be considered as the price of a homogeneous good purged from store and market heterogeneity. It can therefore be interpreted as a measure of the distance (or deviation) from the Bertrand outcome.

We investigate the importance of search costs on explaining price dispersion in two ways. First, we study the effect of a decrease in search costs on price differentials. As suggested by the literature (see Warner and Barsky, 1995), we assume periods of high aggregate demand are periods of lower search costs per product since the fixed component of the search costs is divided by a longer list of items to be purchased. We thus regress alternative measures of price dispersion on seasonal dummies and dummies indicating periods of exogenous peak in demand, such as Christmas and weekends.

Second, we test the effect of consumers' opportunity cost of time on the prices they pay. If search costs are relevant, consumers with a high cost of time have a less intense searching activity and thus pay higher prices on average. The opportunity cost of time is capture by income, number of children, age, and whether the consumer has a professional activity.

Finally, we present a model of consumer choice with sequential search costs and develop an empirical strategy to identify the parameters of the search costs distribution, which we estimate along with the parameters of the utility function. Products are considered to be heterogeneous with both vertical and horizontal attributes. Hence, consumers search for the product with the highest indirect utility instead of the lowest price.

As far as we are concerned, this is the first paper in the literature to identify search costs in a context of horizontally differentiated products. The horizontal dimension is particularly important when dealing with physical (not online) stores. In this context, the relative geographical location of the store, which is consumer specific, is clearly an important characteristic affecting choices, and ignoring this differentiation dimension will bias estimated parameters. Also, unlike previous methodologies, our's does not require any assumption on how firms set prices. This means that once we recover the demand parameter estimates, we can test between alternative models of firm behavior. In particular, we could test whether equilibrium prices are a result of pure or mixed strategy Nash equilibrium.

The empirical investigation is performed on a comprehensive individual level dataset which includes every food product purchased by a representative survey of french households during 3 years, 1999, 2000, and 2001. We have information on product and store characteristics, as well as household demographics. This dataset is complemented by information on store location from INSEE (the french National Institute of Statistics and Economic Studies).

Reduced-form tests show that price dispersion is important in the french food market, even after controlling for unobserved store attributes. The price dispersion is also persistent over time. Stores frequently change positions in the cross-sectional distribution of prices, which is evidence in favor of firms playing mixed strategies. Moreover, there is evidence of a negative correlation between average price of the product and price dispersion, which is consistent with the idea that consumers have more incentives to search for high valued items since, due to a fixed cost component, search costs are relatively (to the high price of the product) lower in this case.

Periods of aggregate demand peaks, where search costs are expected to decrease, are also periods of lower price dispersion. This result indicates that search costs are a major cause of price dispersion. Furthermore, we find that prices paid increase with the opportunity cost of time, indicating that search costs are an important component of consumer behavior and that consumers who are time constrained search less intensively and end paying higher prices for identical products.

Finally, results from the structural estimation show that consumers obtain at most three utility quotes before purchasing the product. The vast majority of consumers (more than 90%) do not search at all, purchasing the first product drawn.

The paper is organized as follows. In Section 2, we review the literature on price dispersion and search costs. We start by discussing reduced-form studies and we focus on traditional stores, leaving out search in the internet. We then review the few papers that structurally identify search costs. The third section describes the data and the product choice. Section 4 presents the reduced-form tests, whereas Section 5 describes consumer choice behavior with sequential search and describes the empirical identification strategy. Results of the structural empirical analysis are in Section 6. Finally, the last section concludes and discusses extensions.

3.2 Literature Review

3.2.1 Reduced-Form Studies on Price Dispersion and Search Costs

In this section, we discuss empirical studies of price dispersion, concentrating on those which use data from traditional stores (as opposed to virtual or online stores). See Baye, Morgan, and Scholten (2006) for a review of theoretical models on price dispersion and search, as well as of empirical studies of price dispersion online.

To empirically identify price dispersion in supermarkets in France we follow closely the exercise proposed by Lach (2002), who looks at store-level data on monthly prices of four homogeneous products in Israel. Lach studies the existence and characteristics of the price dispersion across stores, as well as its persistence over time. Main results show that price dispersion across stores is prevalent. Furthermore, it differs across products, with dispersion decreasing with the price of the good. Also, price dispersion is shown to prevail even after controlling for observed and unobserved product heterogeneity, where the heterogeneity is related to the different stores and periods of purchase since all other product attributes are identical. To clear prices from heterogeneity, Lach regresses prices on fixed effects for chain store, month, city (where the store is located), and type of store, and claims that the residuals thus obtained are the prices of a homogeneous good. The variance of the residuals shows that price dispersion cannot be explained solely by product differentiation. Finally, he finds price dispersion to be persistent since the ranking of stores in the price distribution fluctuates over time, which means that consumers cannot learn about which stores have consistently lower prices. This result supports the idea that firms play mixed strategies in prices.

Sorensen (2000) focus on the empirical importance of price dispersion due to costly search using data on retail prices for prescription drugs in two markets. He finds that prices vary considerably across pharmacies in the same market, and that differences in pharmacy service or location do not appear to explain fully the price variation (pharmacies' price ranking are inconsistent across drugs and hedonic price equations on pharmacy characteristics are not totally successful in explaining the price variation). Actually, he finds that pharmacy effects account for at most one third of price differences. The central result of the paper is that the price dispersion (and markups) is significantly lower for drugs which are frequently purchased. This is consistent with models based on consumer search, which predict that consumers' incentives to price-shop are greater for frequently purchased prescriptions.

Zhao (2006) studies pricing patterns in six supermarkets in a Chicago suburb and studies its consistency with existing price dispersion theory based on costly consumer search, competition, and consumer heterogeneity. Price dispersion is empirically defined as the coefficient of variation of prices. Consumer search costs, that are not directly observed, are proxied by frequency of purchases[1]. Competition is empirically defined as an increase in the number of stores in the market, whereas consumer heterogeneity is measured as the coefficient of variation of some consumer demographics. He focuses on price dispersion for a certain UPC (universal product code) across stores in a certain week, within a product category in a store across UPCs in a certain week, and over time for a certain brand in a store. The main results are that the observed price dispersion is positively correlated with search costs, competition, and consumer heterogeneity.

The results above are coherent with those found by Lewis (2008), who measures price dispersion among differentiated retail gasoline sellers. The paper focuses on how the local competitive environment (represented by the number of nearby competitors) affects price dispersion. Results show that significant dispersion remains after controlling for station characteristics (price is regressed on station fixed effects and the residuals are interpreted as the price of the product once store heterogeneity is controlled for). As in Lach (2002), there is evidence that stations play mixed strategies, with stations changing position in the price ranking very frequently. Finally, Lewis finds that the relationship between price dispersion and seller density varies across different types of stations. For discount brands and independent (unbranded) stations, the relationship is negative and quite strong, whereas it is insignificant (and in some cases weakly

[1] The problem with the idea of proxying search costs by the frequency of purchase in the case of grocery products is that the frequency of purchase may also be capturing a measure of consumers' taste for the store, category, or brand. Notice that this problem does not arise when the product considered is a prescription drugs, as in Sorensen (2004).

positive) for high-brands premium branded stations.

The link between competition and price dispersion is studied also by Giulietti, Otero and Waterson (2007). The authors look at dispersion in electricity tariffs in the UK, where since May 1999 all consumers may choose their electricity supplier, and since March 2002 there exist no price regulation. They focus on the evolution of switching and search cost. Their identification hypothesis is that changes in electricity supplier involve both switching and searching costs. To disentangle the two effects, price divergences between incumbent and others is assumed to reflect both switching and searching costs, whereas price divergences between non-incubents are assumed to reflect search cost influences. Price dispersion is measured in two alternative ways: as the difference between the median non-incumbent price (the price that would be achieved by a single price inquiry) and the lowest price (which would be revealed by a complete search costs), and as the range of prices. Since the complete list of tariffs is available in the internet, the paper also explores the effect of potentially better informed consumers (or the effect of lower search costs) on the price variance. Empirical results show that search costs were decreasing under price regulation. However, from March 2002 on, there is evidence that search costs start to increase, probably due to increased product differentiation. There is also indication of a positive relationship between search costs and the number of firms in the market.

Delgado and Waterson (2003) find evidence of substantial price dispersion across outlets in the retail market for car tyres. Most interestingly, they find empirical evidence on the importance of considering vertical linkages in order to understand the pattern of prices across retailers. Indeed, their analysis show that manufactured-owned outlets sell rivals' tyre brand nearly 20% more expensively than they sell their own tyres, ceteris paribus.

