The London School of Economics and Political Science

THREE ESSAYS
IN
EMPIRICAL INDUSTRIAL ORGANIZATION

Ricardo Ribeiro

Dissertation submitted to the Department of Economics
of the London School of Economics and Political Science
for the degree of Doctor of Philosophy

London

August 2010

Declaration

I certify that the thesis I have presented for examination for the MPhil/PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Abstract

This dissertation consists of three essays in empirical industrial organization. Chapter 1 introduces the three essays and describes their main results. In chapter 2, I examine consumer demand for variety. The consumption of a good typically creates satiation that diminishes the marginal utility of consuming more. This temporal satiation induces consumers to increase their stimulation level by seeking variety and therefore substitute towards other goods (substitutability across time) or other differentiated versions (products) of the good (substitutability across products). The literature on variety-seeking has developed along two strands, each focusing on only one type of substitutability. I specify a demand model that attempts to link these two strands of the literature. This issue is economically relevant because both types of substitutability are important for retailers and manufacturers in designing intertemporal price discrimination strategies.

In chapter 3, which draws upon joint work with Peter Davis, we specify a new method of uncovering demand information from market level data on differentiated products. In particular, we propose a globally consistent continuous-choice demand model with distinct advantages over the models currently in use and describe the econometric techniques for its estimation. The proposed model combines key properties of both the discrete- and continuous-choice traditions: *i*) it is flexible in the sense of Diewert (1974), *ii*) it is globally consistent in the sense it can deal with entry and exit of products over time, and *iii*) incorporates a structural error term.

In chapter 4, I examine market dominance and barriers to competition in financial trading venues. As of 1 November 2007, the Market in Financial Instruments Directive introduced venue competition in the European cash trading market. I argue that, although positive, the impact on the degree of actual competition may be limited due to two barriers to competition: *i*) direct network effects together with increasing returns to scale and *ii*) post-trading constraints.

Acknowledgements

First and foremost I would like to thank my family for their unconditional support and love: my mother Emilia, my father Serafim, my brother Filipe, and my dearest niece Gabriela. They make it all worthwhile.

I feel particularly indebted to Peter Davis and John Van Reenen for their guidance, continuous help and, in particular, for never giving up on me. I am also grateful to my friends at LSE for making my everyday life at school much more enjoyable: Claudia Custódio, Alberto Galasso, Pedro Gomes, Sónia Gonçalves, Ana Matos, Hans Naudts, Kênia Parsons, Carlos Santos and Stephan Seiler. I want to give special thanks to Stefan Penczynski for his continuous encouragement.

I also want to thank Peter Brown for accepting me at Netherhall House and all the residents that helped me feel at home while away from home: Rui Agostinho, Amir Ahmadi, Ali Ahmed, Dimitris Akrivos, Aamir Asad, Guilherme Bandeira, João Bettencourt, Archan Boonyanan, Fernando Bueno, Manuel Castro, Miguel Constante, Joaquín Crespo, Roland Fisher, Ignácio Gil, Giorgos Grigoriou, Kevin Gouder, Santiago Herrasti, Florian Keppler, Vibhor Kumar, Maximilian Majewski, Domingo Martin, Fabrício Mattos, Gregor McWilliams, Pedro Muñoz, Kamil Olesiejuk, Reuven Proença, Prakarsh Sigh, Giriraj Singh, Alexandre Silva, Jan Stjernsward, Supachoke Thawornkaiwong and Juan Villalobos. I am truly grateful for the long night talks with Kevin and for all the cinema evenings with João and Manuel.

Finally, I gratefully acknowledge financial support from the Fundação para a Ciência e a Tecnologia (scholarship SFRH/BD/18629/2004/5A6F), from the Kauffman Foundation and the NET Institute (www.netinst.org).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Ithaka gave you the marvelous journey.
Without her you would not have set out.
She has nothing left to give you now.

And if you find her poor, Ithaka won't have fooled you.
Wise as you will have become, so full of experience,
you will have understood by then what these Ithakas mean.[1]

This dissertation consists of three essays in empirical industrial organization. Chapter 1 introduces the three essays and describes their main results. In chapter 2, I examine consumer demand for variety. The consumption of a good typically creates satiation that diminishes the marginal utility of consuming more. Temporal satiation induces consumers to increase their stimulation level by seeking variety and therefore substitute towards other goods (substitutability across time) or other differentiated versions (products) of the good (substitutability across products).

The literature on variety-seeking has developed along two strands, each focusing on only one type of substitutability. I specify a demand model that attempts to link these two strands of the literature. This issue is economically relevant because both types of substitutability are important for retailers and manufacturers in designing intertemporal price discrimination strategies. The consumer demand model specified allows consumption to have an enduring effect and the marginal utility of the different products to vary over consumption occasions. Consumers are assumed to make rational purchase decisions by taking into account, not only current and future satiation levels, but also prices and product choices. I can then use the model to simulate the demand implications of a major pricing policy change from hi-low pricing to an everyday low pricing strategy. To my knowledge, there is only one study that structurally addresses consumer response to such major policy changes, Erdem *et al.* (2003). However, they studied storable goods and do not allow for switching costs. I find similar patterns deriving from an entirely different source of dynamics, the stock of past consumption.

[1] Ithaka, C.P. Cavafy (translated by E. Keeley and P. Sherrard).

I apply the model to an indulgence good: ice cream (and related frozen desserts). The reason is twofold. First, ice cream constitutes the textbook illustration of the diminishing marginal utility concept and the industry is characterized by a high degree of product differentiation. Second, the temptation nature of the good makes stockpiling less relevant. This is important because in a context where temporary price promotions are a key marketing tool, if consumers respond to temporary price cuts by accelerating (anticipating) purchases and hold inventories for future consumption (i.e. stockpile), the separate identification of satiation and stockpiling would be somewhat problematic. I show below that, even though consumers do anticipate purchases in response to temporary price promotions, they do not stockpile, maybe because of the temptation feature of the good.

I find evidence that consumption has a lasting effect on utility that induces substitutability across time and that the median consumer has a taste for variety in her product decisions. Consumers are found to be forward-looking with respect to the duration since the last purchase, to price expectations and product choices. Pricing policy simulations suggest that retailers may increase revenue by reducing the variance of prices, but that lowering the everyday level of prices may be unprofitable.

In chapter 3, which draws upon joint work with Peter Davis, we specify a new method of uncovering demand information from market level data on differentiated products. We propose a globally consistent continuous-choice demand model with distinct advantages over the models currently in use and describe the econometric techniques for its estimation. The proposed model combines key properties of both the discrete- and continuous-choice traditions: *i*) it is flexible in the sense of Diewert (1973, 1974), *ii*) it is globally consistent in the sense it can deal with entry and exit of products over time, and *iii*) incorporates a structural error term. In order to encompass different possible real-world applications, we consider two alternative specifications of the baseline model depending on the degree of flexibility the researcher is willing to accept for the substitution patterns between inside and outside goods. The estimation procedure follows an analog to the algorithm derived in Berry (1994), Berry, Levinsohn and Pakes (1995). Depending on the specification considered, the contraction mapping for matching observed and predicted budget shares may be analytical or not. The case for which the contraction is analytic is relatively simple and fast to estimate which can prove a key advantage in some applications such those in competition policy, where time and transparency are important. For the case it is not analitic, we propose an alternative to Berry, Levinsohn and Pakes (1995)'s contraction mapping with super-linear rate of convergence. Finally, we provide a series of Monte Carlo experiments to illustrate the estimation properties of the model and discuss how it can be extended to cope with consumer dynamic behaviour.

In chapter 4, I examine market dominance and barriers to competition in financial trading venues. The interaction between competition and economic growth is a well established fact in the literature, with competition impacting economic growth via a more efficient allocation of market resources that contributes to "better economic performance, better prices and better services for consumers and businesses" (Kroes (2007)). Recent years have witnessed a strong and ferocious promotion of competition in a large spectrum of markets and the financial trading industry is no exception. As of 1 November 2007, the Market in Financial Instruments Directive (MiFID) aims to increase competition and to foster client protection in the European financial market. Among other provisions, it abolishes the concentration rule and challenges the market power of existing trading venues. The directive introduces venue competition in order to achieve better execution and ultimately lower trading costs. I argue that, although positive, the impact on the degree of actual competition may be limited due to two barriers to competition: *i*) direct network effects together with increasing returns to scale and *ii*) post-trading constraints since venues typically bundle trading and post-trading services.

The trading decision can be decomposed in two stages. First, investors decide the characteristics of the order and send it to a financial intermediary to be executed. Second, after receiving the order, the intermediary decides the trading venue where to execute it, conditional on the order characteristics received. I take the first stage as given and propose to model the choice of financial intermediaries in the second stage. I specify a structural discrete-choice multinomial random-coefficients logit demand model for the choice of venue that takes into account the trade-off between the different costs incurred during a trade. These costs can be divided into two broad categories: explicit and implicit costs. Explicit trading costs denote the transaction costs of a venue and include the costs of executing the order (trading fees) and the costs of post-trading (clearing and settlement fees). Implicit trading costs relate to the liquidity of a venue and typically include the bid-ask spread, the potential impact of a trade and the opportunity cost of missed trades. Implicit trading costs are important since cash trading exhibits direct network effects. The valuation of financial intermediaries for a venue is increasing in the number of other agents that choose the same venue as it reduces the costs of finding a counterpart. A more liquid venue translates into lower implicit trading costs as it *i*) stabilizes the market price of a financial instrument, and *ii*) reduces the extent to which placing an order has an adverse effect on the corresponding price.

I apply the model to the set of 16 most traded securities in the FTSE 100 following the list of liquid securities published (and updated regularly) by CESR after the implementation of MiFID. The results imply that financial intermediaries tend to value liquidity more than total fees when deciding to which venue to route a given order for execution. For this reason the incumbent venue has a clear advantage relative to its competitors and can, as a result, exert market power when setting its fees level. After estimating the degree of substitutability between the different trading venues, I examine the impact of the mentioned barriers to competition. First, I study the role of direct network effects by computing the counterfactual market shares that would arise if there were no liquidity differences across venues. Then, I evaluate the impact of the post-trading constraints induced by the typical bundle of trading and post-trading services. I simulate the equilibrium market shares that would arise if the different trading services were fungible. In both cases, the results suggest that eliminating the corresponding barrier to competition is associated with a significant decrease (of similar magnitude) in the asymmetry of the industry.

Chapter 2

Consumer Demand for Variety: Intertemporal Effects of Consumption, Product Switching and Pricing Policies

"*Rachel*: Hi!
Chandler: Another cheesecake came! They delivered it to the wrong address again!
Rachel: So just bring it back downstairs, what's the problem?
Chandler: I can't seem to say goodbye.
Rachel: Are you serious?! Chandler, we ate an entire cheesecake two days ago and you want more?"

Friends, Episode 7-11, The one with All The Cheesecakes

The concept of diminishing marginal utility is a cornerstone of economic theory. The consumption of a good typically creates satiation that diminishes the marginal utility of consuming more. The length of time that the marginal utility is diminished is likely to vary across goods and, as Rachel and Chandler's cheesecake episode illustrates, across consumers. While it only took Chandler two days for his cheesecake marginal utility to return to pre-consumption levels, Rachel seemed still satiated (suggesting that the utility provided by her prior cheesecake consumption had not yet faded).

Temporal satiation induces consumers to increase their stimulation level by seeking variety and therefore substitute towards other consumption alternatives. In this chapter, I define a forward-looking dynamic discrete choice model of demand that, similarly to Hartmann (2006), allows consumption to have an intertemporal effect: consuming produces a consumption capital stock that provides utility over time until it gradually depreciates. However, unlike Hartmann (2006), I consider a differentiated products setting that allows consumers to switch not only towards other goods, but also towards other differentiated versions (products) of the good.

In each shopping trip, consumers decide whether or not to purchase the good, and in case they decide to purchase, which quantity and product to buy. Consumers are assumed to make rational purchase decisions by taking into account, not only current and future satiation levels, but also prices and product choices. Price expectations are an important determinant of intertemporal substitution. If prices are expected to be higher in the future, consumers may anticipate their purchase decisions and vice-versa. Another important feature is consumer product choice. The marginal utility of different products is allowed to change over consumption occasions, depending on the switching costs of the individual. Consumers with low switching costs may exhibit shorter interpurchase durations than those that incur in high switching costs whenever they alter their product choice.

This chapter relates to the literature on variety-seeking developed from Jeuland (1978) and McAlister (1982). Jeuland (1978) explains variety-seeking behaviour by proposing that prior experience with a good decreases the consumer's utility for that good, which constitutes a direct application of the diminishing marginal utility concept. This explanation is predictive of the consumer's tendency to switch away from the most recently consumed good. McAlister (1982) then refined the explanation by proposing that prior experience with the *attributes* of a good decreases the consumer's utility for goods with similar attributes, which refocuses the diminishing marginal utility concept over attributes rather than goods. These two explanations have governed the subsequent development of an extensive literature centered on the implications of switching costs for consumer choice (see, for example, Keane (1997)). Recently, Hartmann (2006) extended the variety-seeking literature by allowing intertemporal effects of consumption, that is, by allowing consumption to have a lasting effect that diminishes the marginal utility of future consumption. However, the homogeneous nature of the good studied, golf, did not allow him to focus on product switching.

This chapter attempts to link the two strands of the literature on variety-seeking by allowing consumers to substitute towards other goods (substitutability across time), as well as to other differentiated products of the same good (substitutability across products). This issue is economically relevant because both types of substitutability are important for retailers and manufacturers in designing intertemporal price discrimination strategies. I specify a consumer demand model which allows consumption to have an enduring effect and allows the marginal utility of the different products to vary over consumption occasions. The main contribution of the chapter is to study how different pricing policies affect consumer demand for goods with such enduring effects of consumption and that are characterized by a high degree of differentiation. The model can then be used to simulate the demand implications of major pricing policy changes like a shift from hi-low pricing to everyday low pricing. To my knowledge, there is only one study that structurally addresses consumer response to such major policy changes, Erdem *et al.* (2003). However, they studied storable goods and do not allow for switching costs. I find similar patterns deriving from an entirely different source of dynamics, the stock of past consumption.

The state space implied by a dynamic problem where forward-looking consumers make optimal decisions in light of current and future satiation levels, prices and product choices is, in a differentiated products setting, extremely large for practical estimation. In order to reduce the dimensionality of the state space, I adopt a multi-stage budgeting approach that decomposes the consumers decision into a quantity choice and a product choice (see Aguirregabiria (2002) and Hendel and Nevo (2006a) for similar dynamic applications of Gorman (1971)'s approach). Under the set of assumptions discussed below, I show that the consumer product choice *conditional on the quantity purchased* does not depend on dynamic considerations. This simplifies the estimation of many of the parameters of the demand model, while the remaining ones are estimated solving a simplified dynamic problem that involves only quantity and timing decisions.

I estimate the different stages of the model by maximum-likelihood and solve the dynamic programming problem by using value function parametric approximation with policy function iteration in the lines of Benitez-Silva *et al.* (2000). In order to control for unobserved heterogeneity, I incorporate a rich specification. This is important since in both stages unobserved consumer heterogeneity may confound the inference of true state-dependence effects. As Heckman (1981) points out, if households have different preferences "and if these differences are not properly controlled, previous experience may appear to be a determinant (...) of future experience solely because it is a proxy for temporally persistent unobservables that determine choices." The state dependence in the product decision arises because the marginal utility of different products is allowed to change over consumption occasions, while the state dependence in the quantity decision is induced by the consumption capital stock. For reasons I discuss below, I incorporate *observable* heterogeneity in the product choice and a *continuous distribution* of consumer heterogeneity in the quantity decision.

I apply the model to an indulgence good: ice cream (and related frozen desserts). The reason is twofold. First, ice cream constitutes the textbook illustration of the diminishing marginal utility concept and the industry is characterized by a high degree of product differentiation. Second, the temptation nature of the good can (and in fact does, for the empirical application considered) make stockpiling limited in relevance (and in particular duration). This is important because in a context where temporary price promotions are a key marketing tool, if consumers respond to temporary price cuts by accelerating (anticipating) purchases and hold inventories for future consumption (i.e. stockpile), the separate identification of satiation and stockpiling would be somewhat problematic. I show below that, even though consumers do anticipate purchases in response to temporary price promotions, they do not stockpile, maybe because of the temptation feature of the good.

I find evidence that consumption has a lasting effect on utility that induces substitutability across time and that the median consumer has a taste for variety in her product decisions. Consumers are found to be forward-looking with respect to the duration since the last purchase, to price expectations and product choices. Pricing policy simulations suggest that retailers may increase revenue by reducing the variance of prices, but that lowering the everyday level of prices may be unprofitable.

2.1 Data Description and Preliminary Analysis

I use Information Resources Inc. (IRI) scanner data collected from June 1991 to May 1993 in two separate submarkets of a large Midwest city. The dataset covers 24 different product categories at both the store and household levels. The former includes weekly sales, prices, and promotional activities for each universal product code (UPC) in nine supermarkets, belonging to different chains, while the latter tracks the store visits of 548 households and includes when and how much each household spent in her shopping trips.