The focus of Aguiar and Hurst (2007) is on the correlation between consumers characteristics and prices. They show how consumers with high opportunity costs of time pay higher prices on average. This is consistent with a model of search costs where more time constrained individuals have higher search costs, thus searching less intensively for the lowest price and paying on average higher prices than individuals with lower opportunity costs. The cost of time is captured by observable variables such as income, household size, and age of household head. They find that for identical goods, prices paid increase with income and household size. They also find evidence that prices paid increase with the age of the household head, reaching a peak around the forties, when it starts to decrease again. Finally, they find that households that purchase more frequently pay lower prices

In the reduced form part of this paper, we repeat the exercise in Lach (2002) considering 5 different product categories and two tightly defined products within each product category. We find important and persistent price dispersion, even once we control for store differences. Inspired by Aguiar and Hurst, we also perform reduced form tests of the effect of consumers search costs (proxied by variables capturing the opportunity cost of time) on prices paid. Results indicate that time constrained individuals tend to pay higher prices. Finally, we analyze the effect of an exogenous decrease in search costs on observed price dispersion. More specifically, we look at the effect of exogenous aggregate peaks, which are assumed to be periods of lower search costs, on observed price differentials. We find evidence that periods of high exogenous aggregate demand, like Christmas, are periods of lower price dispersion. This result is interpreted as corroborating the idea that price dispersion is at least partly driven by informational issues.

3.2.2 Structural Identification of Search Costs

There are two main strands of the literature on the empirical identification of search costs. One based both on price and quantity data and one which exploits the equilibrium supply-demand restrictions of the theoretical models to estimate search costs distributions using price data alone. The latter was inaugurated by Hong and Shum (2006) who generalize Burdett and Judd (1983)'s consumer search model by adding consumer search cost heterogeneity. They consider equilibrium models of sequential and non sequential search (fixed sample size) where there is a continuum of firms and consumers and the equilibrium price distribution is interpreted as the symmetric equilibrium in mixed strategy employed by firms. Goods are homogeneous so that only search frictions (arising from consumers' imperfect information about store prices) and heterogeneity in search costs in the consumer population generate price dispersion. Demand is inelastic for a single unit of the good. They assume each price observed is real in the sense that it generates positive demand. The identical firms play mixed strategies in price which implies that the characterization of the equilibrium price distribution starts with the mixed strategy condition that firms be indifferent between charging the monopoly price (and selling only to people who never search but receive an initial draw equal to the monopoly price), and any other price in the equilibrium price support.

Hong and Shum (2006) methods are illustrated using observed price data on four Economics textbooks sold online. Results for the non sequential search model show that more than half of the consumers never search (they shop at the store where they received their initial quote, assumed to be costless). The high proportion of people who don't search implies that they cannot identify the shape of the distribution for these people. The sequential search model predicts higher magnitudes for the search costs (for some books, more than ten times higher) and lower marginal costs. The authors argue that the more sensible estimates implied by the non-sequential search model do not necessarily imply that the non-sequential assumption better describes consumer search behavior in this market. The extreme magnitudes of the sequential search model may be in part related to the stronger parametric assumptions required for identification of distributions in this environment.

Building on Hong and Shum's methodology, Moraga-Gonzales and Wildenbeest (2006) propose an alternative maximum likelihood estimation method on which the asymptotic theory for computing search costs cutoffs and conducting test of hypothesis remains standard. They apply their method to a data set on prices for four personal computer memory chips, obtained from a web-based search machine. They find evidence that consumers either search too much or too little (between 4% and 13% of consumers in the market search for all prices, but the majority of consumers search for at most three prices). The high search costs imply high market shares that are estimated to be around 25%.

It is worth noticing that the semiparametric identification of search costs can be problematic. Since prices reflect the behavior of a group of consumers, not individuals, the search cost distribution can only be identified at critical points determined by consumers' optimal search. So if there are N firms in a market, there are only N points of the search cost distribution that can be nonparametrically identified. This problem is studied by Moraga-Gonzales, Sandor, and Wildenbeest (2008), who work with an oligopolistic version (finite instead of infinite number of firms) of the Burdett and Judd (1983) model. They show that considering a large number of firms in the market is not sufficient for identification of the search costs distribution in its full support. A better solution to overcome the identification problem would be to work with price data from several oligopolistic markets that share a common search cost distribution but whose consumers value products differently. They show an example where they include a period specific component to the consumers' utility and look at the same market observed at different periods. The authors also propose a method to estimate the search cost density function by a flexible polynomial type function. They argue that compared to existing methods, this is an easier way of estimating parameters in a framework where price data needs to be pooled from multiple markets.

Still in the literature of identification of non-sequential search costs using price data alone, there is Wildenbeest (2009), who allows for vertically differentiated products. He uses a data set for grocery items from supermarkets in the UK. Results indicate that most observed price variation can be explained by supermarkets heterogeneity and that the amount of search is low in this market. Furthermore, there is evidence that ignoring vertical product differentiation leads to overestimation of search costs.

Our paper is closest to Hortaçsu and Syverson (2004), who study price dispersion and search costs in the mutual fund industry. In their model, consumers are identical, except on search costs, which means that goods are vertically differentiated. Consumers search with replacement and purchase at most one unit of the good in a given period. Also, they know the distribution of realized utilities before any search is actually conducted. There are N firms offering N (vertically differentiated) products. Observed prices and quantities are the outcome of pure strategy Nash equilibrium between firms. In this framework, they are able to estimate the search cost distribution non parametrically. However, to study policy implications, specifically the effect of entry on welfare, they assume search costs are lognormally distributed. Empirical results indicate that price dispersion cannot be explained by product differentiation alone nor search costs alone, but a combination of product heterogeneity and information frictions. In terms of welfare implications, the authors find that total costs sunk into search process are quite relevant. These costs could be avoided by restricting entry into the sector to a single monopolist. The obvious trade-off is the welfare losses associated with the reduction of product variety and the deadweight loss from increased market concentration. However, rough calculations suggest that the net social effect of imposing entry restriction is positive.

The sequential search cost model presented in this paper is an extension of Hortaçsu and Syverson (2004) in the sense that we allow for horizontal differentiation as well as vertical differentiation between products. This means that consumers differ not only on their search costs but also on their valuation of product attributes. As discussed below, however, we only consider observed consumer heterogeneity. Furthermore, we do not need to make assumptions on how firms behave in order to identify the search cost distribution. Hence, although this is out of the scope of this study and we leave it for future work, in principle our method enables us to use the estimated demand parameters to test among different supply models as in Nevo (2001), Bonnet and Dubois (2006), Berto Villas Boas (2007), among others. Specifically, it would be interesting to test whether price distributions are a results of firms playing pure or mixed strategies.

3.3 Data and Product Choice

To study price dispersion and to identify search costs, we use a comprehensive database which is a representative survey of households distributed across all regions of France. We have information on three years: 1999, 2000, and 2001. Each household was given a scanner with which it should register every food product purchased. For each product purchased, we have information on its brand and characteristics, including price and pack size, label, the date of the purchase and the brand and surface of the retailer where it was purchased. We also have comprehensive information on household demographics.

Matching these data with geographical information on retail outlets locations, we are also able to compute distances from home to supermarkets, which are important for the structural identification of search costs. The data on location is taken from the INSEE database.

We consider 10 products within 5 product categories. The categories are beer, cola, milk, coffee, and whisky. From each category, we chose two products so we could compare intra category price dispersion. The choice of product within a category was done bearing in mind that we wanted to consider brands of high market share to be sure to have a big number of stores selling them. Hence, among each category, we chose the product most frequently purchased. The second product chosen in a category was a product that, among those most frequently purchased, had an average price clearly higher or lower than the first product chosen. In terms of the choice of categories, we considered relatively cheap and relatively expensive categories, e.g., milk and whisky, to check if price dispersion decreases with average price. Finally, we also wanted to compare categories that are "necessities" and frequently consumed (e.g., coffee, milk) against "luxury" items (e.g., cola, beer, whisky).

We define products very tightly, allowing for only one source of differentiation which is the store where they were purchased. So, for example, within the Cola category, a product is defined by its brand, whether it comes in a bottle or in a can, the pack size, whether it is light cola, and the size of the bottle or can. A full description of each of the products considered can be found in the Appendix.

Table 3.1 shows some descriptive statistics for the products studied. The first column brings the number of observations in the data set, which corresponds to the number of times the product was purchased by one of the households during the 3 year data span. In the second column, we have the average quantity purchased in one purchase occasion, measured in liters for liquid products and kilograms in the case of coffee. This average quantity can vary a lot within one product category and across categories because different products come in different pack sizes.

So, for instance, beer A comes in packs of 24 bottles of 250 ml, whereas beer B comes in packs of 10 bottles of the same 250 ml[2]. The last column gives the total quantity purchased during the whole time span, per product. Again, units of measurement are liters for liquids and kilograms for coffee.

Table 3.1: Descriptive Statistics of Quantities Purchased per Product

Category	Product	Observations	Avg. Quantity per Purchase	Total Quantity
Beer	A	4201.00	7.00	29412.00
	B	6891.00	2.69	18538.00
Coffee	A	8270.00	0.61	5038.50
	B	5736.00	0.44	2508.25
Cola	A	35338.00	2.46	87087.00
	B	1708.00	9.14	15606.00
Milk	A	4308.00	2.42	10426.00
	B	43185.00	5.99	258754.00
Whisky	A	1972.00	0.72	1428.70
	B	1023.00	0.71	730.10

Note: quantities are in liters for liquids and in kilograms for coffee.

Descriptive statistics for the variables used to capture the opportunity cost of time are displayed in Table 3.2. They are age of the household head (measured in years), a dummy variable indicating the presence of a baby (a child of less than 48 months of age) in the household, the number of children of 16 or less years of age, the education level of the household head, the household size measured as the number of household members, the socioeconomic class, and a dummy indicating whether the household head is professionally inactive. The education level variable is organized in three levels: no diploma, high school, undergraduate, and graduate diploma. The socioeconomic class variable is constructed as a function of the number of people in the household and INSEE unities of consumption. The variable indicating whether the household head is inactive is equal to one if the household head is either a student, retired, long term unemployed, or has no professional activity.

[2] See the appendix for details on pack sizes of each product.

Table 3.2: Descriptive Statistics of Consumer Characteristics

Variable	Mean	Std. Dev.	Min.	Max.	N
Age person of household head	47.86	14.04	19	94	5963
Baby in the household	0.13	0.34	0	1	5963
Nb of kids younger than 16	0.71	1.02	0	4	5963
Educational level = 1	0.24	0.43	0	1	5963
Educational level = 2	0.12	0.32	0	1	5963
Educational level = 3	0.25	0.43	0	1	5963
Educational level = 4	0.39	0.49	0	1	5963
Upper Class	0.12	0.32	0	1	5963
Upper Middle Class	0.27	0.44	0	1	5963
Lower Middle Class	0.45	0.5	0	1	5963
Lower Class	0.16	0.37	0	1	5963
Household Size	3.05	1.41	1	9	5963
Household head is inactive	0.41	0.49	0	1	5963

3.4 A Map of Price Dispersion in French Supermarkets

3.4.1 Uncontrolled Price Dispersion

Price dispersion has been frequently observed in a number of markets. The French groceries market is no different. Table 3.3 brings some descriptive statistics on the price of the products considered in this study. Those statistics are the average price per liter in the case of liquids or per kilo in the case of coffee, the standard deviation, the coefficient of variation, and the ratio of the third to the first quartile, as well as the ratio of the 95% to the 5% quantile. Notice that, here, there is no control for store heterogeneity. Hence, even if we are dealing with tightly defined products, they may still be differentiated as discussed earlier because they embed potentially differentiated characteristics of the store where they were purchased.

The statistics provided in Table 3.3 show that the price dispersion is important in all categories considered. Indeed, looking at the interquartile difference, we see that 50% of prices in the middle of the distribution differ up to 30%. This difference is less important for Whisky (1%), which is the most expensive product under study. This is consistent with the idea that search costs have a fixed cost component. In the case of expensive products, the search cost is low relatively to the price of the good and consumers have more incentives to search more intensively for the best price. Since more search is undertaken, consumers are better informed about practised prices, forcing stores towards the Bertrand Equilibrium, in which price dispersion is minimal.

Table 3.3: Descriptive Statistics of Pricing Patterns

Category	Product	Avg Price	Std Dev	Coef of Variation	Q75/Q25	Q95/Q5
Beer	A	0.98	0.10	0.10	1.21	1.27
	B	1.30	0.10	0.07	1.04	1.26
Colas	A	0.705	0.048	0.07	1.08	1.18
	B	0.494	0.114	0.23	1.28	1.81
Coffee	A	8.73	0.64	0.07	1.10	1.22
	B	4.49	0.768	0.17	1.34	1.70
Milk	A	0.495	0.033	0.07	1.05	1.16
	B	0.815	0.136	0.17	1.24	1.63
Whisky Bourbon	A	15.5	1.65	0.11	1.01	1.44
	B	14.9	0.63	0.04	1.02	1.05

3.4.2 Price Dispersion of Homogeneous Products

Part of the price dispersion observed in prices may be explained by product differentiation and time variation. Although we compare goods with exactly the same physical attributes, they are sold at different stores and different time periods, which means that the products cannot be considered to be homogenous. To clear prices from the heterogeneity due to the location of purchase and period, we run product by product regressions of prices, measured as log deviations with respect to the weekly mean, on month and year fixed effects that capture observed and unobserved effects of the time-period, on store chain identity dummies, on store type and regional dummies. The residuals of these regressions are considered to represent the price of a homogeneous product. This method which allows to obtain measures of the price of the common attributes of the good (all attributes excluding store and period) has become standard in the literature (see Lach, 2002, Zhao, 2006, and Sorensen, 2000, among others). However it is important to notice that by arguing that the residuals of the fixed effects regression described above can be interpreted as the price of the homogenous product, we are implicitly assuming that the final retailer price is a linear combination of the prices of individual attributes (specifically, the sum of the price of the homogenous product and the price of the differentiated services offered by the retailer).

Table 3.4 shows descriptive statistics for the dispersion of the residuals of the fixed effects regressions for each product considered. Those descriptive statistics are weekly averages of the standard deviation of the residuals, the first and third quartiles, as well as differences between the first and third quartile, and the 95% and 5% quantile. We do not report the mean value of the residuals because they are by construction equal to zero.

Comparing Table 3.3 and 3.4, we notice that the dispersion in prices as measured by the dispersion coefficient drops significantly once we control for observed and unobserved product heterogeneity. In terms of interquartile differences, the difference in prices in the middle of the price distribution decrease by more than 50% for half of the products. However, we are still far away from the law of unique price predicted by the standard Bertrand Equilibrium.

Table 3.4: Dispersion of Prices cleared from Product Heterogeneity

Category	Product	Q25	Q75	Q75-Q25	Q95-Q5
Beer	A	-0.06	0.06	0.12	0.23
	B	-0.01	0.02	0.03	0.14
Cola	A	-0.01	0.01	0.02	0.06
	B	-0.10	0.11	0.21	0.48
Coffee	A	-0.03	0.02	0.06	0.16
	B	-0.17	0.12	0.29	0.45
Milk	A	-0.08	0.08	0.17	0.46
	B	-0.01	0.01	0.01	0.06
Whisky	A	-0.04	0.03	0.06	0.21
	B	-0.01	0.01	0.02	0.06

3.4.3 Stores Position in the Price Ranking

It is important to establish whether price dispersion is persistent over time. If this is the case, then stores cannot be consistently selling at a higher price or a lower price than the others[3].

[3] We are here refering to the price cleared of store heterogeneity, of course. If a store offers higher quality than others, then setting consistent higher (total) prices is not inconsistent with persistence in price dispersion.

Otherwise, consumers would learn which are the cheapest stores and nobody would purchase at the expensive stores. This means that (if price dispersion is persistent) consumers must be unable to perfectly predict which stores offer the lowest prices. A way of checking this is to study the position of stores in the cross sectional price distribution, and check whether stores frequently change position in this ranking over time or rather tend to remain in the same position. Tables 3.5 to 3.9 show the average (across periods and stores) of the probabilities of changing positions in the price ranking. At each week, we assign stores to one of the three price intervals limited by the quartiles of the price distribution[4]. The transition probabilities in Tables 3.5 to 3.9 are simply the empirical probabilities of changing from position j to position k, with $j = 1, 2, 3, 4$ and $k = 1, 2, 3, 4$.

The probability that a store remains in the same position in the price ranking is highest for milk, and is around 60-70%. For the other products, the probability of remaining in the same position ranges from 20 to 60%. In general, this probability is higher for stores in the top and bottom ranks. In any case, these probabilities are far from 1, indicating that stores are not consistently offering high or low prices (remember that here we are dealing with prices cleared from store heterogeneity). Thus, consumers cannot learn before searching where they should go for the best deal. The empirical evidence presented here is therefore consistent with persistent price dispersion and equilibrium prices (for the homogeneous product) resulting from firms playing mixed strategies.