I estimate the model for an indulgence good category: ice cream and related frozen desserts. Frozen desserts are offered in four segments: regular ice cream, diet ice cream, frozen yoghurt and ice milk. Regular ice creams account for 67% of the volume purchased, with diet ice creams and frozen yoghurt roughly splitting the remaining of the market. The market share of ice milk is less than one percent. Ice creams come in a limited number of package sizes, with the top four sizes accounting for more than 99% of the market: 64 *oz.* (72.3%), 16 *oz.* (11.5%), 160 *oz.* (10.8%) and 32 *oz.* (4.8%). The choice set available to the households is substantial. The average supermarket in the sample carries 170 different frozen dessert products (from 20 brands) on a weekly basis. I defined a product as a segment-brand-flavour combination so that, for example, Häagen-Dazs Vanilla Ice Cream, Häagen-Dazs Chocolate Ice Cream, and Häagen-Dazs Vanilla Frozen Yoghurt are classified as distinct products.

Kemps is the dominant brand with 23% volume market share, followed by Breyers and Wessanen's Value Pack (both with 12%), Dreyer's (10%) and Häagen-Dazs (6%). Store private labels account for 3.5% of the market. The most popular flavours are vanilla (21%), chocolate (9%), neapolitan (7%), strawberry (5%) and chocolate chip (5%), although a typical supermarket would carry an average of 84 different flavours, each week. In contrast with the moderated brand and flavour concentration, there is substantial market fragmentation at the product level. Breyers Vanilla Ice Cream is the market leader with a 2.6% volume market share.

The median household has two members and an income between 25,000 and 35,000 dollars. I conduct the subsequent analysis using a subset of the sampled households selected on the basis of three criteria. First, I eliminated consumers recorded purchasing in supermarkets for which no price data is available. An alternative approach could have been to include those households and either (i) eliminate the purchases in unsampled stores as if they never happened, or (ii) assume some cross-store price pattern and generate price data to be imputed for those purchases. All solutions potentially could introduce bias in the analysis. I opted for the elimination after ensuring the subset sample was representative, an issue I discuss below. Second, computational barriers compelled me to eliminate consumers that purchased more than two items of ice cream in a shopping visit or bought non-representative package sizes. Their inclusion would increase the dimensionality of the state space to a degree that made the structural estimation computationally infeasible. Finally, I eliminated households that made less than 10 purchases of ice creams over the total sample period since they are likely to be either (i) not regularly in the market, or (ii) purchasing in alternative stores. This reduced the sample size from 548 to 115 consumers, who made a total of 17,899 supermarket trips and 2,822 ice-cream purchases.

An important question is obviously whether the subset sample is representative of the whole population buying at these supermarkets. Table 2.1 addresses this question by reporting, for the different samples, the top-8 products, brands and flavours in terms of their volume market share. The simple comparison of the columns show that, with minor exceptions, the product, brand and flavour market shares in the different samples are very similar, which is suggestive that the subset sample is reasonably representative.

2.1.1 Substitutability Across Time

In this section, I examine the shopping behaviour of consumers and the frequency of their purchasing patterns for the ice cream category as a whole. Table 2.2, Panel A presents summary statistics for the consumers supermarket trips. Although there is evidence of substantial heterogeneity across consumers with regard to their shopping behaviour, the median consumer in the sample visits a supermarket every three days to a total of 98 times over the

TABLE 2.1

*Volume Market Shares**

Panel A: Product level		Household				Store	
		All		Subset			
		S	CS	S	CS	S	CS
1	Wessanen's Value Pack NY Vanilla IC	3.85	3.85	4.14	4.14	2.03	2.03
2	Wessanen's Value Pack Vanilla IC	2.71	6.56	1.96	6.10	2.58	4.61
3	Wessanen's Value Pack Neapolitan IC	2.06	8.62	2.59	8.69	1.78	6.39
4	Fieldcrest Vanilla IC	2.05	10.67	2.81	11.50	1.22	7.61
5	Kemps Vanilla FY	1.97	12.64	5.26	16.76	1.13	8.74
6	Kemps Vanilla IC	1.69	14.33	1.37	18.13	2.15	10.89
7	Breyers Vanilla IC	1.51	15.84	1.62	19.75	2.60	13.49
8	Wessanen's Value Pack Chocolate IC	1.41	17.25	1.26	21.01	1.14	14.63

Panel B: Brand level		Household				Store	
		All		Subset			
		S	CS	S	CS	S	CS
1	Kemps	21.22	21.22	19.59	19.59	23.37	23.37
2	Wessanen's Value Pack	13.69	34.91	12.71	32.30	11.73	35.10
3	Breyers	9.22	44.13	10.04	42.34	11.86	46.96
4	Dreyer's	8.89	53.02	5.34	47.68	9.88	56.84
5	Sealtest	6.90	59.92	10.70	58.38	4.27	61.11
6	Fieldcrest	4.10	64.02	5.66	64.04	3.43	64.54
7	Dean Foods	3.75	67.77	3.85	67.89	2.71	67.25
8	Häagen-Dazs	3.62	71.39	1.53	69.42	5.81	73.06

Panel C: Flavour level		Household				Store	
		All		Subset			
		S	CS	S	CS	S	CS
1	Vanilla	22.20	22.20	26.38	26.42	21.10	21.10
2	Chocolate	8.00	30.02	9.04	35.46	8.59	29.69
3	Neapolitan	7.21	37.41	11.47	46.93	6.50	36.19
4	New York (NY) Vanilla	5.49	42.90	5.52	52.45	3.63	39.82
5	Strawberry	5.36	48.26	4.48	56.93	4.99	44.81
6	Butter Pecan	4.01	52.27	3.21	60.14	3.79	48.60
7	Chocolate Chip	2.69	54.96	2.59	62.73	4.63	53.23
8	Pistachio	2.39	57.35	2.94	65.67	1.31	54.54

* Columns labeled S denote market shares and columns labeled CS denote cumulative market shares. IQ stands for an ice cream product and FY for a frozen yoghurt product.

TABLE 2.2
*Consumer Category Purchasing Patterns**

Panel A: Supermarket trips	Mean	Median	Std	Min	Max
Number of Trips	114	98.0	59.9	36.0	317
Days from Previous Trip	4.75	3.00	5.05	0.00	75.0
Number of Stores Visited	1.95	2.00	0.79	1.00	4.00
Store HHI	0.86	1.00	0.20	0.36	1.00
Panel B: Ice cream purchases	**Mean**	**Median**	**Std**	**Min**	**Max**
Number of Purchases	17.7	13.0	15.0	3.00	126
Volume	63.3	64.0	31.5	16.0	160
Multiple-item Purchases	0.12	0.00	0.32	0.00	1.00
Days from Previous Purchases	30.0	16.0	37.8	0.00	311

* For Number of Trips, Number of Stores Visited and StoreHHI, an observation is a household. For Days from Previous Trip, an observation is a trip instance. For all other statistics, an observation is a purchase instance. Store HHI denote the household's Herfindahl-Hirschman index for ice-cream volume purchases.

observed sample. This consumer shops in two different supermarkets, but concentrates her purchases on a single one. In order to compute the consumers intertrip duration, I use the first six months in the sample to generate an initial trip for each household. I will discuss below that these first six months will also be instrumental in generating an initial product choice for each consumer to avoid spurious switching.

Table 2.2, Panel B displays some summary statistics of households ice-cream purchasing patterns. The results suggest substantial heterogeneity also at this level, with the median consumer making a single-item purchase of 64 *oz.* of ice-cream every 16 days to a total of 13 purchases over the sample period.

I now move on to examine the hypothesis that consumption has a lasting effect that diminishes the marginal utility of future consumption. If the magnitude of this effect is such that induces consumers to vary their choice of dessert, the probability of purchase will be related to how long it has been since their last purchase. If, on the other hand, the magnitude of the effect is small, then the probability of purchase will not depend on interpurchase duration. Figure 2.1 displays the purchase hazard rate by no-purchase spell duration in days. The hazard rate denotes here the probability that you purchase if you have not purchased up to now. The pattern illustrated provides some support for the duration dependence argument: there is evidence of a non-linear relationship between the probability of purchase and the duration since the last purchase. The hazard rate is quite low immediately after a purchase, then gradually increases until day 7, after what it exhibits a gradual, although rather jagged, downward trend (an interesting aspect of this hazard rate relates to its recurring spikes, an issue I address below). However, the downward trend of the hazard rate suggests that, in contrast to the initial argument of this chapter, the probability of purchase seems to decrease (and not increase) with the duration since last purchase. There are two possible explanations for this behaviour (and that illustrate the well known problem that unobserved heterogeneity can be confounded with state-dependence). Either the utility from consuming ice cream does in fact decrease with duration (positive state-dependence), or alternatively, the utility increases with duration from last purchase (negative state-dependence), but there exists an over-representative group of low-demand consumers who are more likely to exhibit longer interpurchase durations (heterogeneity). In order to evaluate the degree of consumer unobserved heterogeneity, I re-compute the hazard rate at the *consumer-level.* The consumer-level hazard rate denotes here the probability that a consumer purchases if she has not purchased up to now. Figure 2.2 displays, for each duration spell, the mean of the probability of purchase across consumers, as well as the interval limited by that mean value plus and minus one standard deviation (subject to a non-negativity constrain). The high standard deviation around the mean indicates substantial heterogeneity across consumers, which is suggestive of the importance of controlling for unobserved heterogeneity in the structural estimation.

While the duration dependence of ice cream purchases implied by Figures 2.1 and 2.2 can be consistent with variety-seeking behaviour induced by a diminishing marginal utility, it can also be consistent with the main alternative theory: if consumers respond to temporary price cuts by accelerating (anticipating) purchases and hold inventories for future consumption (i.e. stockpile), the probability of purchase will also be duration-dependent. Temporary price promotions are an important marketing tool in the pricing strategy of many nondurable goods and ice creams are no exception. The ice cream prices in the sample display a classic high-low pattern: products have a "regular (modal) level" that remains constant for long periods of time with occasional temporary reductions. Figure 2.3 displays, as an illustration, the price of Dreyer's Vanilla Ice Cream 64 oz. over the sample weeks in a typical supermarket. The price is at the "regular level" ($4.59) for 57% of the weeks. Defining a sale to be a price reduction of at least 5% (below the modal level), it is on sale for 31% of the time, with the average price discount being $1.61. If we consider the sample as a whole, untabulated statistics show that prices are, on average, 66% at the "regular level" and 26% on sale (with an average discount of $0.70). In such an environment, consumers may respond to temporary price cuts by accelerating (anticipating) purchases and stockpile.

Table 2.3 addresses the purchase acceleration effect by comparing household level sale and nonsale purchasing patterns. The first column displays averages during nonsale purchases.

FIGURE 2.1
Purchase Hazard

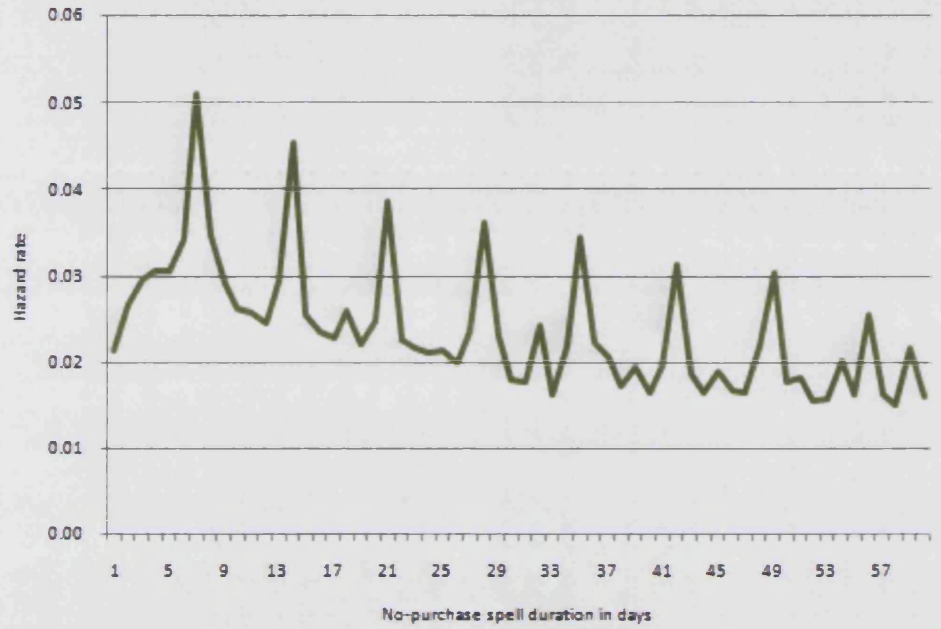


FIGURE 2.2
Consumer-level Purchase Hazard

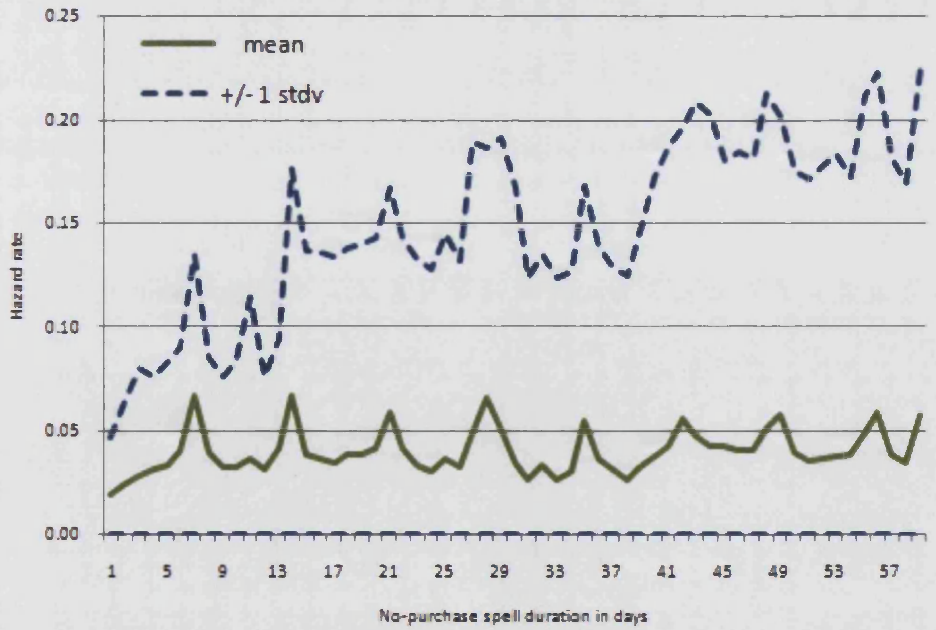
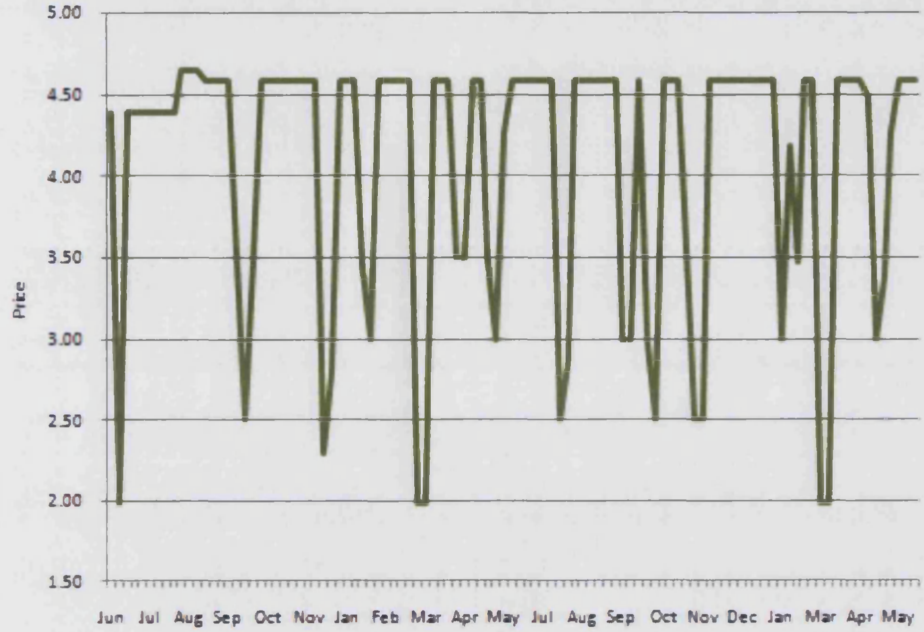


FIGURE 2.3

Price Example: Dreyer's Vanilla Ice Cream



The following columns examine the difference towards a sale purchase, decomposing the total difference into a within and a between households effects. As before, a sale is defined as any price at least 5% below the modal price of a store-UPC combination over the observed period. I focus the analysis on the within column,[2] that compares the household purchasing patterns over time. The evidence seems to indicate that consumers do respond to temporary price cuts. Unsurprisingly, the results suggest that households tend to shorten their duration from previous purchase (between 3-4 days) and to increase their volume purchases (by roughly 17%), when buying on sale.