Table 3.5: Probabilities of Transition between Positions in Store Price Rankings - Beer

		Position at $t+1$							
		Beer A				Beer B			
		1	2	3	4				
Position at t	1	0.329	0.188	0.126	0.140	0.428	0.184	0.088	0.079
	2	0.196	0.242	0.211	.136	0.202	0.393	0.169	0.061
	3	0.166	0.204	0.303	0.147	0.076	0.178	0.410	0.143
	4	0.157	0.148	0.173	0.342	0.096	0.053	0.153	0.464

[4] That is, if a store is in position 1 at t, it means that its price at t is lower or equal to the first quartile of the price distribution at that period. The store is in position 2 if its price is between the first and second quartiles, and in position 3 if its price is between the second and third quartiles. Finally, the store is in position 4 if its price is greater than the third quartile.

Table 3.6: Probabilities of Transition between Positions in Store Price Rankings - Coffee

		Position at $t+1$							
		Coffee A				Coffee B			
		1	2	3	4				
Position at t	1	0.328	0.169	0.130	0.139	0.508	0.218	0.112	0.106
	2	0.183	0.287	0.207	0.143	0.228	0.365	0.149	0.149
	3	0.138	0.203	0.290	0.187	0.143	0.227	0.426	0.163
	4	0.176	0.150	0.195	0.310	0.142	0.157	0.184	0.467

Table 3.7: Probabilities of Transition between Positions in Store Price Rankings - Cola

		Position at $t+1$							
		Cola A				Cola B			
		1	2	3	4				
Position at t	1	0.602	0.117	0.094	0.091	0.303	0.172	0.138	0.091
	2	0.160	0.565	0.105	0.080	0.169	0.174	0.190	0.133
	3	0.070	0.159	0.552	0.127	0.184	0.151	0.193	0.135
	4	0.080	0.065	0.160	0.601	0.137	0.137	0.167	0.209

Table 3.8: Probabilities of Transition between Positions in Store Price Rankings - Milk

		Position at $t+1$							
		Milk A				Milk B			
		1	2	3	4				
Position at t	1	0.660	0.106	0.077	0.080	0.378	0.155	0.097	0.090
	2	0.138	0.625	0.095	0.066	0.167	0.309	0.175	0.095
	3	0.067	0.140	0.623	0.093	0.080	0.178	0.355	0.175
	4	0.093	0.051	0.130	0.672	0.109	0.093	0.187	0.362

Table 3.9: Probabilities of Transition between Positions in Store Price Rankings - Whisky

		Position at $t+1$							
		Whisky A				Whisky B			
		1	2	3	4				
Position at t	1	0.286	0.144	.0848	0.083	0.279	0.105	0.058	0.048
	2	0.174	0.186	0.172	0.078	0.163	0.178	0.133	0.043
	3	0.117	0.169	0.197	0.138	0.102	0.110	0.185	0.114
	4	0.094	0.086	0.199	0.225	0.041	0.053	0.177	0.207

3.4.4 Search Costs and Price Dispersion: Reduced Form Tests

Having established that price dispersion is important in the french food market, we would like to identify the source of the price dispersion that cannot be accounted by store and market characteristics. In particular, we test if there is evidence that search costs are driving price differentials, i.e., we test the hypothesis that price dispersion is consistent with predictions of models based on consumer search. Our identification hypothesis is that periods of exogenous increase in aggregate demand are periods of relatively lower search costs. As Warner and Barsky (1995) argue, periods of high aggregate demand are also periods where consumers are willing to invest more on information and transportation to find the lowest price. In other words, since consumers will usually have a longer list of items to purchase during Christmas, for instance, they will be willing to pay the extra search cost for the best price. They are then better informed and more vigilant and both mark ups and price dispersion should go down. Hence, we test the search costs theory by regressing alternative measures of price dispersion on dummies indicating periods of high aggregate demand. The measures of price dispersion used are the coefficient of variation of prices and the interquartile difference ($Q75/Q25$). We also consider the interquartile difference of the residuals obtained in the last section ($Q75 - Q25$), that is, the interquartile differences of the prices cleared of unobservable heterogeneity due to store and period of purchase.

Results are in Table 3.10. We are mainly interested in the coefficients for Christmas (the three weeks preceding Christmas) and weekends. There is clear evidence that price dispersion drops significantly during Christmas. Indeed, the Christmas coefficient is negative and significant for the three alternative measures of price dispersion. This is not the case for weekends though. As a matter of fact, weekends seem to be periods of higher rather than lower price dispersion.

If search costs are important, households with a higher opportunity cost of time will search less and will, on average, pay higher prices for identical products. A positive correlation between the cost of time and prices paid is therefore evidence of the existence and importance of search

Table 3.10: Effect of a Decrease in Search Costs on Alternative Measures of Price Dispersion

	(1) Coef of Variation	(2) Q75/Q25	(3) Q75-Q25 of Residuals
Christmas	-0.22***	-0.05***	-0.45***
	(-7.70)	(-12.68)	(-12.36)
Weekend	0.05**	0.01***	0.10***
	(2.96)	(5.77)	(4.83)
Spring	0.11***	0.02***	0.09**
	(5.32)	(7.80)	(3.19)
Summer	-0.01	0.02***	-0.01
	(-0.28)	(6.61)	(-0.28)
Autumn	0.07**	0.04***	0.13***
	(3.29)	(12.81)	(4.44)
Year 2000	-0.04	-0.00	-0.15***
	(-1.92)	(-0.72)	(-6.27)
Year 2001	0.02	0.01***	-0.04
	(1.16)	(5.16)	(-1.63)
Constant	-2.59***	0.10***	-2.53***
	(-135.91)	(36.85)	(-103.24)
N	13070	13565	13417

t statistics in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$
Pooled for all products.

costs as part of the demand behavior of consumers. We investigate this possibility by regressing prices paid on household characteristics assumed to capture the opportunity cost of time. Those characteristics are age and age square, education level and professional activity status of the household head (the variable inactive is equal to 1 if the household head is a student, retired or unemployed), socioeconomic class, presence of a baby of less than 48 months, number of kids aged 16 or less, and household size. We also include controls for region of residence and frequency of purchases which should capture a learning effect and is therefore expected to be negatively correlated to prices paid.

As can be seen in Table 3.11, prices decrease significantly with the frequency of purchases. As in Aguiar and Hurst, there is evidence that prices are hump-shaped over the lifecycle, which can be seen by the positive and significant coefficient on age and negative and significant coefficient on age square. Having kids at home also affects prices paid. Specifically, having a baby of less than 4 years, increases prices paid by 40 cents of an euro. However, we do not observe a monotonic relationship between the number of kids under 16 and prices. Prices paid seem to go up only for 4 kids: having at least 4 kids under 16 increases prices paid by $1,20$ euros. Finally, prices are a positive function of income (as measured by the socioeconomic variables) and of the education level of the household head.

3.5 Consumer Behavior with Search Costs

The previous sessions bring evidence not only that the law of one price is not observed in the French food market, even when store unobservable characteristics are accounted for, but also that search costs seems to be an important driving force of the observed price differentials. We therefore consider a model of consumer choice with search behavior. We also develop an empirical strategy to identify the search cost distribution and other parameters of the model.

	price
Freq of Purchase	-0.03***
	(-35.36)
Age	0.05***
	(4.76)
Age2	-0.00*
	(-2.48)
Baby	0.37***
	(5.48)
Inactive	-0.11**
	(-2.78)
Nb kids aged $<16 = 1$	-0.41***
	(-7.34)
Nb kids aged $<16 = 2$	-0.12
	(-1.90)
Nb kids aged $<16 = 3$	-0.88***
	(-9.51)
Nb kids aged $<16 = 4$	1.20***
	(6.68)
Upper Middle Class	-0.67***
	(-11.25)
Middle Class	-1.36***
	(-22.10)
Lower Middle Class	-2.13***
	(-27.29)
Fam size ≤ 4	0.39***
	(8.17)
Fam size ≤ 6	0.06
	(0.47)
Educ 1	0.62***
	(9.13)
Educ 2	0.47***
	(9.13)
Educ 3	0.17***
	(3.95)
Regional Dummies	Yes
Constant	6.71***
	(21.68)
N	13565

t statistics in parentheses

Table 3.11: Effect of Opportunity Cost of Time on Prices Paid

3.5.1 Modelling the Consumer Choice Behavior

Our demand model with search costs is based on Hortaçsu and Syverson (2004)'s extension of the framework developed by Carlson and McAfee (1983). However, in their model consumers are identical except for search costs, whereas we allow for some observable heterogeneity in preferences. This means that in our model, product attributes have both a vertical and an horizontal differentiation component (consumers do not all agree on the value of each attribute, as in Hortaçsu and Syverson). The horizontal dimension will be related to some observable characteristics of consumers. We allow for a number I of consumer types in terms of their valuation of the product.