The interesting question is whether this response by consumers translates into a consumption effect or merely represents a demand-anticipation effect with households stockpiling for future consumption. In order to examine this question, I follow Hendel and Nevo (2006a) and examine households interpurchase duration to the next purchase. The idea here is that if consumers do stockpile, then the duration to next purchase is expected to be longer for large volume purchases (like purchases on sale). The results from Table 2.3 show that there is no significant difference in the duration forward to next purchase between sale and non-

[2] The results from the between column in Table 2.3 suggest substantial heterogeneity in how consumers respond to temporary price cuts, with households that purchase more frequently on sale, buying larger volumes and less frequently. This reinforces the need for the structural model to control for consumer heterogeneity.

TABLE 2.3

*Category Purchasing Patterns: Comparison between Sale and Nonsale**

	Average during Nonsale	Difference during Sale			Week F.E.
			Consumers		
		Total	Within	Between	
Volume (oz.)	57.7 (3.08)	11.8 (2.26)	9.91 (1.42)	23.3 (7.70)	No
	–	12.1 (2.30)	10.1 (1.51)	22.5 (12.7)	Yes
Units	1.05 (0.01)	0.16 (0.02)	0.17 (0.02)	0.13 (0.04)	No
	–	0.16 (0.02)	0.17 (0.02)	0.07 (0.08)	Yes
Average Package Size	56.1 (2.95)	2.27 (1.94)	-0.31 (0.91)	14.3 (6.91)	No
	–	2.22 (1.97)	-0.18 (0.90)	13.0 (12.2)	Yes
Days from Previous Purchase	27.0 (2.40)	0.22 (1.83)	-3.69 (1.54)	17.6 (7.04)	No
	–	-0.13 (1.77)	-4.00 (1.51)	19.8 (9.84)	Yes
Days to Next Purchase	25.9 (2.31)	2.55 (1.65)	-0.14 (1.39)	9.23 (7.03)	No
	–	2.59 (1.64)	-0.14 (1.44)	28.3 (11.5)	Yes

* An observation denotes a purchase instance. Standard errors clustered by household in parentheses (except for the between analysis).

sale purchases. The comparison of the quantity and duration effects seem to indicate that stockpiling may not be a relevant feature of ice-cream demand and that the quantity effect induced by temporary price reduction substantiates a consumption effect. In order to examine the robustness of this conclusion, I also compared the duration forward to the next purchase when consumers buy an above average volume. The results of those regressions (which are untabulated) are consistent with the above conclusion. The difference in inter-purchase duration is again not significantly different from zero. The alternative theory that state-dependence in the probability of purchase is due to stockpiling can not explain these results. Furthermore, the analysis of how the additional quantity is bought is consistent with the variety-seeking theory. When purchasing on sale, consumers do not significantly change their average package size. Instead, they purchase more units of ice cream. This supports the variety-seeking story since if the increased volume translates into increased consumption, then purchasing multiple-items is a sensible strategy to deal with the diminishing marginal utility from consumption.

Having addressed the issue of eventual stockpiling behaviour, I now move on to address another somewhat problematic issue. In this chapter, I model consumption to have a lasting effect that diminishes the marginal utility of future consumption. However, I do not observe the actual time and magnitude of consumption. So, I am forced to infer it from purchase choices. The results from Table 2.3 provide some evidence that, not only consumers do not anticipate purchases to hold inventories for future consumption, as already discussed, but also that utility does not depend on the stock of past consumption. If consumption did create a stock, then duration to next purchase would increase with the size purchased, which it does not. That said, following Table 2.3, the only relevant variable that may affect the marginal utility of future consumption is the timing of current consumption. As I discuss below, due to the temptation nature of the good, assuming that the time of consumption coincides with the time of purchase is not unreasonable. At least for most people, in line with what Erdem *et al.* (2003) argue, ice creams are technologically, but not practically storable over more than a few days. It may seem inconsistent to assume that consumption has a lasting affect that induces intertemporal substitution in purchases while assuming that the good held in inventory has a temptation feature. These assumptions are however consistent with observed behaviour, since consumers seem to depreciate the costs of goods they have in inventory (see Gourville and Soman (1998) and Prelec and Loewenstein (1997)).

I now move on to describe two other timing aspects of consumers ice cream category purchasing patterns. I begin by addressing seasonality. If the decisions of consumers are seasonal, then the structural model must reflect this feature. Table 2.4 addresses this question by comparing household level summer and nonsummer purchasing patterns. The first column displays averages during nonsummer purchases while the following columns examine the difference towards a summer purchase, again decomposing the total difference into a within and a between consumers effects. The results suggest that summer does not induce a significant difference in the purchasing patterns of households, at any dimension: volume, units, average package size, days from previous purchase or days to next purchase. This holds both within and across consumers. In order to examine the robustness of this conclusion, I replicated this analysis to compare the consumers purchasing patterns in winter and nonwinter seasons. The results of those regressions (which are untabulated) are consistent with the above conclusion. They show no significant difference in the associated purchasing patterns. Surprisingly, seasonality does not seem therefore to be an important feature in the purchasing decision of ice cream and related frozen desserts.

Another timing aspect of consumers ice cream category choice patterns relates to the purchase day. If consumers are more likely to purchase on a particular day of the week or weekend, then the structural model must somehow incorporate it. Untabulated statistics show no evidence of a clear preference towards a given day of the week, when comparing *across* consumers. However, Figure 2.1 illustrated an interesting pattern. The probability of purchase spikes at every seven days (and *exactly* every seven days), which suggests that even though no preference exists *across* consumers, each consumer seems to have a preferred day of the week to purchase ice creams - maybe at their main weekly shopping trip. This constitutes a feature of consumer behaviour that must be incorporated into the structural

TABLE 2.4
*Category Purchasing Patterns: Seasonality**

	Average during Nonsummer	Difference during Summer		
			Consumers	
		Total	Within	Between
Volume (oz.)	63.1 (0.78)	0.76 (1.74)	1.86 (1.25)	2.54 (18.7)
Units	1.12 (0.01)	0.00 (0.02)	0.00 (0.02)	-0.17 (0.10)
Average Package Size	57.0 (0.63)	0.94 (1.39)	1.58 (0.86)	12.1 (16.4)
Days from Previous Purchase	27.0 (0.86)	0.74 (1.86)	-0.72 (1.68)	28.8 (13.8)
Days to Next Purchase	27.1 (0.86)	0.09 (1.86)	-1.01 (1.68)	15.2 (16.4)

* An observation denotes a purchase instance. Standard errors clustered by household in parentheses (except for the between analysis).

model.

2.1.2 Substitutability Across Products

Having described the shopping behaviour and purchasing patterns of consumers for the ice cream category as a whole, I now move on to examine their product choice patterns. Table 2.5 displays household-level concentration and variety-seeking measures for ice cream product and brand choices. Table 2.5, Panel A displays the descriptive statistics for the product level measures. The median consumer buys 8 different products over the sample period and fragments her volume purchases considerably as the relatively low household-level concentration ratios (CRm) and Herfindahl-Hirschman index (HHI) suggest. Nevertheless, there is evidence of substantial heterogeneity across consumers as indicated by the large range intervals and standard deviation of the several concentration measures. So, although some households show evidence of considerable product fragmentation, others concentrate their purchases on a relatively small number of products.

Having examined product choice concentration, I now move on to examine a measure of product switching, following Menon and Kahn (1995). The probability of *successive* product switching denotes the proportion of consumer purchases that involved switching, where a switch is defined as occurring each time the product(s) chosen on a purchase occasion is different from those chosen on the immediately preceding purchase instance. This is consistent with Faison (1977) and Venkatesan (1973). Counting switching from the beginning of the sample period would generate spurious switching. Therefore, as discussed previously, I use the first six months in the sample to generate an initial product choice for each consumer. This approach is similar to Shum (2004) and Pozzi (2009). The descriptive statistics for the probability of successive product switching suggest substantial heterogeneity across con-

TABLE 2.5
*Consumer Product Choice Behaviour**

Panel A: Product level	Mean	Median	Std	Min	Max
CR1	0.38	0.33	0.19	0.11	0.95
CR5	0.83	0.84	0.16	0.46	1.00
HHI	0.26	0.20	0.17	0.07	0.91
Number of different products	8.27	8.00	4.18	2.00	26.00
Probability of successive product switching	0.77	0.82	0.22	0.15	1.00
Probability of product exploration switching	0.45	0.42	0.26	0.00	1.00
Panel B: Brand level	Mean	Median	Std	Min	Max
CR1	0.57	0.53	0.23	0.19	1.00
CR5	0.97	1.00	0.06	0.73	1.00
HHI	0.46	0.38	0.25	0.13	1.00
Number of different brands	4.43	4.00	2.10	1.00	11.00
Probability of successive brand switching	0.57	0.62	0.27	0.03	1.00
Probability of brand exploration switching	0.33	0.30	0.20	0.03	1.00

* An observation is a household. CRm and HHI denote the household's m-product (brand) volume concentration ratio, and Herfindahl-Hirschman index, respectively.

sumers, with the median household switching from the immediately preceding products in 82% of her purchases.

An alternative approach would be to define a switch as occurring each time the product chosen on a purchase occasion is different from any of the preceding choices, following Faison (1977) and Pessemier (1985). The two definitions differ in the idea of variety-seeking that may capture. While the latter definition implicitly assumes that the level of stimulation of a household can only be increased by exploring new products, i.e. products that the consumer never tried before, the successive switching definition assumes that the level of stimulation of a household can be increased by alternating from one product to another, even if the products in the switching set are all familiar. A simple comparison of the probability of switching according to the two definitions suggests that the proportion of switching involving familiar products should not be neglected.

Table 2.5, Panel B presents descriptive statistics for the same concentration and variety-seeking measures, but aggregated at the brand level. The purchases of the median consumer show a higher degree of concentration and a lower probability of switching when compared with her product choice patterns, which may be suggestive of the relative importance of different flavours and product types in increasing the level of stimulation of a household.

One problem with inferring variety-seeking from product switching is that unobserved heterogeneity can be confounded with product state-dependence. The identification problem arises because a consumer may exhibit high product switching by repeatedly alternating products in her purchases either because of a weak unobserved, idiosyncratic preference for the different products or because she has a taste for variety. In order to evaluate the importance of product preferences, I examine the association between product switching behaviour and product choice. The dependent variable is a product preference measure in the lines of Simonson and Winer (1992). Each consumer purchase is associated with a score equal to the volume market share of the corresponding product (or products in case it is a multiple-item purchase) in the consumer's shopping history.[3] Products with high consumer-level market shares are assumed to correspond to products for which the consumer has a strong preference, given their weight in the household shopping basket. This assumption is of course problematic, but it allows me to illustrate the high degree of unobserved product heterogeneity. Table 2.6 presents the OLS results of regressing the product preference score of each purchase on a product state-dependence variable that keeps track of the number of product switches from the immediately preceding purchase instance. I include marketing-mix variables as covariates: price and two types of promotional activities - feature (defined as any type of retailer product advertising) and display (defined as any type of special store display).[4] The different specifications vary on the degree of controls included. The results suggest a significant negative association between product switching and the product preference score, implying that consumers tend to switch more towards less preferred products. This association weakens as additional heterogeneity is incorporated, which is indicative of the importance of controlling for unobserved product heterogeneity in the structural estimation.

A final problem with inferring variety-seeking from product switching is that product unavailability may generate spurious switching. In order to address this concern, I need to separate *true* product switching from switching *induced* by product unavailability. However, as the IRI dataset does not include information on product availability, I have to infer it from store product sales. I consider that a product was available in a given week and supermarket combination if the store sold at least one unit of the product in that week.[5] I then compute, for each consumer, the proportion of *true* product switching, i.e. switching that occurs

[3] As an illustration consider the hypothetical example of a consumer that, over her shopping history, purchases 100 *oz.* of product A, 50 *oz.* of product W and 50 *oz.* of product B. Each item purchase of product A will receive a product preference score of 0.50 (100/200).

[4] The data includes several categories of feature and display. I aggregate across the different categories to feature/no feature and display/no display.

[5] The product availability proxy will obviously overestimate the induced product switching.

TABLE 2.6

*Product State-Dependence vs Product Preference**

	(1)	(2)	(3)	(4)
Product State-Dependence	-0.28	-0.31	-0.17	-0.10
	(0.06)	(0.06)	(0.03)	(0.01)
Marketing-Mix				
Price		0.04	0.07	0.01
		(0.01)	(0.01)	(0.01)
Feature		0.05	0.07	0.01
		(0.02)	(0.02)	(0.01)
Display		-0.00	0.01	-0.02
		(0.03)	(0.02)	(0.02)
Household F.E.	No	No	Yes	Yes
Product F.E.	No	No	No	Yes
R2	0.26	0.28	0.64	0.85

* An observation is a purchase instance by a household. Standard errors clustered by households in parentheses.

despite the products chosen on the immediately preceding purchase instance being available (as measured by the proxy). This analysis (which is untabulated) seems to suggest that most product switching is not induced by unavailability: the median proportion of *true* product switching is 83%, the average is 80%, the 25th percentile is 66%, and the 75th percentile is 98%.

Having described the important features of the purchasing behaviour of consumers, I now move on to specify the demand model and the proposed estimation procedure.

2.2 Demand Model

This section introduces the utility function and the assumptions of the model. I study the demand for a temptation good in a setting similar to Hartmann (2006) where consumption creates a stock that diminishes over time. This creates in the consumer an incentive to variety-seek and thus intertemporal substitute consumption for the good. Unlike Hartmann (2006), I extend the analysis to address the differentiated nature of the good and examine not only substitution across time, but also substitution across products. In order to do so, I adapt Aguirregabiria (2002) and Hendel and Nevo (2006a) multi-stage budgeting approach.

2.2.1 The Setup

There are I consumers who are indexed by i. In each shopping trip t, consumer i chooses whether or not to purchase the good, and in case she decides to purchase, which product and size to buy. Let $j = 1, \ldots, J$ index the inside product alternatives to the consumer, with each product alternative being (possibly) offered in a variety of different sizes x. Multiple-item purchases are included by expanding the choice set to allow for bundles. If in a particular trip a consumer buys, for example, both 64 $oz.$ of Häagen-Dazs Vanilla Ice Cream *and* 16 $oz.$ of Häagen-Dazs Chocolate Ice Cream, the purchase size is given by $x = 80$ $oz.$ and product j denotes the bundle of the two products. The no purchase choice (outside alternative) is indexed by $j = 0$.

2.2.2 Consumer Flow Utility

The consumer flow utility is expressed in terms of the indirect utility from each of the available alternatives. I begin by specifying the indirect utility from not purchasing (the outside option). I follow Hartmann (2006) and relax the common assumption of additively-separable utility in consumption by considering a frequency of purchase model where past choices affect current utility. In particular, I assume the utility of the outside option to be a function of the depreciated stock of past consumption:

$$u_{i0t}\left(y_{it}, \varepsilon_{i0t}\right) = z_{it} + \varepsilon_{i0t}, \tag{1}$$

where z_{it} denotes the stock of past consumption of individual i at time t and ε_{i0t} is a random shock to consumer choice. The depreciated stock of past consumption will, in full generality, depend on both the time elapsed since the previous consumption and the magnitude (or size) of past consumption. However, because the stock of past consumption is intangible and unobservable, I am required to infer it from (observed) past purchase choices. In order do so, I make the following assumptions.

Assumption 2.1 *Consumption takes place at the time of purchase.*

Assumption 2.1 is motivated by the temptation nature of the good. Since the data description analysis has shown that consumers do not anticipate purchases to hold inventories for future consumption, inferring that individuals do consume their purchased ice cream *before* their next purchase occasion is not unreasonable. However, the actual time of consumption is unobserved. Due to the temptation nature of the good, I assume that the time of consumption coincides with the time of purchase. At least for most people, in line with what Erdem *et al.* (2003) argue, ice creams are technologically, but not practically storable over more than a few days. As discussed above, it may seem inconsistent to assume that consumption has a lasting affect that induces intertemporal substitution in purchases while assuming that the good held in inventory has a temptation feature. These assumptions are, however, consistent with observed behaviour, since consumers seem to depreciate the costs of goods they have in inventory (see Gourville and Soman (1998) and Prelec and Loewenstein (1998)).

Assumption 2.2 *The stock of past consumption fully depreciates after a new consumption occasion.*

Assumption 2.2 implies that the stock of past consumption does not accumulate across multiple consumption occasions and that only the last consumption occasion is relevant. The motivation behind this assumption is twofold. First, it significantly reduces the state space since, instead of keeping track of all past choices, only the last consumption (which is observable under Assumption 2.1) is relevant to the decision of consumers. An alternative approach would consist of constructing an accumulated stock index, a strategy that would also have the advantage of a simplified state space. However, it carries a disadvantage related to the second justification for Assumption 2.2. Under this assumption, the initial stock of past consumption is observable and does not need to be inferred, which would not be true if I allowed the stock to accumulate across multiple consumption occasions.

One concern with the simplification implied by Assumption 2.2 is that it comes at a cost: the stock of past consumption is measured with error. Although measurement error is a potentially troublesome problem, it may not be too problematic here. The error introduced will, at best, underestimate the incidence of intertemporal substitution, rather than falsely induce finding intertemporal substitution. In order to understand why this is the case, note that because the estimated stock of past consumption will not exceed the *true* stock of past consumption, the utility of the outside option will be underestimated (in a setting where this alternative is the most common choice made by consumers). As a consequence, intertemporal substitution is underestimated.