Consumers purchase at most one unit of the product. Before purchasing, consumers sequentially search for the product with the highest indirect utility. Search is costly and its cost is heterogeneously distributed across the population of consumers. The cost of the first quote is zero. This is a standard assumption in the literature and ensures that everyone willing to purchase the product will do so independently of their level of search costs. The indirect utility of a consumer of type i from purchasing product j at period t is denoted u_{ijt}. Notice that within each i-type, search costs may vary, though valuations may not.

We assume consumers search with replacement. Let $F(.)$ be the belief distribution of indirect utilities u_{ijt} of a type i consumer. Then, the optimal search rule for i with search cost c_i^t is to search once more if:

$$c_i^t \leqslant \int_{u_{it}^*}^{\bar{u}_{it}} \left(u_{ijt} - u_{it}^*\right) dF\left(u_{ijt}\right)$$

where $\bar{u}_{it}$ is the upper bound of the support of $F\left(.\right)$, and u_{it}^* is the indirect utility of the highest-utility product already found by i. The above condition means that the marginal cost of searching one more time is smaller than or equal to the expected gain of searching one more time.

We assume consumers know $F(.)$ which means that they know the support of the distribution of indirect utilities so that they can label the N_i available products in ascending order with respect to the indirect utility: $u_{i1t} < u_{i2t} < ... < u_{iN_it}$. For simplicity, we assume that there does not exist any two products (stores) that provide the same indirect utility. Notice that we index the number of available products by the consumer type i. This is to make clear that consumers do not necessarily have access to the same products. Since the stores we consider are traditional stores, as opposed to virtual stores, we only allow consumers to purchase from stores inside a circle of radius ρ of distance around their home, which we call consumer i's "catchment area"[5]. Notice also that we index the ranking of indirect utilities ($j = 1, ..., N_i$) by t, the period of the purchase, since the ranking mays change from one period to the other. For notation simplicity however, in what follows we drop the time subscript of the store ranking.

As all indirect utilities of stores are strictly different, we get that:

$$F(u) = P(u_{ijt} \leqslant u) = \sum_{k=1}^{N_i} \phi_{ik} I_{\{u_{ikt} \leqslant u\}}$$

where ϕ_{ik} is the probability that store k is visited by consumer i sampled (this probability belief is known by consumers and common to all consumers of type i).

This, yields the following cut-off points on the search cost distribution:

$$c_{ij}^{t} = \sum_{k=j}^{N_i} \phi_{ik} (u_{ikt} - u_{ijt}) \tag{3.1}$$

where c_{ij}^{t} is the search cost level that makes any consumer of type i indifferent between purchasing at store j and searching once more (i.e., it is the lowest possible search cost of any type i consumer who purchases product j). Given that j was already quoted, products with indirect utility lower than u_{ijt} do not enter the calculation of the expected gain of searching once more (right-hand side of the above equation) because we assume that consumers can revisit previously searched stores without cost. This assumption means that we consider that the consumer has no cost to visit back previously visited stores either because knowing where to find a product allows her to save the most part of the cost of search or because coming back home to consume the consumer has anyway to follow the initial path of store visits. Remark that although search

[5] For instance, we do not allow a product sold at a store in the North of France to enter the consideration set of a consumer who lives in the South of the country.

costs are assumed to be time invariant, cut-off points will depend on the period of purchase. Notice also that $c_{iN}^t = 0$ and that the expected gain of an additional search decreases with the index of the product, so that $0 = c_{iN}^t < c_{iN-1}^t < ... < c_{i1}^t$.

Then, a consumer will purchase the lowest indirect utility product if she can search only once and find the lowest utility product in her first and only search. At t, the proportion of consumers of type i who will search only once is $G\left(c_{i1}^t\right)$, where G is the probability distribution function of the search costs. The probability of sampling the product $j = 1$ is equal to ϕ_{i1}. Therefore, the demand for the lowest indirect utility product at period t, aggregated for type-i consumers is equal to:

$$q_{i1}^t = \phi_{i1}\left[1 - G\left(c_{i1}^t\right)\right] \tag{3.2}$$

Following the same kind of reasoning, we get (see proof in the Appendix of Hortaçsu and Syverson):

$$q_{i2}^t = \phi_{i2}\left[1 + \frac{\phi_{i1}G\left(c_{i1}^t\right)}{1-\phi_{i1}} - \frac{G\left(c_{i2}^t\right)}{1-\phi_{i1}}\right] \tag{3.3}$$

and for $j = 3, .., N_i$:

$$q_{ij}^t = \phi_{ij}\left[\sum_{k=1}^{j} \frac{G\left(c_{ik-1}^t\right) - G\left(c_{ik}^t\right)}{1 - \sum_{l=0}^{k-1}\phi_{il}}\right]$$

which can be re-written

$$q_{ij}^t = \phi_{ij}\left[1 + \sum_{k=1}^{j-1} \frac{\phi_{ik}G\left(c_{ik}^t\right)}{\left(1 - \sum_{l=0}^{k}\phi_{il}\right)\left(1 - \sum_{l=0}^{k-1}\phi_{il}\right)} - \frac{G\left(c_{ij}^t\right)}{1 - \sum_{l=0}^{j-1}\phi_{il}}\right] \tag{3.4}$$

where by convention $G\left(c_{i0}^t\right) = 1$ and $\phi_{i0} = 0$.

3.5.2 Identification of the Search Costs

Let's start by assuming that we know the probabilities of finding a store ϕ_{ij} (we will actually estimate the parameters of a parametric function which depends on the distance to the store and on the number of stores in the catchment area). If we observed the indirect utility, then it would be straightforward to calculate $G\left(c_{ij}^t\right)$, for $j = 1, .., N_i$, from equations (3.2) to (3.4) above, using observed purchases. The problem is that indirect utilities are unobservable to the econometrician. Thus, unlike the consumer, we are not able to rank utilities. Although we observe the demand for the product in each store, we do not know the position of the store in the utility ranking at period t (we observe q_{ij}^t for every j but we do not observe j). However, we notice that

$$\frac{q_{ij}^t}{\phi_{ij}} - \frac{q_{ij-1}^t}{\phi_{ij-1}} = \frac{G\left(c_{ij-1}^t\right) - G\left(c_{ij}^t\right)}{1 - \sum_{l=0}^{j-1} \phi_{il}} > 0$$

which means that

$$\frac{q_{i1}^t}{\phi_{i1}} < \frac{q_{i2}^t}{\phi_{i2}} < ... < \frac{q_{iN_i}^t}{\phi_{iN_i}}$$

Then, knowing quantities q_{ij}^t and the probabilities of finding a store ϕ_{ij}, we know the elements of the vector of ratios $\left\{\frac{q_{ij}^t}{\phi_{ij}}\right\}_{j=1,..,N_i}$ and thus can order them to identify the indirect utility ranking of stores.

Knowing the $\frac{q_{ij}^t}{\phi_{ij}}$ and ϕ_{ij} we can solve the following triangular system in the unknowns $G\left(c_{ij}^t\right)$:

$$\begin{cases} \frac{q_{i1}^t}{\phi_{i1}} = 1 - G\left(c_{i1}^t\right) \\ \frac{q_{i2}^t}{\phi_{i2}} = 1 + \frac{\phi_{i1}}{1-\phi_{i1}} G\left(c_{i1}^t\right) - \frac{1}{1-\phi_{i1}} G\left(c_{i2}^t\right) \\ \frac{q_{i3}^t}{\phi_{i3}} = 1 + \frac{\phi_{i1}}{1-\phi_{i1}} G\left(c_{i1}^t\right) + \frac{\phi_{i2}}{(1-\phi_{i1})(1-\phi_{i1}-\phi_{i2})} G\left(c_{i2}^t\right) - \frac{1}{(1-\phi_{i1}-\phi_{i2})} G\left(c_{i3}^t\right) \\ .. \\ \frac{q_{ij}^t}{\phi_{ij}} = 1 + \sum_{k=1}^{j-1} \frac{\phi_{ik} G\left(c_{ik}^t\right)}{\left(1-\sum_{l=0}^{k} \phi_{il}\right)\left(1-\sum_{l=0}^{k-1} \phi_{il}\right)} - \frac{G\left(c_{ij}^t\right)}{1-\sum_{l=0}^{j-1} \phi_{il}} \end{cases}$$

which gives:

$$\begin{cases} G\left(c_{i1}^t\right) = 1 - \frac{q_{i1}^t}{\phi_{i1}} \\ G\left(c_{i2}^t\right) = 1 - q_{i1}^t - (1-\phi_{i1}) \frac{q_{i2}^t}{\phi_{i2}} \\ ... \\ G\left(c_{ij}^t\right) = 1 - \sum_{k=1}^{j} \left(\frac{q_{ik}^t}{\phi_{ik}} - \frac{q_{ik-1}^t}{\phi_{ik-1}}\right)\left(1 - \sum_{l=0}^{k-1} \phi_{il}\right) \end{cases} \tag{3.5}$$

The above system enables identification of the height of the search cost distribution evaluated at the cutoffs points, that is $G\left(c_{ij}^t\right)$ for $j = 1, .., N_i$.

As for $j = 1, .., N_i$

$$c_{ij}^t = \sum_{k=j}^{N_i} \phi_{ik} \left(u_{ikt} - u_{ijt}\right) = G^{-1}\left(G\left(c_{ij}^t\right)\right) \tag{3.6}$$

if we know the cumulative distribution function G_{ij}^t, we obtain c_{ij}^t and the indirect utilities up to a constant. We normalize the lowest utility at each period to be equal to 0 so that we are left with N_{i-1} equations and N_{i-1} unknown values to be calculated and obtain:

$$u_{ik}^{t} = \frac{c_{i1}^{t}}{\sum_{j=2}^{N} \phi_{ij}} + \sum_{k'=2}^{k-1} \frac{\phi_{ik'} c_{ik'}^{t}}{\left(\sum_{j=k'+1}^{N} \phi_{ij}\right)\left(\sum_{j=k'}^{N} \phi_{ij}\right)} - \frac{c_{ik}^{t}}{\sum_{j=k}^{N} \phi_{ij}} \tag{3.7}$$

with $u_{i1}^{t} = 0$.

Now, let's assume that indirect utilities u_{ij}^{t} depend on joint characteristics of the consumer and store X_{ijt}, common parameters γ_t, price p_{jt} and a consumer-store specific random deviation to mean utility v_{ijt} such that:

$$\ln\left(u_{ij}^{t}\right) = X_{ijt}\beta + \gamma_t - \alpha p_{jt} + v_{ijt} \tag{3.8}$$

In practice, X_{ijt} are observable characteristics of the store (that may vary with the identity of the consumer), γ_t are time-period fixed effects, p_{jt} is the price paid for product j at period t as before and α and β are parameters. Consumers' valuation of product characteristics has both horizontal and vertical dimensions. The horizontal dimension is captured by the distance of the consumer's home to the store and the number of stores in the consumer's catchment area (defined as a circle of radius R around the consumer's home). This distance and catchment area will be the same for all consumers living in a certain community. Therefore, all consumers living in a certain community are identical, and the community defines the type i of the consumer. All other product attributes aside from number of competitors and geographical location (i.e., store and seasonal attributes) are vertical attributes in the sense that all consumers value those attributes in the same way.

Finally, to complete the discussion on identification, we still have to deal with two issues: the parametric assumptions and estimations of both the probability of drawing utility quote u_{ij}^{t} (ϕ_{ij}) and of the distribution of search costs.

We assume first that G belongs to a family of c.d.f. parameterized by θ such that $G(c, \theta)$ is known. We also assume that the probability of a consumer of type i finding store j has the following form:

$$\phi_{ij}(\eta) = \frac{\exp(\eta_1 N_i + \eta_2 d_{ij})}{\sum_k \exp(\eta_1 N_i + \eta_2 d_{ik})}$$

or

$$\phi_{ij}(\eta) = \frac{N_{ij}^{\eta}}{\sum_k N_{ik}^{\eta}}$$

where $\eta = (\eta_1, \eta_2)$ is a vector of parameters, N_i is the total number of stores in consumers i catchment area, N_{ij} is the number of stores j in consumers i catchment area, d_{ij} is the distance of i's home to the store j.

Then (3.5) gives

$$c_{ij}^t(\eta,\theta) = G^{-1}\left(1 - \sum_{k=1}^{j}\left(\frac{q_{ik}^t}{\phi_{ik}(\eta)} - \frac{q_{ik-1}^t}{\phi_{ik-1}(\eta)}\right)\left(1 - \sum_{l=0}^{k-1}\phi_{il}(\eta)\right), \theta\right)$$

Using (3.7),

$$u_{i2}^t(\eta,\theta) = \frac{c_{i1}^t(\eta,\theta) - c_{i2}^t(\eta,\theta)}{\sum_{j=2}^N \phi_{ij}(\eta)}$$

and for $k \geq 3$,

$$u_{ik}^t(\eta,\theta) = \frac{c_{i1}^t(\eta,\theta)}{\sum_{j=2}^N \phi_{ij}(\eta)} + \sum_{k'=2}^{k-1} \frac{\phi_{ik'}(\eta) c_{ik'}^t(\eta,\theta)}{\left(\sum_{j=k'+1}^N \phi_{ij}(\eta)\right)\left(\sum_{j=k'}^N \phi_{ij}(\eta)\right)} - \frac{c_{ik}^t(\eta,\theta)}{\sum_{j=k}^N \phi_{ij}(\eta)}$$

Remark that by construction

$$u_{ik}^t(\eta,\theta) - u_{ik-1}^t(\eta,\theta) = \frac{c_{ik-1}^t(\eta,\theta) - c_{ik}^t(\eta,\theta)}{\sum_{j=k}^N \phi_{ij}(\eta)} > 0$$

Then, we could find the parameters of interest using (3.8) and using that $E[v_{ijt}] = 0$. As some endogeneity problem may generate a correlation between v_{ijt} and the prices (for example), we can instead assume that we know some variables Z_{ijt} that are uncorrelated with v_{ijt}, such that (3.8) gives the following moment condition

$$E\left[\left(\ln u_{ij}^t(\eta,\theta) - X_{ijt}\beta - \gamma_t + \alpha p_{jt}\right) Z_{ijt}\right] = 0$$

In what concerns the distribution of the search costs, we assume three alternative parametric forms. First, we assume that search costs are uniformly distributed (in which case there are no parameters to be estimated). We also consider a lognormal distribution, with parameters μ_i and σ_i, and a Gamma distribution with parameters θ_i. Note that parameters are allowed to vary with consumer type.

3.6 Structural Model Estimation Results

The empirical strategy described in the above section was performed in a subset of the products we have considered so far. For each product category, we picked the most frequently purchased product to ensure that we have observations for the majority of periods and communities. Table 3.12 displays the average (across consumer types and periods) of G_1 and G_2, the heights of the distribution function evaluated at the search cost cutoffs c_1 and c_2. We do not show G_3, G_4 etc. because they are all estimated to be equal to zero, implying that consumers search at most twice. Results indicate that consumers search activity is not very intense. More than 90% of consumers do not search at all. For all products except cola, they search at most once for the highest indirect utility. G_1, the proportion of consumers who search at least once (that is, they purchase after obtaining the second utility quote) varies from 0% for beer to 12% for cola, respectively.

Table 3.13 brings the search cost cutoffs c_1 and c_2 obtained by inverting G_1 and G_2. Remark that comparing the coefficient of prices and the difference in cutoff estimates of the search cost distribution between one search and two searches (c_1 and c_2) gives some idea about the price variation needed to compensate the utility difference between consumers who are at the margin willing to incur one additional search. Expressing this in percentage of the retail price of the product in store 2 amounts to compute $\frac{\ln(c_1^t - c_2^t)}{\alpha p_{2t}}$. The last column of Table 3.13 displays the average per product of this percentage change, where αp_{2t} is calculated using the estimated value of α.

Finally, Table 3.14 shows the results of the estimation of utility parameters. The variables that enter the utility specification are price, store surface, distance from home to the store, time period fixed effects, store brand effects, and controls for region of residence. The instruments Z_{ijt} used are the total number of stores in the catchement area of the consumer, and the proportion of stores of the same brand in the catchment area. Those variables are related to the competitive environment and are therefore bound to be correlated with prices, but not with indirect utilities, what makes them valable instruments. The sign of the price coefficients are of course expected to be negative, as is the case for the five products considered. The fact that most of those coefficients are not significant may be a problem because it may be an indication that the instruments for price are weak and that the price coefficients are therefore not well identified. The best results are for cola, which is also the product most frequently purchased. We believe that the availability of purchase observations plays an important role on well identifying the search costs cutoffs and consequently the parameters of the utility function.

Table 3.12: Average (across periods, region, and consumer type) Height of the Log Normal Distribution Function at Search Cutoffs Points, per product

Product	G1	G2
Beer	0.00	0.00
	(0.04)	(0.01)
Coffee	0.02	0.00
	(0.11)	(0.01)
Cola	0.12	0.01
	(0.25)	(0.07)
Milk	0.01	0.00
	(0.05)	(0.00)
Whisky	0.06	0.01
	(0.15)	(0.06)

Standard deviation in between parentheses.