Assumption 2.3 *The stock of past consumption is independent of the quantity purchased.*

Assumption 2.3 relates to a previous discussion since the descriptive analysis of the data has shown that the consumer interpurchase duration is not affected by the quantity purchased. Even though consumers respond to price promotions by increasing their purchased volume, the effect on the duration to the next purchase is not significantly different from zero.

A consequence of Assumptions 2.1 to 2.3 is that only the time elapsed from the last purchase is relevant to infer the depreciated stock of past consumption. The utility of the outside option can then be specified as:

$$u_{i0t}\left(H_{it}, \varepsilon_{i0t}\right) = \varphi\left(H_{it}\right) + \varepsilon_{i0t}, \tag{2}$$

where H_{it} is the number of days since the last purchase occasion and $\varphi\left(H_{it}\right)$ denotes the function that allows me to infer the (unobserved) stock of past consumption from (observed) past purchase choices.

I now move on to specify the indirect utility from choosing an inside alternative. I assume the utility to individual i in time period t from choosing a product j of size $x > 0$ that belongs to h_{it-1} is:

$$\begin{aligned} u_{ijxt}\left(p_{jxt}, a_{jxt}, h_{it-1}, \varepsilon_{ijxt}\right) &= \bar{u}_{ijxt}\left(p_{jxt}, a_{jxt}, h_{it-1}\right) + \varepsilon_{ijxt} \\ &= \gamma_{ix} + \alpha_i p_{jxt} + \xi_{ijt} + \beta_i a_{jxt} + \lambda_i m_{jxt} + \eta_i y_{ijt-1} + \varepsilon_{ijxt}, \end{aligned} \tag{3}$$

where h_{it-1} indicates the set of products purchased by consumer i in her previous purchase event, γ_{ix} denotes the (dis)utility from making a purchase of size x, which could be interpreted as a carrying cost associated with that particular purchase, p_{jxt} is the price of product j in size x, ξ_{ijt} is consumer i taste for for product j that could be a function of product characteristics (like, for example, size), a_{jxt} denotes a vector of indicator variables that control for other promotional activities, m_{jxt} is an indicator variable that takes the value 1 if product j denotes a multiple-item purchase, and ε_{ijxt} is a random shock to consumer choice. The variable y_{ijt-1} keeps track of the number of products that do not belong to the set h_{it-1} if consumer i purchases product j in purchase event t.

The term η_i accounts for state-dependence effects. A positive η_i implies that consumer i has a taste for variety-seeking, since switching to products different from those included in the h_{it-1} set increases the consumer's utility (see McAlister and Pessemier (1982)). The marketing literature provides several explanations for such variety seeking behaviour. Consumers may have an internal desire for change due to satiation or need for stimulation, or they may be balancing the different tastes within the household (see Kahn, 1995, for a comprehensive review of the variety seeking literature). A negative η_i, on the other hand, implies consumer i incurs in a switching cost, since switching to products that do not belong to the set h_{it-1} decreases the consumer's utility (see Pollack (1970) and Spinnewyn (1981)). Klemperer (1995) provides a number of possible reasons for switching costs. Consumers may have shopping search costs and, therefore, do not reoptimize the set of products purchased at every purchase occasion, or they may keep repurchasing the same product as part of a learning process. I do not attempt to distinguish here between these alternative explanations. Rather, I focus on whether state dependence in fact exists and can be identified from observed purchasing behaviour. This approach is similar to Osborne (2007).

2.2.3 Consumer Dynamic Optimization Problem

Consumers in each period decide if or not to purchase, and in case they opt to purchase, which product or products to choose. I make the following assumptions about how consumer expectations of the future affect current period decisions.

Assumption 2.4 *Consumers are forward-looking with regard to their purchase decisions, but myopic with respect to their product choices.*

The myopic assumption implies that consumers maximize their per-period expected utility when making their product choices and is motivated by solely pragmatism. A forward-looking consumer, who experiences state-dependence in her choices of product, considers the future consequences of those choices. The state space of the dynamic problem without this assumption would be extremely large, making the structural estimation computationally infeasible. The development of a framework that incorporates such forward-looking behaviour into a feasible computational estimation procedure is a very interesting potential area for future research. That said, the myopic assumption seems a reasonable assumption about consumer formation of expectations with regard to product choice. While some consumers may plan the whole sequence of product decisions accounting for the consequence of state-dependence in future periods, I tend to believe such forward-looking behaviour to be rare. I should note, however, that Assumption 2.4 does not imply that dynamics are absent from product choice. As I discuss below, current product choices impact the expected future flow utility of the different inside alternatives and, as a consequence, influence the purchase size decision. In other words, even though consumers are myopic with regard to product choice, their decisions have dynamic implications for current and future purchase size choices.

Assumption 2.4 implies a multi-stage budgeting approach to model the purchase and product decisions of consumers (see Aguirregabiria (2002) and Hendel and Nevo (2006a) for similar dynamic applications of Gorman (1971)'s approach). The consumer's expected discounted utility in purchase occasion t can therefore be represented as:

$$\begin{aligned} V(s_{it}) &= \max_{\Pi_i} \sum_{\tau=t}^{\infty} \delta^{\tau-t} E\Big[\sum\nolimits_{x>0} \max_{\Phi_i} \sum\nolimits_j d_{ix\tau} d_{ij/x\tau} u_{ijx\tau}(p_{jx\tau}, a_{jx\tau}, h_{i\tau-1}, \varepsilon_{ijx\tau}) \\ &\quad + d_{i0\tau} u_{i0\tau}(H_{i\tau}, \varepsilon_{i0\tau}) \,|s_{it}, \Pi_i, \Phi_i \Big], \end{aligned} \tag{4}$$

where s_{it} denotes the state at time t and $\delta > 0$ the discount factor. The state s_{it} in each period consists of the vector of current prices and promotional activities for all products and sizes, the set of of products purchased by consumer i in her previous purchase event, the stock of past consumption as measured by the time since the last purchase, and the vector of random shocks to consumer choices, $s_{it} \equiv (p_t, a_t, h_{it-1}, H_{it}, \varepsilon_{it})$. For convenience, I define also the state space s_t^* that consists only of the vector of current prices and promotional activities for all products and sizes, and the set of products purchased by consumer i in her previous purchase event, $s_{it}^* \equiv (p_t, a_t, h_{it-1})$.

Π_i and Φ_i denote a set of decision rules mapping states, s_{it}, to choices, d_{ixt} and $d_{ij/xt}$, respectively, where d_{ixt} is an indicator variable equal to 1 if the choice of consumer i is a purchase of size x (with $x = 0$ standing for no purchase) and $d_{ij/xt}$ is an indicator variable equal to 1 if consumer i chooses to buy product j when purchasing size x. The product of the two indicator variables, $d_{ijxt} = d_{ixt} d_{ij/xt}$, denotes the purchase of product j and size x. I assume that $\sum_{x,j} d_{ijxt} = 1$.

At every state, s_{it}, the consumer faces the same infinite-horizon maximization problem. The value function $V(s_{it})$ defined in equation (4) above is, therefore, the solution to the following Bellman's equation:

$$\begin{aligned} V(s_{it}) &= \max_{d_{ixt}} \Big\{ \sum\nolimits_{x>0} \max_{d_{ij/xt}} \sum\nolimits_j d_{ixt} d_{ij/xt} u_{ijxt}(p_{jxt}, a_{jxt}, h_{it-1}, \varepsilon_{ijxt}) \\ &\quad + d_{i0t} u_{i0t}(H_{it}, \varepsilon_{i0t}) + \delta E\left[V(s_{it+1}) \,|s_{it}, d_{ixt}, d_{ij/xt} \right] \Big\}. \end{aligned} \tag{5}$$

In order to complete the specification of the demand model, I make the following assumptions about the beliefs of consumers regarding the uncertain future prices (and promotional activities) and future utility random shocks.

Assumption 2.5 *Consumers have rational expectations.*

The assumption of rational expectations implies that consumers take all available information into account in forming expectations. Though expectations may turn out incorrect, they will not be systematically wrong. In particular, Assumption 2.5 implies that consumers know both the true transition probability of prices and promotional activities, and the true distribution of the utility random shocks.

Assumption 2.6 *The transition probability of prices and promotional activities are exogenous from the point of view of consumers. Furthermore, they follow a first-order Markov process.*

Assumption 2.6 is consistent with the view that retailers inter-temporal price discriminate by playing mixed strategies that are exogenous from the point of view of consumers (Conslik *et al.* (1984), Sobel (1984), Varian (1980), Pesendorfer (2002)). This assumption implies that, conditional on the control variables, price and promotional activities are independent of the unobserved random shocks, which might be unreasonable if consumers stockpile and inventories are not accounted for. If prices are persistent over time and consumers anticipate purchases in order to hold inventories for future consumption, then unobserved inventories will be correlated with current prices causing an endogeneity problem. Another concern with this assumption might be seasonality. If the likelihood of a temporary promotion is affected by seasonality and it is not accounted for into the transition probability, then unobserved random shocks will be correlated with current prices causing (again) an endogeneity problem. However, as discussed in the previous section, both issues are probably not a concern here.

The first-order Markov process assumption reduces the state space and, although probably inconsistent with equilibrium prices, it is not unreasonable with regard with observed consumers' memory and formation of expectations. The assumption can be relaxed to allow higher order processes, with an increase in the associated computational burden.

Assumption 2.7 ε_{ix_jt} *is independently and identically distributed extreme value type 1.*

Assumption 2.7 is motivated by pragmatism as it significantly reduces the computational burden. The main concern with this type of assumption might be to preclude correlation between products. This is not probably a concern here since the model accounts for product heterogeneity and product state-dependence. Incorporating correlation between the unobserved random shocks of different products can, in principle, be allowed, but at a significant increase in the computational costs of the estimation procedure.

2.3 Maximum Likelihood Estimation

This section presents the estimation details. I estimate the parameters of the model via maximum likelihood. The standard approach would begin by specifying the probability of observing consumer i's choices at time t, which is given by the following likelihood function:

$$l_{it}\left(d_{it}|s_{it}\right) = \prod\nolimits_{x,j} \left[\Pr\left(d_{ijxt} = 1\right)|s_{it}\right]^{d_{ijxt}}, \tag{6}$$

where $d_{it} \equiv \{d_{ijxt}\}$ denotes the vector of her choices. The likelihood of consumer i choices across all time periods would then be:

$$L_i\left(s_{i1}, \ldots, s_{iT}, d_{i1}, \ldots, d_{iT}|s_{i0}, d_{i0}\right) = \int \prod\nolimits_t l_{it}\left(d_{it}|s_{it}\right) dF\left(s_{it}|s_{it-1}, d_{it-1}\right), \tag{7}$$

where s_{i0} and d_{i0} denote the initial conditions, which are observed, and $F\left(s_{it}|s_{it-1}, d_{it-1}\right)$ is the transition probability.

The problem with the standard approach relates to the computation of $\Pr\left(d_{ijxt} = 1\right)$, the probability of observing consumer i purchasing product j and size x in period t, due to the dimensionality of the state space. In order to understand why this is the case, note that given the extreme value assumption on the unobserved utility random shocks (Assumption 2.7) this probability can be defined as:

$$\Pr\left(d_{ijxt} = 1|s_{it}\right) = \frac{\exp\left\{\bar{u}_{ijxt}\left(p_{jxt}, a_{jxt}, h_{it-1}\right) + \delta E\left[V\left(s_{it+1}\right)|s_{it}, d_{ixt}, d_{ij/xt}\right]\right\}}{\sum_{y,k} \exp\left\{\bar{u}_{ikyt}\left(p_{kyt}, a_{kyt}, h_{it-1}\right) + \delta E\left[V\left(s_{it+1}\right)|s_{it}, d_{ixt}, d_{ij/xt}\right]\right\}}, \tag{8}$$

where the summation is over all products from all sizes. The state space includes the vector of current prices and promotional activities for all products and sizes, the set of products purchased by consumer i in her previous purchase event, the stock of past consumption as measured by the time since the last purchase, and the vector of random shocks to consumer choices. Given the multitude of products and sizes available to consumers, the state space is extremely large for practical estimation of $\Pr\left(d_{ijxt} = 1|s_{it}\right)$.

In order to simplify the estimation procedure, I propose a three-stage budgeting approach in the lines of Aguirregabiria (2002) and Hendel and Nevo (2006a).

Step 1 *Estimation of Product Preferences*

I begin by noting that $\Pr\left(d_{ijxt} = 1|s_{it}\right)$ can, in full generality, be decomposed into the product of two components: the probability of choosing product j conditional on the size x

purchased *and* the probability of choosing a purchase of size x:

$$\Pr\left(d_{ijxt}=1|s_{it}\right)=\Pr\left(d_{ij/xt}=1|s_{it},d_{ixt}\right)\Pr\left(d_{ixt}=1|s_{it}\right). \tag{9}$$

The myopia of consumers with regard to product choice (Assumption 2.4), implies that consumers maximize their per-period expected utility when making their product decisions. As a consequence, $\Pr\left(d_{ij/xt}=1|s_{it},d_{ixt}\right)$ can be computed without solving the full dynamic problem. Furthermore, given the extreme value assumption on the utility shocks (Assumption 2.7), that probability can be defined as:

$$\begin{aligned}\Pr\left(d_{ij/xt}=1|s_{it},d_{ixt}\right) &= \frac{\exp\left[\bar{u}_{ijxt}\left(p_{jxt},a_{jxt},h_{it-1}\right)\right]}{\sum_k \exp\left[\bar{u}_{ikxt}\left(p_{kxt},a_{kxt},h_{it-1}\right)\right]} \\ &= \frac{\exp\left(\gamma_{ix}+\alpha_i p_{jxt}+\xi_{ijt}+\beta_i a_{jxt}+\lambda_i m_{jxt}+\eta_i y_{ijt-1}\right)}{\sum_k \exp\left(\gamma_{ix}+\alpha_i p_{kxt}+\xi_{ikt}+\beta_i a_{kxt}+\lambda_i m_{kxt}+\eta_i y_{ikt-1}\right)} \\ &= \Pr\left(d_{ij/xt}=1|p_{xt},a_{xt},h_{it-1}\right),\end{aligned} \tag{10}$$

where the summation is now only over the products of size x.

The parameters in $\bar{u}_{ijxt}\left(p_{jxt},a_{jxt},h_{it-1}\right)$ - with the exception of γ_{ix} that cancels out - can therefore be recovered by maximizing the likelihood of consumer product choice conditional on the size purchased. Let $L_i^{step1}\left(h_{i1},\ldots,h_{iT-1},d_{ij/x1},\ldots,d_{ij/xT}|h_{i0}\right)$ denote the likelihood of consumer i's conditional choices across all time periods:

$$L_i^{step1}\left(h_{i1},\ldots,h_{iT-1},d_{ij/x1},\ldots,d_{ij/xT}|h_{i0}\right)=\prod_t\prod_{j/x}\left[\Pr\left(d_{ij/xt}=1|p_{xt},a_{xt},h_{it-1}\right)\right]^{d_{ij/xt}}, \tag{11}$$

where h_{i0} denotes the initial set of products purchased by consumer i, which is observed. Taking the product of this likelihood function across consumers yields the likelihood function to be maximized in step 1:

$$L=\prod_i L_i^{step1}\left(h_{i1},\ldots,h_{iT-1},d_{ij/x1},\ldots,d_{ij/xT}|h_{i0}\right). \tag{12}$$

In making the utility of choosing a given product state-dependent from the set of products bought in the previous purchase occasion, h_{it-1}, I introduce an identification problem since unobserved consumer heterogeneity may confound the inference of true state-dependence effects. As Heckman (1981) points out, if households have different preferences "and if these differences are not properly controlled, previous experience may appear to be a determinant (...) of future experience solely because it is a proxy for temporally persistent unobservables that determine choices."

State-dependence is usually identified by testing the null hypothesis that the current choice, after accounting for consumer-level heterogeneity, is independent of the previous choice. One approach to introduce heterogeneity is to include *observed* consumer heterogeneity. This approach assumes the existence of a finite number of types or segments, with each type consisting of a set of consumers with identical overall choice preferences (Kamakura and Russell (1989)). As the number of types assumed increases, so will the degree of heterogeneity accounted for under this approach. Goldfarb (2006) presents the extreme case where the number of types exactly coincides with the number of consumers. He makes use of a rich dataset containing nearly 1,000 observations per household to estimate a fully flexible model of consumer preferences, by allowing for consumer-specific regressions.

Another approach is to introduce heterogeneity by considering consumer preferences to be realizations of *random variables.* These random variables are assumed in the literature to follow a multitude of distributional assumptions. For example, Chintagunta *et al.* (1991), Gonul and Srinivasan (1993) and Keane (1997) consider preferences to follow a continuous probability distribution, while Jain *et al.* (1994) consider a discrete probability distribution approximation. An intermediate assumption is presented by Dubé *et al.* (2006) that considers a flexible semi-parametric, but continuous model of consumer heterogeneity.

The estimation procedure in step 1 can not allow for random effects in the lines of the latter approach. If consumer preferences are assumed to be realizations of random variables that follow a probability distribution (either parametric or semi-parametric), then computing $\Pr\left(d_{ij/xt} = 1 | p_{xt}, a_{xt}, h_{it-1}\right)$ requires integration over the assumed distribution. Although conditional on the type of consumer, this probability will still be independent of the dynamic purchase decision, computing this probability unconditional on the type of consumer requires integration over the distribution of types conditional on the size bought. And working out this distribution requires solving the dynamic problem.

Consumer-level heterogeneity can, however, be allowed in the lines of the former approach: either by using observable household demographics to segment consumers into types or, in the lines of Goldfarb (2006) and Hendel and Nevo (2006a), by considering household-level product and state-dependence fixed effects. One concern with the latter solution might be the dimensionality of the parameters to estimate. However, since the likelihood function in equation (12) is well behaved, the estimation of a considerable number of consumer-level fixed effects is feasible and involves very slight increases in computational costs. Furthermore, the consumer-product fixed effects need only to include those products that belong to each consumer shopping history. Consumer-level product preferences can not be estimated for products never purchased by the household. This reduces the number of consumer-product fixed effects substantially since each household typically purchases a relatively small number of products when compared with the full supermarket assortment. Another concern might be the standard incidental parameters problem. However, given the large number of consumer shopping trips in the typical scanner panel datasets, this issue is probably not a concern and therefore assuming T grows asymptotically is not unreasonable.

Step 2 *Estimation of the Inclusive Values Transition Process*

Having outlined the procedure to estimate the probability of choosing product j conditional on the size x purchased, I now move on to specify the two remaining steps required to estimate the probability of choosing a purchase of size x.

The consumer decision with regard to purchase size (whether and what quantity to purchase) is the solution to the dynamic problem characterized by Bellman's equation (5). However, instead of solving this problem, I follow Hendel and Nevo (2006a) and consider a simplification of the state space that makes use of the extreme value assumption on the utility shocks (Assumption 2.7). This simplification involves summarizing a *subset* of the consumer state space, $s_{it}^{*} \equiv (p_t, a_t, h_{it-1})$, into a single index per size, an index representing the utility expected by the consumer, *before* seeing the realization of the utility shocks, from all products of each size. Under Assumption 2.7, this expected utility is given by the inclusive value w_{ixt} (McFadden (1981a)):

$$w_{ixt} = \log\left[\sum\nolimits_{k} \exp\left(\alpha_i p_{kxt} + \xi_{ikt} + \beta_i a_{kxt} + \lambda_i m_{kxt} + \eta_i y_{ikt-1}\right)\right], \tag{13}$$

which can be computed with the parameter estimates from step 1.

In order to show that the original dynamic problem can be written in terms of the simplified state space, I make the following additional assumption, where w_{it} denotes the vector of inclusive values at time t:

Assumption 2.8 $F\left(w_{it}|s_{t-1}^{*}\right)$ can be summarized by $F\left(w_{it}|w_{it-1}\right)$.

Solving the consumer dynamic programming decision requires solving the associated Bellman's equation, which in turn involves working out the expectation of the value function. In order to compute such expectation, I need to specify the transition probabilities for the different state variables. Assumption 2.8 simplifies these processes. The motivation is twofold.

First, the transition probabilities of prices and promotional activities from a multitude of different products of the same size are summarized into the transition probability of a single index. Second, it also simplifies the transition probabilities of product state-dependence. Although consumer product choice is, by Assumption 2.4, myopic (which means that current product choices do not impact future product choices), it does not mean that dynamics are absent. Current product choices impact the expected future flow utility of the different inside alternatives and, as a consequence, impact the expected future inclusive values that influence the purchase size decision. In other words, even though consumers are myopic with regard to product choice, their decisions have dynamic implications for current and future purchase size choices. Assumption 2.8 summarizes the transition probabilities regarding product choice into the inclusive values processes. Because product-choices are consumer-specific, the inclusive values and their transition processes will necessarily be consumer-specific, requiring that the Bellman's equation is solved separately for each consumer

One concern with Assumption 2.8 might be that shopping trips involving different prices, promotional activities and/or previous period product choices can be reflected in a same inclusive value, which in turn yields the same future transition probabilities. This restriction can, to some extent, be relaxed, although at a substantial computational cost.

Step 3 *Estimation of the Intertemporal Effects of Consumption*

Step 3 addresses the computation of the probability of choosing a purchase of a given size x. Not by solving the dynamic problem characterized by Bellman's equation (5), but by solving a simplified problem, where the subset of the consumer state space, s_{it}^{*}, is summarized into the vector of each single size indexes, w_{it}. In this simplified problem, the consumer observes only H_{it} and w_{it} and decides whether and how much to purchase.

I now move on to specify the details of this simpler problem. The utility of consumer i in time period t is given by:

$$\begin{aligned} u_{i0t}^{step3}\left(H_{it}, \varepsilon_{i0t}\right) &= \varphi\left(H_{it}\right) + \varepsilon_{i0t}, && \text{if } x = 0 \\ u_{ixt}^{step3}\left(w_{ixt}, \varepsilon_{ixt}\right) &= \gamma_{ix} + w_{ixt} + \varepsilon_{ixt}, && \text{if } x > 0, \end{aligned} \tag{14}$$

where, as before, $x = 0$ stands for no purchase. The consumer is assumed to be forward-looking and, therefore, to maximize the expected discounted utility:

$$V_{step3}\left(H_{it}, w_{it}, \varepsilon_{it}\right) = \max_{\Pi_i^{step3}} \sum_{\tau=t}^{\infty} \delta^{\tau-t} E\left[\sum\nolimits_{x>0} d_{ix\tau} u_{ix\tau}^{step3}\left(w_{ix\tau}, \varepsilon_{ix\tau}\right) + d_{i0\tau} u_{i0\tau}^{step3}\left(H_{i\tau}, \varepsilon_{i0\tau}\right) | H_{it}, w_{it}, \varepsilon_{it}, \Pi_i^{step3}\right], \tag{15}$$

where Π_i^{step3} denotes a set of decision rules mapping states to choices, d_{ixt}. The Bellman's equation associated with the consumer's simpler dynamic problem is given by:

$$V_{step3}\left(H_{it}, w_{it}, \varepsilon_{it}\right) = \max_{d_{ixt}} \left\{\sum\nolimits_{x>0} d_{ixt} u_{ixt}^{step3}\left(w_{ixt}, \varepsilon_{ixt}\right) + d_{i0t} u_{i0t}^{step3}\left(H_{it}, \varepsilon_{i0t}\right) + \delta E\left[V_{step3}\left(H_{it+1}, w_{it+1}, \varepsilon_{it+1}\right) | H_{it}, w_{it}, \varepsilon_{it}, d_{ixt}\right]\right\}. \tag{16}$$

It remains to be shown that the probability of purchasing size x computed from the simplified problem is equivalent to the one computed from the original problem. Establishing this equivalence involves two steps. In the first step, I show that the Bellman's equations associated with the original and simplified problems have the same solution. The second step involves actually showing the equivalence of the probability of purchasing size x from the two problems.

The two-step proof adapts the one presented in Hendel and Nevo (2006a) to this variety-seeking framework.

Proposition 2.1 *The Bellman's equations associated with the original and simplified problems have the same solution.*

Proof. I begin by addressing the original dynamic problem. The Bellman's equation associated with this problem is given in equation (5), reproduced here for convenience:

$$V\left(s_{it}\right) = \max_{d_{ixt}} \left\{\sum\nolimits_{x>0} \max_{d_{ij/xt}} \sum\nolimits_{j} d_{ixt} d_{ij/xt} u_{ijxt}\left(p_{jxt}, a_{jxt}, h_{it-1}, \varepsilon_{ijxt}\right) + d_{i0t} u_{i0t}\left(H_{it}, \varepsilon_{i0t}\right) + \delta E\left[V\left(s_{it+1}\right) | s_{it}, d_{ixt}, d_{ij/xt}\right]\right\}.$$

Given Assumption 2.7, the expected value of $V\left(s_{it+1} | s_{it}, d_{ixt}, d_{ij/xt}\right)$ will be a function of H_{it}, s_{it}^*, d_{ixt} and $d_{ij/xt}$. Recall that s_{it}^* denotes the state space that consists only of the vector of current prices and promotional activities for all products and sizes, and the set of products purchased by consumer i in her previous purchase event. Let $V^e\left(H_{it}, s_{it}^*\right)$ denote such function to simplify notation: $V^e\left(H_{it}, s_{it}^*\right) = E\left[V\left(s_{it+1}\right) | s_{it}, d_{ixt}, d_{ij/xt}\right]$.

Computing the expected value of $V(s_{it})$ conditional on the information available at time $t-1$ yields:

$$\begin{aligned} V^e\left(H_{it-1}, s^*_{it-1}\right) &= \int \left\{ \max_{d_{ixt}} \left\{ \sum\nolimits_{x>0} \max_{d_{ij/xt}} \sum\nolimits_j d_{ixt} d_{ij/xt} u_{ijxt}\left(p_{jxt}, a_{jxt}, h_{it-1}, \varepsilon_{ijxt}\right) \right\} \right. \\ &\quad \left. + d_{i0t} u_{i0t}\left(H_{it}, \varepsilon_{i0t}\right) + \delta V^e\left(H_{it}, s^*_{it}\right) \right\} dF\left(s_{it} | s_{it-1}, d_{it-1}\right). \end{aligned}$$

The myopic assumption with regard to consumer product choice together with the extreme-value assumption allows this expected value to be re-written in terms of the inclusive values defined in equation (13):

$$\begin{aligned} V^e\left(H_{it-1}, s^*_{it-1}\right) &= \int \log \left\{ \sum\nolimits_{x>0} \exp\left(\gamma_{ix} + w_{ixt} + \delta V^e\left(H_{it}, s^*_{it}\right)\right) \right. \\ &\quad \left. + \exp\left[\varphi\left(H_{it}\right) + \delta V^e\left(H_{it}, s^*_{it}\right)\right] \right\} dF\left(s^*_{it} | s^*_{it-1}, d_{it-1}\right). \end{aligned}$$

where the expression inside the integral represents integration over the vector of random utility shocks. From the analysis of the above equation, it is possible to conclude that $V^e\left(H_{it-1}, s^*_{it-1}\right)$ can, under Assumption 2.8, be iterated using the following Bellman's equation rewritten in terms of w_t instead of s^*_{it-1}:

$$\begin{aligned} V^e\left(H_{it-1}, w_{it-1}\right) &= \int \log \left\{ \sum\nolimits_{x>0} \exp\left[\gamma_{ix} + w_{ixt} + \delta V^e\left(H_{it}, w_{it}\right)\right] \right. \\ &\quad \left. + \exp\left[\varphi\left(H_{it}\right) + \delta V^e\left(H_{it}, w_{it}\right)\right] \right\} dF\left(w_{it} | w_{it-1}, d_{it-1}\right). \end{aligned}$$

I now address the simplified problem. The Bellman's equation associated with this problem is given in equation (16). After substituting for $u^{step3}_{i0\tau}\left(H_{it}, \varepsilon_{i0t}\right)$ and $u^{step3}_{ixt}\left(w_{ixt}, \varepsilon_{ixt}\right)$ yields:

$$\begin{aligned} V_{step3}\left(H_{it}, w_{it}, \varepsilon_{it}\right) &= \max_{d_{ixt}} \left\{ \sum\nolimits_{x>0} d_{ixt}\left(\gamma_{ix} + w_{ixt} + \varepsilon_{ixt}\right) + d_{i0t}\left[\varphi\left(H_{it}\right) + \varepsilon_{0t}\right] \right. \\ &\quad \left. + \delta E\left[V_{step3}\left(H_{it+1}, w_{it+1}, \varepsilon_{it+1}\right) | H_{it}, w_{it}, \varepsilon_{it}, d_{ixt}\right] \right\}. \end{aligned}$$

Taking expectations given the information available at time $t-1$ and integrating out the utility random shocks making use of the extreme-value assumption (Assumption 2.7) allows me to write the expected value, $V^e_{step3}\left(H_{it-1}, w_{it-1}\right)$, as:

$$\begin{aligned} V^e_{step3}\left(H_{it-1}, w_{it-1}\right) &= \int \log \left\{ \sum\nolimits_{x>0} \exp\left[\gamma_{ix} + w_{ixt} + \delta V^e\left(H_{it}, w_{it}\right)\right] \right. \\ &\quad \left. + \exp\left[\varphi\left(H_{it}\right) + \delta V^e\left(H_{it}, w_{it}\right)\right] \right\} dF\left(w_{it} | w_{it-1}, d_{it-1}\right). \end{aligned}$$

Thus, as the proposition claims, the solution to the Bellman's equations associated with the original and simplified problems is the same. ■

I now address the second step of the proof, showing the equivalence between the probability of purchasing size x computed from the original problem to the one computed from the simplified problem.

Proposition 2.2 $\Pr\left(d_{ixt}=1|H_{it},s_{it}^{*}\right)=\Pr\left(d_{ixt}=1|H_{it},w_{it}\right).$

Proof. The probability of purchasing size x computed from the simplified problem is given by:

$$\Pr\left(d_{ixt}=1|H_{it},w_{it}\right)=\frac{\exp\left[\gamma_{ix}+w_{ixt}+\delta V^{e}\left(H_{it},w_{it}\right)\right]}{M_{i0t}+\sum_{y>0}\exp\left[\gamma_{iy}+w_{iyt}+\delta V^{e}\left(H_{it},w_{it}\right)\right]},$$

where for notational simplicity $M_{i0t}=\exp\left[\varphi\left(H_{it}\right)+\delta V^{e}\left(H_{it},w_{it}\right)\right]$. If, on the other hand, this probability is computed from the original problem, it is given by:

$$\Pr\left(d_{ixt}=1|H_{it},s_{it}^{*}\right)=\frac{\sum_{j}\exp\left\{\bar{u}_{ijxt}\left(s_{it}^{*}\right)+\delta V^{e}\left(H_{it},s_{it}^{*}\right)\right\}}{M_{i0t}^{*}+\sum_{y,k}\exp\left\{\bar{u}_{ikyt}\left(s_{it}^{*}\right)+\delta V^{e}\left(H_{it},s_{it}^{*}\right)\right\}},$$

where $M_{i0t}^{*}=\exp\left[\varphi\left(H_{it}\right)+\delta V^{e}\left(H_{it},s_{it}^{*}\right)\right]$. The summation in the numerator is over all products of size x, and the summation in the denominator is over all products of all sizes. As discussed in Proposition 2.1, $V^{e}\left(H_{it},s_{it}^{*}\right)$ can, under Assumption 2.8, be re-written in terms of w_{it} instead of s_{it}^{*}. This implies that the expected value function depends on the purchase size chosen, but not on the particular product choice. Furthermore, $M_{i0t}=M_{i0t}^{*}$. As a consequence, the above probability can be decomposed and simplified as follows:

$$\begin{aligned}
&\Pr\left(d_{ixt}=1|H_{it},s_{it}^{*}\right)=\\
&=\frac{M_{ixt}\sum_{j}\exp\left(\alpha_{i}p_{jxt}+\xi_{ijt}+\beta_{i}a_{jxt}+\lambda_{i}m_{jxt}+\eta_{i}y_{ijt-1}\right)}{M_{i0t}+\sum_{y>0}M_{iyt}\sum_{k}\exp\left(\alpha_{i}p_{kyt}+\xi_{ikt}+\beta_{i}a_{kyt}+\lambda_{i}m_{kyt}+\eta_{i}y_{ikt-1}^{e}\right)}\\
&=\frac{M_{ixt}\exp\left\{\log\left[\sum_{j}\exp\left(\alpha_{i}p_{jxt}+\xi_{ijt}+\beta_{i}a_{jxt}+\lambda_{i}m_{jxt}+\eta_{i}y_{ijt-1}\right)\right]\right\}}{M_{i0t}+\sum_{y>0}M_{iyt}\exp\left\{\log\left[\sum_{k}\exp\left(\alpha_{i}p_{kyt}+\xi_{ikt}+\beta_{i}a_{kyt}+\lambda_{i}m_{kyt}+\eta_{i}y_{ikt-1}^{e}\right)\right]\right\}}\\
&=\frac{\exp\left(w_{ixt}+M_{ixt}\right)}{M_{i0t}+\sum_{y>0}\exp\left(w_{iyt}+M_{iyt}\right)}\\
&=\frac{\exp\left[\gamma_{ix}+w_{ixt}+\delta V^{e}\left(H_{it},w_{it}\right)\right]}{M_{i0t}+\sum_{y>0}\exp\left[\gamma_{iy}+w_{iyt}+\delta V^{e}\left(H_{it},w_{it}\right)\right]}\\
&=\Pr\left(d_{ixt}=1|H_{it},w_{it}\right),
\end{aligned}$$

where $M_{x}=\exp\left[\gamma_{ix}+\delta V^{e}\left(H_{it},w_{it}\right)\right]$.

Thus, as the proposition claims, the probabilities computed from the original and simplified problems are equivalent. ■

Having established the equivalence of the probability of a purchase of size x between the two problems, I move on to specify the estimation procedure. I estimate the remaining parameters by maximizing the likelihood of consumer purchase choices. Let the likelihood of consumer i's purchase choices across all time periods be denoted by:

$$L_i^{step3}\left(H_{i1},\ldots,H_{iT},d_{ix1},\ldots,d_{ixT}|H_{i0}\right)=\prod\nolimits_t\prod\nolimits_x\left[\Pr\left(d_{ixt}=1|H_{it},w_{it}\right)\right]^{d_{ixt}}dF\left(w_{it}|w_{it-1},d_{it-1}\right), \tag{17}$$

where H_{i0} denotes the initial stock of past consumption as measured by the time since the initial purchase.