3.7 Conclusion and Extensions

Price dispersion is an important characteristic of the food french market. Our empirical results show that only a part of the observed price differentials can be explained by store heterogeneity. We find evidence that the differences that remain are due to incompletely informed consumers who need to engage in costly search in order to find the best available deal. As a result, consumers with a high opportunity cost of time search less and pay higher prices on average.

Table 3.13: Average (across periods, regions, and consumer type) Search Cost Cutoffs, per product

Product	$\mathbf{c_1}$	$\mathbf{c_2}$	$ln(c_1 - c_2)$	$\frac{\mathbf{ln(c_1-c_2)}}{\alpha p_{2y}}$
Bier	0.24	0.46	-4.63	2.24
	(0.15)	(0.29)	(1.31)	(0.67)
Coffee	1.08	0.69	-4.99	0.01
	(1.06)	(0.23)	(2.12)	(0.00)
Cola	1.44	1.08	-1.13	0.01
	(0.53)	(0.48)	(2.20)	(0.02)
Milk	18339.78	354.73	1.81	-0.94
	(45597.44)	(327.70)	(4.23)	(2.19)
Whisky	0.59	0.49	-2.97	0.05
	(0.07)	(0.10)	(1.65)	(0.03)

Standard Deviation in parentheses.

Table 3.14: Estimates of Utility Parameters

	(1) Whisky	(2) Bier	(3) Cola	(4) Milk	(5) Coffee
Price	-4.17	-1.60	-203.94**	-2.26***	-55.32
	(-1.27)	(-1.14)	(-3.09)	(-3.78)	(-0.73)
Distance	-0.03	-0.00	-0.01***	-0.00*	0.07
	(-0.98)	(-1.10)	(-3.47)	(-2.09)	(0.49)
Surface	0.00	-0.00	-0.01	-0.00*	-0.08
	(0.02)	(-0.68)	(-0.92)	(-2.39)	(-0.69)
Time Period FE			YES		
Store Brand FE			YES		
Region FE			YES		
Constant	62.42	2.03	91.40**	1.07*	539.33
	(1.27)	(1.13)	(3.05)	(2.23)	(0.73)
N	525	3378	24849	2384	5591

t statistics in parentheses, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

We develop an empirical strategy to estimate the magnitude and distribution of sequential search costs. The main contribution of our empirical methodology is that we allow for products to be horizontally differentiated. Moreover, we are able to identify the search cost distribution without having to make assumptions on the behavior of firms, which enables us to use the demand parameter estimates to test alternative supply side specifications.

Results of the structural estimation of model parameters show that search costs for the products considered (cola, beer, whisky, milk, and coffee) are quite high and that consumers do not search much. More than 90% of consumers do not search at all, purchasing the first product they encounter. Around 5% of the population search once, and the only product for which a positive proportion of the population searches more than once is cola.

An interesting extension to this work is to model firm behavior and use demand side parameters to test between alternative specifications. We will then be able to study a number of policy implications. In particular, we are interested on the effect of an increase in the number of firms on consumer welfare.

3.8 References

Aguiar, Mark and Erik Hurst (2007) “Lifecycle Prices and Production”, American Economic Review, 97 (5), 1533-59.

Baye, Michael R., John Morgan, and Patrick Scholten (2006) “Information, Search, and Price Dispersion”, in: Terrence Hendershot (Editor), Handbook on Economics and Information Systems, Elsevier.

Baylis, Kathy and Jeffrey M. Perloff (2002) “Price Dispersion on the Internet: Good Firms and Bad Firms”, Review of Industrial Organization, 21: 305-324.

Berto Villas-Boas, Sofia (2007). “Vertical Contracts between Manufacturers and Retailers: Inference with Limited Data”, mimeo, University of California, Berkeley.

Bonnet C. and P. Dubois (2006) "Inference on Vertical Contracts between Manufacturers and Retailers Allowing for Non Linear Pricing and Resale Price Maintenance", IDEI Working Paper, n°519

Burdett, Kenneth and Kenneth L. Judd (1983) “Equilibrium Price Dispersion”, Econometrica, 51: 955-969.

Carlson, John A. and R. Preston McAfee (1983) “Discrete Equilibrium Price Dispersion”, Journal of Political Economy, 91: 480-93.

Delgado, Juan and Michael Waterson (2003) “Tyre Price Dispersion Across Retail Outlets in the UK”, Journal of Industrial Economics, 51 (4), 491-509.

Giulietti, Monica, Jesus Otero and Michael Waterson (2007) "Pricing Behaviour under Competition in the UK Electricity Supply Industry", Warwick Economic Research Papers, no. 790.

Giulietti, Monica, Catherine W. Price and Michael Waterson (2005) "Consumer Choice and Competition Policy: A Study of UK Energy Markets", Economic Journal, 115: 949-968.

Gonzaga-Morales, José Luis, Zsolt Sandor, and Matthijs E. Wildenbeest (2008) “Nonparametric Estimation of the Costs of Non-Sequential Search”, unpublished manuscript.

Hong, Han and Matthew Shum (2006) “Using Price Distributions to Estimate Search Costs”, Rand Journal of Economics, 37 (2), 275-257.

Hortaçsu, Ali and Chad Syverson (2004) "Product differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index funds", Quarterly Journal of Economics, 119: 403-456.

Lach, Saul (2002) "Existence and Persistence of Price Dispersion: An Empirical Analysis", Review of Economics and Statistics, 84 (3): 433-444.

Lach Saul and José Luis Moraga-Gonzalez (2009) "Asymmetric Price Effects of Competition" CEPR Applied IO Conference, Mannheim

Lewis, Matthew (2008) "Price Dispersion and Competition with Differentiated Sellers", Journal of Industrial Organization, 56 (3), 654-678.

Nevo, Aviv (2001) "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, Vol. 69 (2), pp. 307-42.

Salop, Steven C. and Joseph E. Stiglitz (1982) "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents", American Economic Review, 72: 1121-1130.

Sorensen, Alan T. "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs", Journal of Political Economy, 108 (4): 833-850.

Stahl, Dale O., II. (1989) "Oligopolistic Pricing with Sequential Consumer Search", American Economic Review, 79: 700-712.

Stahl, Dale O. (1996) "Oligopolistic pricing with heterogeneous consumer search" International Journal of Industrial Organization, 14(2), 243–268

Stigler, George J. (1961) "The Economics of Information", Journal of Political Economy, 69:213-225.

Warner, Elizabeth J. and Robert B. Barsky. (1995) "The timing and Magnitude of Retail Store Markdowns: Evidence from Weekends and Holidays." Quarterly Journal of Economics, 110(2), pp. 321-352.

Wildenbeest, Matthijs (2009) "An Empirical Model of Search with Vertically Differentiated Products", manuscript.

Zhao, Ying (2006) "Price Dispersion in the Grocery Market", Journal of Business, 79 (3): 1175-1192.

3.9 Appendix

1. Product Description

In what follows, we succinctly describe the characteristics defining each of the products considered in this study, organized by product categories. We do not name brands, only A and B to signal they are different brands (across categories, A and B designate different brands) because of a confidentiality agreement with the provider of the data base.

1.1. Beer:

1.1.1. brand A, bottle size: 250 ml , pack: 24 bottles

1.1.2. brand B, bottle size: 250 ml , pack: 10 bottles

1.2. Coffee

1.2.2. Brand A, no "gamme", arabica, caffeinated, 1 package per pack, package size 250g.

1.2.3. Brand B, degustation, arabica, caffeinated, 1 package per pack, package size 250g.

1.3. Cola

1.3.1. Brand A, plastic bottle, non light, bottle size: 1500 ml, pack: 1 bottle.

1.3.2. Brand A, plastic bottle, non light, bottle size: 1500 ml, pack: 4 bottle.

1.4. Milk

1.4.1. Brand A, paper brick of 1 liter, non organic, semi-skimmed, 1 unit per pack.

1.4.2. Brand B, plastic bottle of 1 liter, non organic, semi-skimmed, 1 unit per pack.

1.5. Whisky

1.5.1. Brand A, 1 liter bottle, no age, blended.

1.5.2. Brand B, 1 liter bottle, 5 years of age, blended.

Bibliography

Aguiar, Mark and Erik Hurst (2007) "Lifecycle Prices and Production", American Economic Review, 97 (5), 1533-59.

Aguirregabiria, Victor (1999) "The Dynamics of Markups and Inventories in Retailing Firms', Review of Economic Studies, 66, 275-308.