In making the utility of purchasing a given size state-dependent from the duration since the last purchase, I introduce an identification problem similar to the one discussed previously for product switching: unobserved consumer heterogeneity may confound the inference of true state-dependence effects. The identification problem arises from the fact the interpurchase duration may be long either due to a low taste for the good or a strong state-dependence effect. In order to control for unobserved heterogeneity, I assume preferences vary across consumers using a random effects specification:

$$\lambda_i = \lambda + \Gamma v_i, \tag{18}$$

where λ_i denotes the vector of the remaining parameters to be estimated (and includes the duration dependence and the size-specific (dis)utilities parameters) and v_i is a independently and identically distributed standard normal. The vector λ denotes the mean values of the different coefficients, while Γ denotes the Cholesky decomposition of the variance-covariance matrix, Σ, which for computational simplicity is assumed diagonal.

I should note that although one may argue that consumer-specific inclusive-values already control for unobserved consumer heterogeneity in their taste for the good, a random effects specification for the size-specific (dis)utilities is required in practice. The justification relates to an unfortunate property of the conditional logit model used in step 1. The model is unable to estimate an intercept since it plays no role in determining the product-choice probability conditional on the size purchased. As a consequence, in order to estimate the consumer-level product preferences in step 1, a normalization is required for each consumer and size (since including dummy variables for all products in the conditional choice would amount to estimate a size-specific intercept). The random effects specification for the size-specific (dis)utilities in step 3 is instrumental in making the inclusive values across sizes and consumers, each estimated using a different normalization, comparable.

With the introduction of the random effects, the likelihood of consumer i's purchase choices across all time periods is now given by:

$$L_i^{step3}\left(H_i, d_{ix}|H_{i0}\right) = \int \prod_t \prod_x \left[\Pr\left(d_{ixt}=1|H_{it}, w_{it}, v_i\right)\right]^{d_{ixt}} dF\left(v_i\right) dF\left(w_{it}|w_{it-1}, d_{it-1}\right). \tag{19}$$

I follow Pakes (1986), Pakes and Pollard (1989), and McFadden (1989) and draw ns pseudo-random consumers to approximate the integral using a (smooth) simulator estimator:

$$L_i^{step3}\left(H_i, d_{ix}|H_{i0}\right) = \int \prod_t \prod_x \left[\frac{1}{ns}\sum_s \Pr\left(d_{ixt}=1|H_{it}, w_{it}, v_s\right)\right]^{d_{ixt}} dF\left(v_i\right) dF\left(w_{it}|w_{it-1}, d_{it-1}\right). \tag{20}$$

Taking the product of this likelihood function across consumers yields the likelihood function to be maximized in step 3:

$$L = \prod_i L_i^{step3}\left(H_i, d_{ix}|H_{i0}\right). \tag{21}$$

2.3.1 Bellman's Equation Solution

The structural estimation is based on Rust (1987)' s algorithm that nests the solution of the consumer's dynamic programming within the estimation parameter search. In this section, I address the computational details of the strategy used to solve the functional equation (16) associated with the consumer's simpler dynamic problem. One strategy to solve dynamic programming problems is by *discrete approximation*. In this type of approach, the value function is solved for numerically by discretizing continuous state spaces into a finite number of n grid points. However, in high-dimensional problems, discretization results in a curse of dimensionality, since n increases exponentially fast in the dimension of the state space. Another approach is to solve dynamic programming problems by *parametric approximation*, where the value function is approximated by a smooth parametric function with k unknown parameters. The latter approach is superior to the former whenever the number of parameters k required to obtain a good global approximation (according to some metric) under parametric approximation is smaller than the value n of grid points required to obtain a comparable fit by discrete approximation.

I follow Hendel and Nevo (2006a) and solve the functional equation (16) by using value function parametric approximation with policy function iteration in the lines of Benitez-

Silva *et al.* (2000). Policy function iteration consists of an alternating sequence of *policy improvement* and *policy valuation* steps:

Policy Valuation

The policy valuation step computes the value function, $V_{step3}(H_{it}, w_{it}, \varepsilon_{it})$, for a given initial guess of the consumer decision, d_{it}. Under a parametric approximation approach, the value function is approximated by a linear combination of k basis functions $(\rho_1, \ldots, \rho_k)$:

$$V_{step3}(H_{it}, w_{it}, \varepsilon_{it}) \simeq \sum_k \theta_k \rho_k (H_{it}, w_{it}, \varepsilon_{it}). \tag{22}$$

Substituting $V_{step3}(H_{it}, w_{it}, \varepsilon_{it})$ in functional equation (16) by the polynomial approximation yields a linear equation in k unknown parameters θ:

$$\begin{aligned} \sum_k \theta_k \rho_k (H_{it}, w_{it}, \varepsilon_{it}) &= \sum_x d_{ixt} u_{ixt}^{step3} (w_{ixt}, \varepsilon_{ixt}) + d_{i0t} u_{i0t}^{step3} (H_{it}, \varepsilon_{i0t}) \\ &\quad + \delta \int \sum_k \theta_k \rho_k (H_{it+1}, w_{it+1}, \varepsilon_{it+1})\, dF(\varepsilon_{it+1})\, dF(w_{it+1}|w_{it}, d_{it}). \end{aligned} \tag{23}$$

This can be solved by ordinary least squares when evaluated at a finite set of $m \geq k$ sample points in the state space $(H_{it}, w_{it}, \varepsilon_{it})$. In order to understand why this is the case, define the $(m \times k)$ matrices r and Er, as well as the $(m \times 1)$ vector u with the following elements:

$$\begin{aligned} r_{mk} &= \rho_k (H_m, w_m, \varepsilon_m) \\ Er_{mk} &= \int \rho_k (H_{m+1}, w_{m+1}, \varepsilon_{m+1})\, dF(\varepsilon_{m+1})\, dF(w_{m+1}|w_m, d_m) \\ u_m &= \sum_x d_{xm} u_{xm}^{step3} (w_{xm}, \varepsilon_{xm}) + d_{0m} u_{0m}^{step3} (H_m, \varepsilon_{0m}). \end{aligned} \tag{24}$$

Equation (23) can then be re-written as a system of linear equations: $u = X\theta$, where $X = (r - \delta Er)$. The solution to this system of equations, which is given by $\hat{\theta} = (X'X)^{-1} X'u$, can then be used to evaluate the approximated value function.

Policy Improvement

The policy improvement step updates the guess of the consumer decision, d_{it}, using the value function approximation from the policy valuation step. The updated consumer decision (purchase size) can be performed analytically by maximizing the sum, evaluated at the same

m sample points, of current utility and the expected discounted future utility:

$$\begin{aligned} d_{xm} &= \arg\max \left\{ \sum\nolimits_{x>0} d_{xm} u_{xm}^{step3} \left(w_{xm}, \varepsilon_{xm} \right) + d_{0m} u_{0m}^{step3} \left(H_m, \varepsilon_{0m} \right) \right. \qquad (25) \\ &\left. + \delta \int \sum\nolimits_k \hat{\theta}_k \rho_k \left(H_{m+1}, w_{m+1}, \varepsilon_{m+1} \right) dF \left(\varepsilon_{m+1} \right) dF \left(w_{m+1} | w_m, d_m \right) \right\}. \end{aligned}$$

The two steps are then iterated until convergence of the parameters of the value function approximation. The consumer decision that it converges to, and the corresponding value functions are approximated solutions to the Bellman's equation. See Puterman and Shin (1978) for sufficient conditions for policy iteration to converge in continuous state spaces.

2.3.2 Identification

In this section, I provide an informal discussion of identification. I begin by addressing the identification of step 1 parameters. The identification of the non-dynamic product preference parameters is standard, with the coefficients being identified through the effect of *current* period's variation in those exogenous variables on *current* period's probability of choosing a given product. Temporary price and non-price promotions provide variation to identify sensitiveness to price and other promotional activities. The (dis)utility from multiple-item purchasing is identified by the share of multiple-item purchases across trips. Consumer-level product effects are identified from variations in consumer shares across products.

Product choice state-dependence is identified, as argued by Chamberlain (1985), through the effect of *previous* period's variation in exogenous variables on *current* period's probability of choosing a given product. If a temporary promotion for product j at time $t-1$ decreases the probability of a given consumer choosing product j at time t, then the consumer may be a variety-seeker. If, on the other hand, such promotion increases that probability, the consumer may be incurring in switching costs. Given a long enough consumer-level price (and other promotional activities) time series, variation in previous period's promotions identifies product-choice state-dependence.

Step 2 parameters are identified through the effect of previous period's variation in each consumer inclusive values on her current period's inclusive values.

I now move on to address the identification of step 3 parameters. The purchase size coefficients help fit the infrequent incidence of purchase across observed trips and are naturally identified from each consumer's propensity to purchase the different sizes. The intertemporal effect of purchasing on the utility of the outside alternative is identified by each consumer's interpurchase duration in days. Because in a discrete choice demand model only the relative utilities are identifiable, an identifying normalization is required. I normalize the utility of the outside option to zero when $H_{it} \leq 1$. Finally, I note that in this frequency of purchase model, the discount factor is not identifiable. I assume it to equal 0.995.

2.4 Empirical Analysis

2.4.1 Step 1: Estimation of Product Preferences

Step 1 estimates product preferences by maximizing the likelihood of observing the sequence of household product choices, conditional on the size purchased. The choice set includes, therefore, only products of the same size as the actual purchase. Table 2.7 presents the results of this analysis, with the different columns reporting distinct specifications that vary on the covariates included. Specification (1) includes as explanatory variables price and a multiple-item purchase dummy variable. The price coefficient is of the expected sign and statistically significant suggesting that the average household is price sensitive. The multiple-item coefficient is not statistically different from zero which seems to indicate that consumers product choice pattern when purchasing a single-item does not significantly differ from when they purchase multiple-items. Specification (2) controls for promotional activities by including feature and display dummy variables as additional covariates. The coefficients on these controls are positive and statistically significant suggesting that consumers do respond to promotional activities. However, the comparison of the price coefficient in the two specifications is suggestive of an endogeneity issue. Prices are negatively correlated with promotional activities since promoted products sell at lower prices and, as a consequence, not including these controls will overestimate consumer price sensitiveness. In specification (3), I include product dummy variables in order to control for market-level unobserved product characteristics. The product dummy variables are interacted with size so that the preference for each specific product is proportional to the package size purchased. The effect of including these controls on the price coefficient is again suggestive of an endogeneity issue. Products with higher unobserved characteristics sell at higher prices inducing a positive correlation that will underestimate consumer price sensitiveness if not accounted for.

Specification (4) addresses the question of whether households incur in switching costs or in variety-seeking by including as covariate the number of products that, in each alternative choice, do not belong to the set of products bought in her previous purchase event. The

TABLE 2.7

*Step 1: Estimation of Product Preferences**

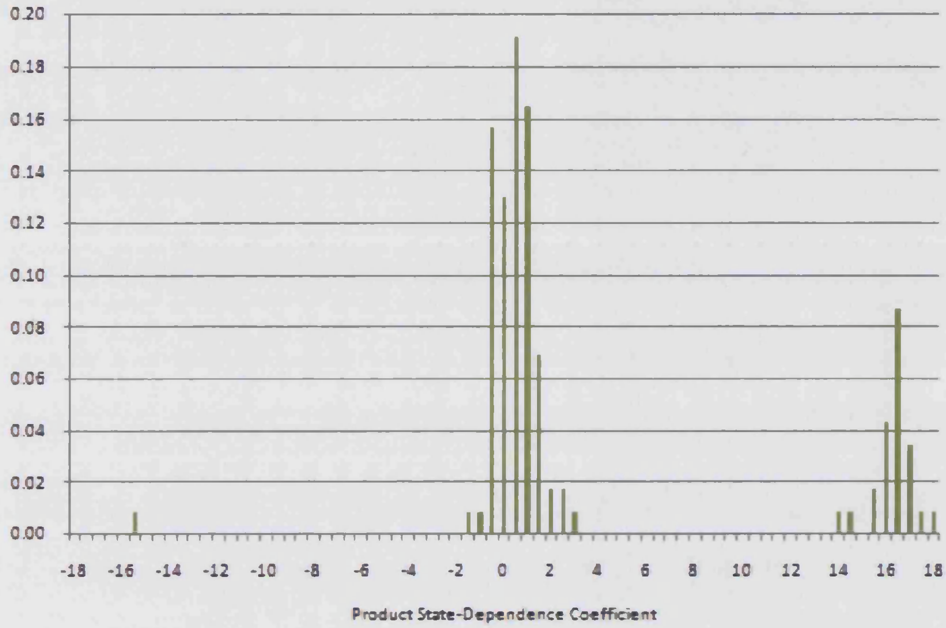
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Price	-0.67	-0.56	-0.67	-0.66	-0.54	-0.54	-0.54
	(0.06)	(0.07)	(0.08)	(0.08)	(0.12)	(0.12)	(0.12)
Price × Single					-0.06	-0.06	-0.07
					(0.16)	(0.16)	(0.16)
Price × Children					0.01	0.01	0.01
					(0.16)	(0.16)	(0.17)
Feature		0.54	0.52	0.54	0.78	0.78	0.79
		(0.13)	(0.13)	(0.12)	(0.15)	(0.15)	(0.16)
Display		0.61	0.63	0.63	0.72	0.73	0.76
		(0.15)	(0.15)	(0.15)	(0.16)	(0.16)	(0.17)
Multiple-Item	-0.47	-0.73	-1.17	-1.06	-3.51	-3.50	-3.94
	(0.40)	(0.45)	(0.57)	(0.56)	(1.00)	(0.99)	(1.18)
Multiple-Item × Single					2.71	2.74	2.88
					(1.30)	(1.30)	(1.55)
Multiple-Item × Children					2.65	2.62	3.01
					(1.20)	(1.21)	(1.39)
Product State-Dependence				-0.39	0.22	0.21	
				(0.08)	(0.09)	(0.11)	
Product State-Dependence × Single						-0.07	
						(0.21)	
Product State-Dependence × Children						0.16	
						(0.18)	
HH Product State-Dependence Dummy Variables							yes
Product Dummy Variables			yes	yes			
HH Product Dummy Variables					yes	yes	yes
Pseudo R2	0.04	0.05	0.15	0.15	0.24	0.24	0.26

* An observation is a purchase instance by a household. Standard errors clustered by households in parentheses.

coefficient is negative and statistically significant suggesting that the average consumer incurs in a cost when switching products in successive purchase occasions. The problem with this specification is that unobserved household heterogeneity will confound the inference of true state-dependence effects. The identification problem arises because a consumer may repeatedly purchase a particular product either because of a strong unobserved, idiosyncratic preference for it or because she dislikes switching. In order to identify true state dependence, I control for household heterogeneity in specification (5). I introduce heterogeneity in two ways. First, I interact price and multiple-item covariates with two observable household demographics: a dummy variable that takes the value 1 if the household is of a single person, and another if children under the age of 18 are present in the household.[6] Second, I introduce household-level product dummy variables. I assume, as before, preference for each specific product to be proportional to the package size purchased. No dimensionality problem arises with this introduction because (*i*) I only consider the products that belong to each household shopping history (I can not expect to estimate household product preferences for products never purchased by the household), (*ii*) each household buys a relatively small number of products, and (*iii*) has a relatively long time sequence of purchases. Most demographic interactions on price are statistically insignificant suggesting observable characteristics are not important in explaining price sensitiveness. Demographic interactions on multiple-item purchases are estimated to be statistically significant and positive. While this result is expected for households with children, it is unexpected and hard to interpret for one person households. Finally, most household-level product dummy variables are statistically significant, with the introduction of such heterogeneity generating substantial changes in the state-dependence coefficient. Households are now estimated to have an average positive taste for variety-seeking. These results seem to indicate that controlling for household heterogeneity matters. Specification (6) and (7) introduce heterogeneity in the variety-seeking/switching cost coefficient. In specification (6), I interact it with observable household demographics, with the interactions being statistically insignificant, while in specification (7) I allow for full household heterogeneity in the coefficient by interacting it with household-level dummy variables. Except for three households, all coefficients are statistically significant. Figure 2.4 plots the coefficient frequency distribution. Most of the households have a taste for variety, but the magnitude is relatively small. Approximately 18% of the consumers actually incurs in a cost when switching products in successive purchase occasions, while approximately 21% are heavily variety-seeking.

[6] I also estimated several specifications that included interactions with household income. Since the results were never significant, I do not consider them here.

FIGURE 2.4
Frequency Distribution for Product State-Dependence Coefficient



2.4.2 Step 2: Estimation of the Inclusive Values Transition Process

Step 2 estimates the transition process for the inclusive values, which were computed for the purchase sizes observed in the data (16 $oz.$, 32 $oz.$, 64 $oz.$, 80 $oz.$, 128 $oz.$ and 160 $oz.$) using step 1's estimates from specification (7) above. I follow Hendel and Nevo (2006a) and assume the following first-order Markov process for the transition probability of the inclusive values:

$$w_{ixt} = \theta_{ix0} + \sum\nolimits_{s \in \Im} \theta_{ixs} w_{ist-1} + \varsigma_{ixt}, \tag{26}$$

where the summation is over the set of package sizes $\Im$ =(16 $oz.$, 32 $oz.$, 64 $oz.$, 160 $oz.$) and ς_{ixt} is distributed normal with mean zero and standard deviation σ_{ix}. Multicollinearity precludes the transition process of being defined over the set of all possible purchase sizes (since consumers that are observed purchasing 80 $oz.$ and 128 $oz.$, do so by buying multiple items: 80 $oz.$ = 16 $oz.$ + 64 $oz.$ while 128 $oz.$ = 64 $oz.$ + 64 $oz.$). Finally, the transition process parameters are index by i because the inclusive values are consumer-specific.

The assumption that the inclusive values are normally distributed may seem somewhat problematic given the evolution of the state variables they summarize. In order to test this assumption, I performed the Shapiro-Wilk W test for normality on the different consumer-level inclusive values. In the untabulated tests, the null hypothesis that the inclusive values

TABLE 2.8

*Step 2: Estimation of the Inclusive Values Transition Process**

	Panel A				Panel B			
	Same Process for All Consumers				Consumer-Level Process			
	ω_{16t-1}	ω_{32t-1}	ω_{64t-1}	ω_{160t-1}	ω_{16t-1}	ω_{32t-1}	ω_{64t-1}	ω_{160t-1}
ω_{16t}	0.97	0.14	-0.05	0.54	0.21	-0.19	-0.20	0.36
	(0.09)	(0.06)	(0.03)	(0.08)	(0.34)	(2.61)	(6.51)	(3.20)
ω_{32t}	0.02	1.08	-0.09	0.15	0.31	0.08	0.08	0.11
	(0.11)	(0.07)	(0.04)	(0.10)	(1.06)	(0.21)	(5.30)	(4.25)
ω_{64}	0.60	-0.27	0.96	0.00	-0.02	-0.01	0.25	0.01
	(0.08)	(0.05)	(0.03)	(0.07)	(0.43)	(0.11)	(0.26)	(0.06)
ω_{80t}	2.21	-0.46	0.48	0.27	0.17	-0.19	-0.05	0.39
	(0.12)	(0.07)	(0.04)	(0.10)	(0.56)	(2.61)	(6.56)	(3.41)
ω_{128t}	0.63	-0.28	1.76	0.16	-0.03	-0.02	0.45	0.02
	(0.08)	(0.05)	(0.03)	(0.07)	(0.43)	(0.20)	(0.47)	(0.13)
ω_{160t}	0.33	-0.10	-0.03	1.24	-0.00	0.00	0.12	0.11
	(0.07)	(0.04)	(0.02)	(0.06)	(0.00)	(0.01)	(5.98)	(0.25)

* An observation is a shoppping trip instance by a household. Also included are a constant and size indicator variables to control for unavailability of a package size at a given shopping trip. Panel A displays point estimates and standard errors in parentheses. Panel B displays the mean and standard deviation across the different consumer estimates.

are distributed normal is accepted only for a small fraction of consumers: 16 *oz.* (3%), 32 *oz.* (3%), 64 *oz.* (17%), 80 *oz.* (3%), 128 *oz.* (13%) and 160 *oz.* (1%). This assumption can be relaxed, although with a substantial increase in the computational burden.

Table 2.8 reports the estimated transition probabilities. Table 2.8, Panel A presents the point estimates (and associated standard errors) under the constraint that all consumers face the same transition probabilities. The results suggest that the lagged inclusive value of own size (or of the two own sizes for those cases that involve multiple-item purchases) is the most important in predicting its future variation. In Table 2.8, Panel B the estimated transition probabilities are consumer-specific, with the results displaying the mean and standard deviation across the different consumer-level estimates. There is evidence of substantial heterogeneity across consumers, as suggested by the large standard deviations, which supports the option for the individual-level transition processes.

2.4.3 Step 3: Estimation of the Intertemporal Effects of Consumption

Step 3 maximizes the likelihood of observing the sequence of consumer purchase choices after solving the consumer-specific Bellman's equations associated with the simplified dynamic programming problem. Even though I solved the Bellman's equation separately for each consumer, the random effects specification for the parameters allowed me to pool the likelihoods across consumers. As discussed previously, I approximated the value function by a linear combination of k basis functions $(\rho_1, \ldots, \rho_k)$, with the approximation basis used being a polynomial in the natural logarithm of the duration in days since the consumer's last purchase and in the levels of the remaining state variables.

In order to estimate the model, I have to specify a functional form for $\varphi(H_{it})$. I assume the following:

$$\varphi(H_{it}) = \kappa_0 \ln(H_{it}) + \kappa_1 H7_{it},$$

where $H7_{it}$ denotes an indicator variable that takes the value 1 if the shopping trip at time t corresponds to the consumer seven days cycle as suggested by the purchase hazard rate.

Table 2.9 reports the results for different specifications of step 3. In specification (1) I do not allow for heterogeneity or forward-looking behaviour. The state-dependence results suggest that the utility of the outside option decreases with the duration since the last purchase, which supports the intertemporal substitution argument. This result should not come as a surprise, despite the opposite suggestion from the raw data (recall the slight downward trend of the hazard rate), because the consumer-specific inclusive values do control in some extent for unobserved heterogeneity. The coefficient on the indicator variable $H7_{it}$ is negative, which suggests that once every 7 days, the value of the outside option decreases. I interpret this result as illustrating potentially reduced transaction costs of consumers purchasing in their main shopping trip. The estimates for the size-specific effects are statistically significant at standard significance levels. Econometrically, they help fit each size frequency of purchase. However, as discussed previously, the magnitude and ordering of these estimates can not be directly interpreted as they capture the different normalizations required for step 1 estimation.

Specification (2) introduces dynamic considerations into the consumers decisions, with this introduction substantively reducing state-dependence. The reason is that the static specification omits price expectations from the consumers purchase decisions. When facing a price promotion, the typical consumer expectation is that the price will go up in the future. This induces her, as I discussed in the descriptive analysis section, to typically take advantage of the price promotion by anticipating purchases. The static specification, by omitting price expectations, bias the results since it interprets this shorter interpurchase durations as stronger state-dependence. The addition of the forward-looking behaviour also impacts the coefficient on the indicator variable, which becomes (significantly) positive. This

TABLE 2.9

*Step 3: Estimation of the Intertemporal Effects of Consumption**

	Static	Dynamic		
	Standard Logit Estimate	Standard Logit Estimate	Random Coefficients Mean Estimate	Random Coefficients Standard Deviation
	(1)	(2)	(3)	
No Purchase (Outside Alternative)				
κ_0	-2.42	-1.63	-1.83	4.57
	(0.00)	(0.00)	(0.02)	(0.03)
κ_1	-2.37	0.43	-0.22	0.26
	(0.02)	(0.02)	(0.19)	(0.15)
Purchase Size				
16 *oz.*	-36.87	-36.84	-38.14	8.76
	(0.02)	(0.02)	(0.34)	(0.18)
32 *oz.*	-40.52	-40.27	-40.56	6.08
	(0.02)	(0.02)	(0.22)	(0.12)
64 *oz.*	-25.56	-25.25	-35.88	13.95
	(0.01)	(0.01)	(0.29)	(0.21)
80 *oz.*	-43.39	-41.57	-41.61	4.29
	(0.20)	(0.04)	(0.27)	(0.43)
128 *oz.*	-49.43	-49.60	-42.89	6.09
	(0.02)	(0.02)	(0.13)	(0.08)
160 *oz.*	-39.42	-39.48	-39.52	5.42
	(0.05)	(0.05)	(0.26)	(0.40)
Log-likelihood	-45,501	-36,294	-14,244	

* An observation is a shopping trip instance by a household. Standard errors in parentheses.

result is unexpected and hard to interpret. The estimates for the size-specific effects maintain the same ordering and magnitude.

Finally, specification (3) estimates the version of the model described in the previous sections that allows for both heterogeneity (via random coefficients) and forward-looking behaviour. The results for the mean estimates do not change substantively. However, the fit of the model, as measured by the log-likelihood, increases significantly. This illustrates the importance of accounting for heterogeneity, not only to control for different degrees of state-dependence (the results suggest substantial heterogeneity at this level), but also to control for the different *consumer*-size normalizations required for step 1 estimation.

2.4.4 Simulation Algorithm and Goodness of Fit

In this section, I present an algorithm to simulate the several dimensions of the model and examine its fit. The need to specify a simulation algorithm arises because the estimation algorithm decomposes the likelihood of the consumer choices into two components: the choice of whether to purchase (and what size) and the decision of which product to buy if purchasing a positive amount. However, the choice of product influences the purchase decision and vice-versa. In order to address this issue, I propose the following consumer-level algorithm:

1. Solve the simplified dynamic programming problem and simulate the sequence of purchase decisions (whether to purchase and not, and what size if they decide to purchase) conditional on the *observed* inclusive values.

2. For each shopping trip that the consumer decides to purchase a positive amount, simulate her product choice(s).

3. Using the sequence of simulated product choices, I then *simulate* the corresponding inclusive values and update the associated transition probabilities.

4. Iterate the previous steps until convergence of the coefficients of the inclusive values transition processes.

I do not provide here a covergence proof for this algorithm. I note, however, that convergence was, in practice, achieved for *all* consumers after a small number of iterations. The remaining section examines several dimensions of the fit between the simulated and observed sequence of consumer choices.

The simulated probability that a consumer makes a purchase in any given week is 13.65%, which fits the observed probability (15.55%) reasonably well. Figure 2.5 analyzes how the model fits the purchase decision dynamics, by comparing the simulated and observed distribution of inter-purchase duration in days. Overall, the fit is very good, although it slightly underestimates the frequency of purchases for duration spells between 3 and 4 days, at the expense of slightly overestimating the frequency of purchases for durations of 7 days. Other than that, the model is quite accurate in simulating this interpurchase duration. Figure 2.6 examines the hazard rate of purchasing by duration in days from the last purchase, i.e. the probability that the consumer purchases a positive amount given that she has not purchased up to now. Again, the model predicts the pattern of the hazard rate quite accurately only

FIGURE 2.5
Observed and Simulated Interpurchase Duration Distribution

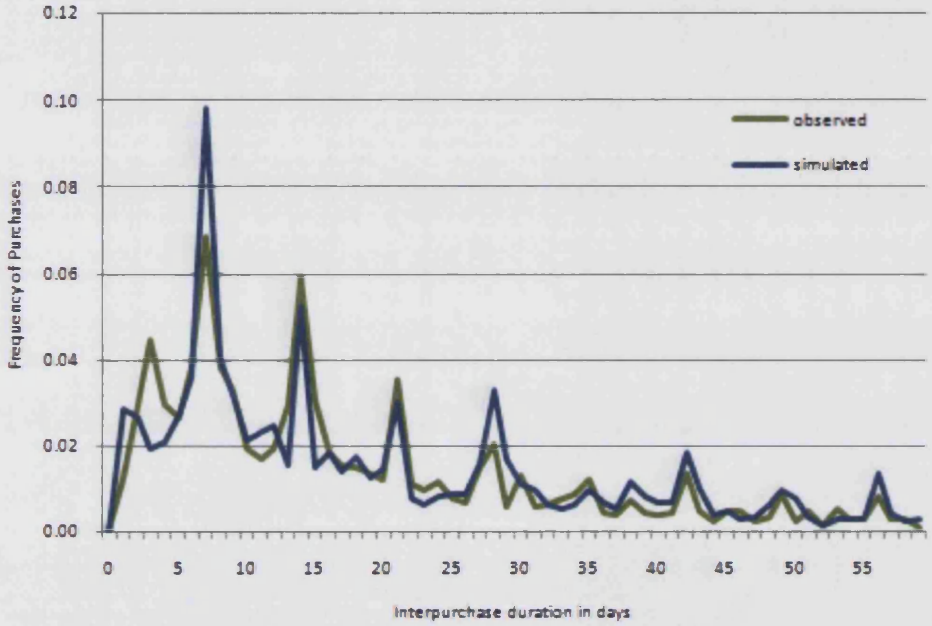


FIGURE 2.6
Observed and Simulated Purchase Hazard

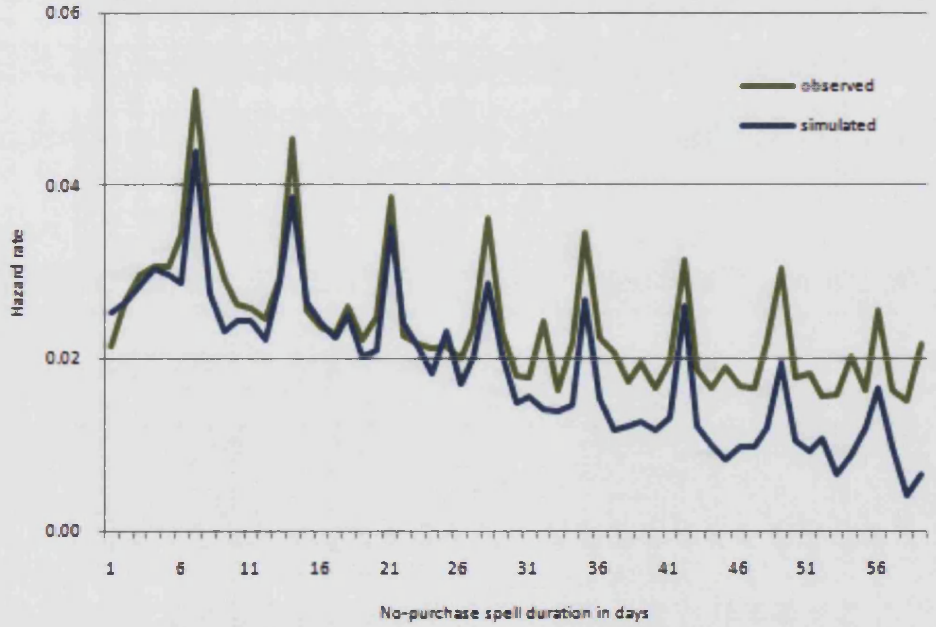
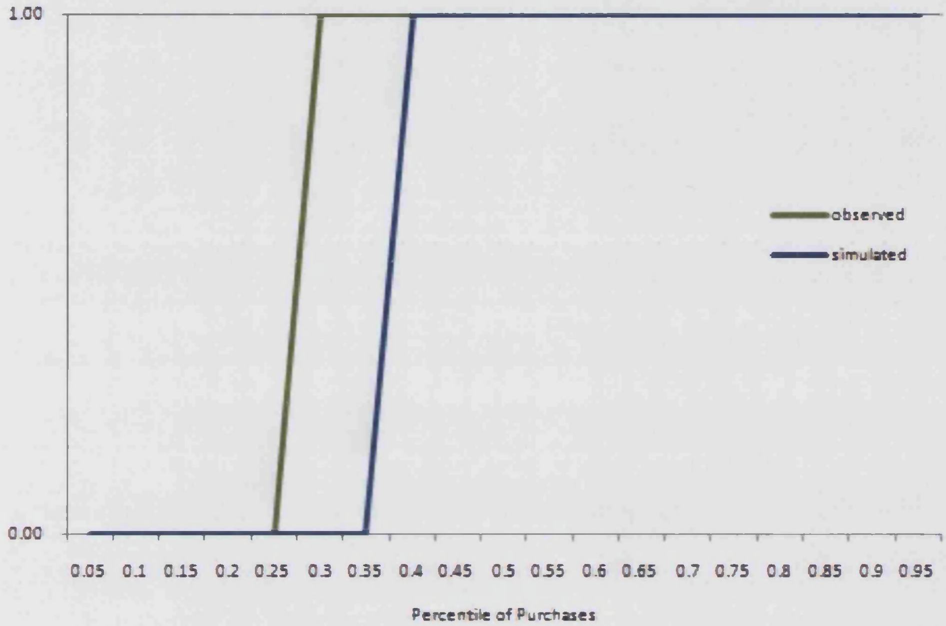


FIGURE 2.7

Observed and Simulated Product Switching Distribution



very slightly underestimating the purchase probabilities for longer durations of no purchase spells, due to the low frequency of purchases with such duration.

Having addressed the purchase decision dynamics, I now move one to examine how the model fits the product switching decision dynamics. The simulated probability that a consumer exhibits a product switch from a purchase instance to the next is 63.79%, which only slightly underestimates the observed probability (72.81%). Figure 2.7 presents the distribution of purchases with regard to product switching. Although there is a slight underestimation of product switching, the fit is reasonably good.