Aguirregabiria, Victor (2005) "Retail Stockouts and Manufacture Brand Competition", mimeo, Boston University.

Ailawadi, Kusum and Scott A. Neslin (1998) "The Effect of Promotion on Consumption: Buying More and Consuming it Faster", Journal of Marketing Research, 35, 390-398.

Assunção, João, Meyer, Robert (1993) "The Rational Effect of Price Promotions on Sales and Consumption", Management Science, 39 (5), pp. 517-535.

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

Baylis, Kathy and Jeffrey M. Perloff (2002) "Price Dispersion on the Internet: Good Firms and Bad Firms", Review of Industrial Organization, 21: 305-324.

Bell, David R., and Christian A.L. Hilber (2005) "An Empirical Test of the Theory of Sales: Do Household Storage Constraints Affect Consumer and Store Behavior?", mimeo.

Benitaz-Silva, H. G. Hall, G. Hilsch, G. Pauletto and J. Rust (2000) "A Comparison of Discrete and Parametric Aproxximation Methods for Continuous-State Dynamic Programming Problems." Yale University, mimeo.

Berry, S., J. Levinsohn, A. Pakes (1995). "Automobile Prices in Market Equilibrium", Econometrica, 63(4), pp. 841-90.

Berto Villas-Boas, Sofia (2007). "Vertical Contracts between Manufacturers and Retailers: Inference with Limited Data", mimeo, University of California, Berkeley.

Bierlaire, Michel, Denis Bolduc, Daniel McFadden (2003). "Characteristics of Generalized Extreme Value Distributions", Working Paper.

Boizot, C., J.M. Robin, and M. Visser (2001). "The Demand for Food Products: An Analysis of Interpurchase Times and Purchased Quantities," Economic Journal, 111(470), pp. 391-419.

Bonnet, C.and P. Dubois (2006). "Non-linear Contracting and Endogenous Market Power between Manufacturers and Retailers: Identification and Estimation on Differentiated Products," mimeo, Université de Toulouse.

Braido, L.H.B. (2009) "Multiproduct Price Competition with Heterogeneous Consumers and Nonconvex Costs", Journal of Mathematical Economics (2007), forthcoming.

Burdett, Kenneth and Kenneth L. Judd (1983) "Equilibrium Price Dispersion", Econometrica, 51: 955-969.

Carlson, John A. and R. Preston McAfee (1983) "Discrete Equilibrium Price Dispersion", Journal of Political Economy, 91: 480-93.

Chevalier, Judith A., Anil K. Kashyap, and Peter E. Rossi (2003) "Why don't prices Rise During Periods of Peak Demand? Evidence from Scanner Data." American Economic Review, 93(1), pp. 345-70.

Delgado, Juan and Michael Waterson (2003) "Tyre Price Dispersion Across Retail Outlets in the UK", Journal of Industrial Economics, 51 (4), 491-509.

Erdem, T., S. Imai, and M. Keane (2003). "Brand and Quantity Choice Dynamics under Price Uncertainty", Quantitative Marketing and Economics, 1, pp. 5-64.

Feenstra, Robert C. and Matthew D. Shapiro (2001) "High-Frequency Substitution and the Measurement of Price Indexes." NBER Working Paper 8176.

Gabszewicz, J., and J. F. Thisse (1979) "Price Competition, Quality and Income Disparities", Journal of Economic Theory, 20: 340-359.

Gabszewicz, J., and J. F. Thisse (1980) "Entry (and Exit) in a Differentiated Industry", Journal of Economic Theory, 22: 327-338.

Giulietti, Monica, Catherine W. Price and Michael Waterson (2005) "Consumer Choice and Competition Policy: A Study of UK Energy Markets", Economic Journal, 115: 949-968.

Giulietti, Monica, Jesus Otero and Michael Waterson (2007) "Pricing Behaviour under Competition in the UK Electricity Supply Industry", Warwick Economic Research Papers, no. 790.

Gonzaga-Morales, José Luis, Zsolt Sandor, and Matthijs E. Wildenbeest (2008) "Nonparametric Estimation of the Costs of Non-Sequential Search", unpublished manuscript.

Hendel, N. and A. Nevo (2006a) "Sales and consumer Inventory." NBER Working Paper 9048.

Hendel, N. and A. Nevo (2006b) "Measuring the Implications of Sales and consumer Inventory Behavior", Econometrica, 74 (6), 1637-73.

Hong, Han and Matthew Shum (2006) "Using Price Distributions to Estimate Search Costs", Rand Journal of Economics, 37 (2), 275-257.

Hortaçsu, Ali and Chad Syverson (2004) "Product differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index funds", Quarterly Journal of Economics, 119: 403-456.

Keane, Michael P. (1993) "Simualtion Estimation for Panel Data Models with Limited Independent Variables." G.S. Maddala, C.R. Rao, H.D. Vinod, eds. Handbook of Statistics, North Holland, 545-572.

Keane, Michael P. and Kenneth I. Wolpin (1994) "Solution and Estimation of Dynamic Programming Models by Simulation", Review of Economics and Statistics, 76, 684-672.

Lach Saul and Jodé Luis Moraga Gonzalez (2009) "Asymmetric Price Effects of Competition" CEPR Applied IO Conference, Mannheim

Lach, Saul (2002) "Existence and Persistence of Price Dispersion: An Empirical Analysis", Review of Economics and Statistics, 84 (3): 433-444.

Lal, Ravij and Carmen Matutes (1994) "Retail Pricing and Advertising Strategies." Journal of Business, 67(3), pp. 345-70.

Lewis, Matthew (2008) "Price Dispersion and Competition with Differentiated Sellers", Journal of Industrial Organization, 56 (3), 654-678.

McFadden, D. (1984) "Econometric Analysis of Qualitative Response Models", in Z. Griliches and M. Intilligator, eds., Handbook of Econometrics, Vol. II, Amsterdam: North-Holland, 1396-1456.

McFadden, Daniel (1980) "Econometric Models for Probabilistic Choice Among Products", Journal of Business, 53(3), Part 2: Interfaces between Marketing and Economics, 13-29.

Mussa, Michael and Sherwin Rosen (1978) "Monopoly and Product Quality", Journal of Economic Theory, 18: 301-317.

Nevo, Aviv (2001) "Measuring Market Power in the Ready-to-Eat Cereal Industry." Econometrica, Vol. 69 (2), pp. 307-42.

Nevo, Aviv and Konstantinos Hatzitaskos (2006) "Why Does the Average Price of tuna Fall During Lent?." manuscript.

Pesendorfer, Martin (2002). "Retail Sales. A Study of Pricing Behavior in Supermarkets," Journal of Business , 75, 33-66.

Rotemberg, Julio J. and Garth Saloner (1986) "A Supergame-Theoretic Model of Price Wars during Booms." American Economic Review, 76(3), pp. 390-407.

Rust, John (1987) "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zucher." Econometrica, 55(5), pp. 999-1033.

Rust, John (1996) "Numerical Dynamic Programming in Economics", in H.M. Amman, D.A. Kendrick, and J. Rust, eds., Handbook of Computational Economics, Vol. I, Elsevier, 731-800.

Salop, Steven C. and Joseph E. Stiglitz (1982) "The Theory of Sales: A Simple Model of Equilibrium Price Dispersion with Identical Agents", American Economic Review, 72: 1121-1130.

Shaked, A. and J. Sutton (1982) "Relaxing Price Competition through Product Differentiation", Review of Economic Studies, 49: 39-13.

Shaked, A. and J. Sutton (1983) "Natural Oligopolies", Econometrica, 51: 1469-1484.

Sorensen, Alan T. "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs", Journal of Political Economy, 108 (4): 833-850.

Stahl, Dale O. (1996) "Oligopolistic pricing with heterogeneousconsumer search" International Journal of Industrial Organization, 14(2), 243–268

Stahl, Dale O., II. (1989) "Oligopolistic Pricing with Sequential Consumer Search", American Economic Review, 79: 700-712.

Stigler, George J. (1961) "The Economics of Information", Journal of Political Economy, 69:213-225.

Sun, Baohong (2005) "Promotion Effect on Endogenous Consumption." Marketing Science, 24 (3), pp. 420-43.

Warner, Elizabeth J. and Robert B. Barsky. (1995) "The timing and Magnitude of Retail Store Markdowns: Evidence from Weekends and Holidays." Quarterly Journal of Economics, 110(2), pp. 321-352.

Wildenbeest, Matthijs (2009) "An Empirical Model of Search with Vertically Differentiated Products", manuscript.

Zhao, Ying (2006) "Price Dispersion in the Grocery Market", Journal of Business, 79 (3): 1175-1192.