2.5 Policy Implications

The pricing decision is one of the most critical for retailers. In this section, I discuss the implications of a major pricing policy change from hi-low pricing strategy to everyday low pricing (EDLP). In a pure EDLP policy, retailers charge a constant everyday price with no temporary price discounts. In contrast, in a hi-low pricing policy, prices have a higher regular level that remains constant for long periods of time, but then retailers run frequent promotions that lower the price below the EDLP level. In practice, however, pure EDLP

TABLE 2.10

Simulated Effects of Pricing Policy Changes *

Panel A: 0% Price Reduction	Variance Reduction		
	25%	50%	75%
Average Interpurchase Duration (Days)	-0.02	-0.13	-0.25
Proportion of Product Switching	-0.24	-1.35	-1.25
Total Volume Purchased	-0.07	-0.33	-0.43
Total Revenue	1.84	2.95	4.07
Panel B: 5% Price Reduction	Variance Reduction		
	25%	50%	75%
Average Interpurchase Duration (Days)	-0.28	-0.73	-0.77
Proportion of Product Switching	-0.30	-1.22	-1.48
Total Volume Purchased	0.14	0.45	0.62
Total Revenue	0.83	1.12	1.82
Panel C: 10% Price Reduction	Variance Reduction		
	25%	50%	75%
Average Interpurchase Duration (Days)	-0.47	-1.14	-1.31
Proportion of Product Switching	-0.36	-1.30	-1.30
Total Volume Purchased	0.35	1.01	1.22
Total Revenue	-0.17	-0.57	-1.31

* The table reports the percentage changes implied by the different policy changes when compared to the actual pricing strategy.

strategies rarely exist (see Information Resources, Inc. (1993)). EDLP retailers typically charge lower prices on an everyday basis, but do engage in some temporary price discounts.

The pricing policy choice is an empirical question. Hi-low pricing policies have been prevalent in the industry since it allows retailers to price discriminate between consumers that are heterogenous in their price sensitiveness (Pigou (1920)), the degree of price information (Varian (1980)), the level of inventory costs (Blattberg *et al.* (1981), Jeuland and Narasimhan (1985)), or the extent of store loyalty (Narasimhan (1988)), just to mention a few dimensions. However, the success of retailers like Wal-Mart, Home Depot and Toys R Us has increased the popularity of EDLP policies. There are various rationales for adopting EDLP. On the supply side, EDLP is assume to lower operating costs through (*i*) better inventory control, warehouse handling and lower in-store personnel costs due to less variability in demand, and (*ii*) lower advertising expenses due to a focus on image rather than price. On the demand side, EDLP is assumed is to restore price credibility with consumers disenchanted with constant changing prices.

Table 2.10 examines this empirical question. I evaluate the *demand* implications of various degrees of a policy change from high-low pricing towards EDLP in four dimensions: the average interpurchase duration, the proportion of product switching, total volume purchased and total revenue. I should note that I do not compute market equilibrium prices, which is beyond the scope of this chapter (although providing such a framework constitutes a very interesting potential area for future research). I consider only ad-hoc changes in the observed pricing strategy. The table reports the percentage changes implied by different policy changes when compared to the actual pricing strategy.

Table 2.10, Panel A addresses the implications of changing *only* the extent of the hi-low pricing policy by simulating a reduction in price variance, while keeping the mean price for each supermarket-product-size combination constant. The results suggest that a pricing policy that exhibits lower price variance slightly decreases the average interpurchase duration. This is the outcome of two opposite effects. On one hand, the magnitude of the promotion price cuts is now smaller, which reduces the response and purchase acceleration of consumers. On the other hand, reducing price variance around the same mean price also decreases its regular level and consumers respond by increasing the frequency of purchases. In this particular case, the latter effect dominates the former, which induces a slight decrease in the average interpurchase duration.

The results also suggest that the proportion of product switching decreases. This is an expected outcome since lower price variance implies worst price deals, that reduce the promotional induced switching. Finally, the results also suggest that total volume sold drops while revenues increase. This is the outcome of a pricing policy that reduces price deals. Before, a proportional large share of the total volume sold was purchased in promotion. Under the new pricing policy, the attractiveness of temporary price promotions is reduced, which decreases the share of volume sold in promotion. The net effect is a decrease in the quantity sold, but an increase in revenues.

Table 2.10, Panels B and C address the implications of changing not only the extent of the hi-low pricing policy, but also the mean price offered. Here I simulate prices that have *both* a lower mean level and variance. The results with regard to average interpurchase duration and proportion of product switching are qualitatively similar to the ones in Panel A. Total volume sold increases, as expected, in response to the reduction in the mean prices. However this positive impact on quantity is not enough to compensate the drop in price, inducing a decrease in revenues when compared with Panel A pricing experiment.

In sum, the results suggest that the demand profitability of a major pricing policy change from hi-low towards EDLP is questionable, which supports the view that retailers are already maximizing profits. There is evidence that changing the extent of the hi-low pricing policy (by only reducing price variance, while keeping the mean price constant) may be revenue-increasing. However, not knowing the cost function, I can't determine the general impact on profits.

2.6 Conclusions

In this chapter, I attempt to link two strands of the literature on variety-seeking: one focusing on substitutability across time and another on substitutability across products. This issue is economically relevant because both types of substitutability are important for retailers and manufacturers in designing intertemporal price discrimination strategies. I specify a consumer demand model which allows consumption to have an enduring effect and allows the marginal utility of the different products to vary over consumption occasions. I then use the model to evaluate the demand implications of a major pricing policy change from hi-low pricing to an everyday low pricing strategy

I find evidence that consumption has a lasting effect on utility that induces substitutability across time and that the median consumer has a taste for variety in her product decisions. Consumers are found to be forward-looking with respect to the duration since the last purchase, to price expectations and product choices. Pricing policy simulations suggest that retailers may increase revenue by reducing the variance of prices, but that lowering the everyday level of prices may be unprofitable.

This chapter leaves many estimation issues yet to be explored. The development of a framework that allows consumers to be forward-looking in their product choice or incorporates the supply side of the market to derive equilibrium pricing strategies seem very interesting potential areas for future research.

Chapter 3

A Simple Globally Consistent Continuous Demand Model for Market Level Data

This chapter considers a new method of uncovering demand information from market level data on differentiated products. In particular, we propose a continuous-choice demand model with distinct advantages over the models currently in use and describe the econometric techniques for its estimation.

When products are differentiated, the number of parameters required to describe a demand system (without *a priori* restrictions on the substitution patterns) tends to be excessively large to estimate, given the number of observations in a typical dataset. As an illustration of the problem, note that even in the simplest and extremely restrictive of the demand specifications - the linear expenditure model - J products would yield at least J^2 parameters to be estimated, just to capture the substitution patterns. Although implied economic theory's restrictions (like the symmetry of the Slutsky matrix) could be imposed to increase the degrees of freedom available, they do not solve the dimensionality issue. The option for a more flexible functional form would only naturally worsen the problem. Some structure must therefore be placed on the estimation procedure.

The recent industrial economics literature has evolved using primarily discrete-choice models of consumer behaviour. That is, the demand models assume that consumers, while heterogeneous, can purchase at most one unit of one of the available products. Moreover, consumer preferences over products are typically mapped onto a space of characteristics (Lancaster, 1971), reducing the number of parameters to be estimated (since the parameter space is defined by the number of characteristics rather than by the number of products). Within this set of assumptions, we can find the multinomial logit (McFadden, 1974), the nested multinomial logit (McFadden, 1978a), the multinomial probit (Hausman and Wise, 1978), the mixed- or random-coefficients multinomial logit (McFadden, 1981b) and the discrete choice analytically flexible (DCAF) models (Davis, 2006a). The most serious drawback of this branch of the literature relates to the typical trade-off between flexibility and computation requirements. On one hand, we have the standard and the nested multinomial logit models which are fully analytic (and thereby relatively simple to estimate), but imply substitution patterns that tend to be model- instead of data-driven. On the other hand, we have the probit and the mixed-coefficients logit multinomial models which provide increased flexibility by introducing unobserved consumer heterogeneity, but require the use of simulation techniques (which in turn increases substantially the computation requirements). The recent discrete choice analytically flexible model seems to present itself as an exception since it appears to combine the good properties from the two groups.

The assumption that consumers purchase at most one unit of one of the available products may be unrealistic in some settings. Cereals, yogurts, soft drinks and wine are all examples where many consumers typically buy more than one product each time they go shopping. While the discreteness assumption can sometimes be rationalised, it is not always natural to do so given the data available. For example, multiple choices are sometimes modelled as discrete by narrowly defining the choice period and, as a consequence, rationalize multiple observed choices by assuming the consumer has made two separate purchasing decisions. Alternatively, a discrete-choice approach can be extended to model multiple choices by defining the choice set to include product bundles. However, often the discrete choice assumption can not easily be motivated given the products being selected and so it may be more natural to model consumers as making continuous quantity choices.

The continuous-choice literature typically assumes a representative agent that might consume all products and uses functional forms that allow flexible substitution patterns. Examples include the translog model of Christensen *et al.* (1975), the almost ideal demand system (AIDS) due to Deaton and Muellbauer (1980) and the distance metric model from Pinkse and Slade (2004). Consumer preferences can be defined directly over products (as in the translog or in the AIDS cases) or mapped onto a space of characteristics in a manner akin to the discrete-choice literature (as in the distance metric model). However, this set of models presents a serious limitation. In contrast to the discrete-choice framework, they can not be used to uncover demand information from markets with significant entry and exit of products. This limitation has been addressed in the literature before, but never in a way that, to the best of our knowledge, we could categorize as adequate. The typical solutions are largely limited to either consider substitution patterns between broad aggregates of products as, for example, in Christensen *et al.* (1975), Deaton and Muelbauer (1980), and Hausman *et al.* (1994), or to estimate the demand system using data only from time periods when all products are present in the market as, for example, in Hausman (1994), Ellison *et al.* (1997), and Pinkse and Slade (2004).

More recently, the continuous-choice literature has evolved to model consumer heterogeneity explicitly. Chan (2006) constitutes an example of such approach. He develops a continuous hedonic-choice model to investigate the demand for soft drinks that, in contrast to the more traditional line of the literature, is able to cope with entry and exit of products. However, his approach, in addition to requiring simulation techniques, does not model unobserved product characteristics.

In this chapter, we follow the traditional continuous-choice literature and develop a representative consumer flexible demand model. We begin by specifying an indirect utility function from which, via Roy's identity, a continuous-choice demand system is derived. The demand function implied by the model is fully analytic and therefore avoids the burden of simulation. The model is flexible in the sense of Diewert (1973, 1974) as the implied own- and cross-price elasticities are capable of capturing the true substitution patterns in the data. In addition, the model can accommodate the use of data on the entry and exit of products. The importance of this last property is twofold as *i*) not being able to cope with entry and exit patterns limits the application of the above models and *ii*) the ability to deal with variation in the set of choices available to consumers provides pseudo-price variation which can be very helpful in the identification of substitution patterns between products.

In order to encompass different possible real-world applications, we consider two alternative specifications of our baseline model depending on the degree of flexibility the researcher is willing to accept for the substitution patterns between inside and outside goods. For estimation, we propose an analog to the algorithm derived in Berry (1994), Berry, Levinsohn and Pakes (1995) (henceforth BLP). Following this line of the literature, the error term is structurally embedded in the model and thereby circumvents the critique provided by Brown and Walker (1989) related to the addition of add-hoc errors and their induced correlations. Depending on the specification considered, the contraction mapping for matching observed and predicted expenditure shares may be analytic or not. For the case it is not, we present an alternative procedure to BLP's contraction mapping with a super-linear rate of convergence.

To sum up, the main contribution of the chapter is to propose a new continuous-choice model that combines the centrally desirable properties of both the discrete- and continuous-choice traditions: *i*) it is flexible in the sense of Diewert (1973, 1974), *ii*) can deal with the entry and exit of products over time, and *iii*) incorporates a structural error term. Furthermore, it is relatively simple and fast to estimate which can prove a key advantage in competition policy issues where time and transparency are always crucial factors.

3.1 The Demand Model

3.1.1 The General Setup

Let $\Im$ denote the set of $J+1$ choices available to consumers, with options $j = 1, \ldots, J$ referring to the inside choices and $j = 0$ referring to an outside option that aggregates all remaining ones, including that of no purchase. We follow the continuous choice literature and define the demand system by specifying an indirect utility function $V(p, y; \theta, \Im)$ for the representative consumer. p denotes the $(J+1)$ vector of $p_i \in \Re^+$ prices, $y \in \Re^+$ denotes the income of the representative consumer and, finally, θ refers to a set parameters which we assume to have support $\Theta \subseteq R^\theta$. For reasons that will become clear later, we also index the utility function by the set of available goods $\Im$.

$V(p, y; \theta, \Im)$ is assumed to satisfy the properties of an indirect utility function, namely to be a continuous function in p and y, strictly increasing in y and nonincreasing in p_j for any p_j, quasiconvex, and homogeneous of degree zero. If $V(p, y; \theta, \Im)$ has the properties of an indirect utility function, then standard duality results imply that the demand system is easily obtained via Roy's identity. In this chapter, we restrict further the class of functions considered by requiring the indirect utility function to satisfy an additional global regularity property, namely the sub-class of indirect utility functions that we define to be globally consistent.

Definition 3.1 *An indirect utility function $V(p, y; \theta, \Im)$ is globally consistent if and only if for any set of products $\Im' \subset \Im$:*

$$V(p, y; \theta, \Im') = \lim_{p_k \to \infty} V(p, y; \theta, \Im),$$

for all $k \in \Im$ and (simultaneously) $k \notin \Im'$.

This corresponds directly to Mcfadden (1981b)'s social surplus condition. We name this property global consistency since it requires that an indirect utility function, defined on the set of all possible products $\Im$, can be specialized down in an entirely consistent fashion to generate demand systems over arbitrary subsets of the goods $\Im'$. It encapsulates the very natural restriction that removing a good from the choice set is entirely equivalent to increasing its price to infinity.

Proposition 3.1 *Any indirect utility function $V(p, y; \theta, \Im)$ which satisfies global consistency* *and non-satiation (that is, non-zero marginal utility of income) generates a demand system via Roy's identity which enjoys the property that:*

$$q_i\left(p, y; \theta, \Im'\right) = \lim_{p_k \to \infty} q_i\left(p, y; \theta, \Im\right),$$

for all $i, k \in \Im$, $i \in \Im'$ *and* $k \notin \Im'$, *where* $q_i\left(p, y; \theta, \Im\right)$ *denotes the Marshallian demand function of good* i. *Furthermore:*

$$q_k\left(p, y; \theta, \Im\right) = 0,$$

for all $k \notin \Im'$.

Proof. If $V\left(p, y; \theta, \Im\right)$ has the properties of an indirect utility function then standard duality results imply that, under non-satiation, the demand system is easily obtained via Roy's identity:

$$q_i\left(p, y; \theta, \Im\right) = \frac{V_i\left(p, y; \theta, \Im\right)}{V_y\left(p, y; \theta, \Im\right)},$$

where $V_i\left(p, y; \theta, \Im\right)$ denotes the marginal utility with respect to p_i and $V_y\left(p, y; \theta, \Im\right)$ denotes the marginal utility of income. It is easy to verify that if $V\left(p, y; \theta, \Im\right)$ satisfies in addition the global consistency property, then:

$$\begin{aligned} V_i\left(p, y; \theta, \Im'\right) &= \lim_{p_k \to \infty} V_i\left(p, y; \theta, \Im\right) \\ V_y\left(p, y; \theta, \Im'\right) &= \lim_{p_k \to \infty} V_y\left(p, y; \theta, \Im\right), \end{aligned}$$

for all $i, k \in \Im$, $i \in \Im'$ and $k \notin \Im'$. This yields the first part of the proposition:

$$q_i\left(p, y; \theta, \Im'\right) = \frac{V_i\left(p, y; \theta, \Im'\right)}{V_y\left(p, y; \theta, \Im'\right)} = \frac{\lim_{p_k \to \infty} V_i\left(p, y; \theta, \Im\right)}{\lim_{p_k \to \infty} V_y\left(p, y; \theta, \Im\right)} = \lim_{p_k \to \infty} q_i\left(p, y; \theta, \Im\right).$$

For all $k \notin \Im'$, we have, under global consistency, that $V_k\left(p, y; \theta, \Im\right) = 0$, which together with non-satiation yields that $q_k\left(p, y; \theta, \Im\right) = 0$. In other words, removing a good from the choice set explicitly forces the level of demand for that good to zero. ■

Proposition 3.1 establishes that a demand system derived from an indirect utility function that satisfies global consistency and non-satiation *can* be estimated using datasets where significant product entry and exit occurs. The possibility of using datasets where significant product entry and exit occurs provides a potentially useful source of pseudo-price variation to help identify substitution patterns.

Surprisingly, the extremely mild and intuitive global consistency condition is *not* satisfied by the vast majority of existing continuous-choice models like the translog model of Christensen *et al.* (1975), the almost ideal demand system (AIDS) due to Deaton and Muellbauer (1980) and the distance metric model from Pinkse and Slade (2004).[7] In all of these examples, demand depends linearly on prices (or on its logarithms) and if a product is not present in a given market such a demand system can not be estimated.

The typical solutions include either consider substitution patterns between broad aggregates of products (a level of aggregation which eliminates product entry and exit) as, for example, in Christensen *et al.* (1975), Deaton and Muelbauer (1980), and Hausman *et al.* (1994), or to estimate the demand system using data only from markets (e.g. time periods) when all goods are present as, for example, in Hausman (1994), Ellison *et al.* (1997), and Pinkse and Slade (2004). The former type of solution involves resorting to an analysis of aggregate data which clearly limits our ability to describe the substitution patterns between the goods actually being purchased by consumers. The latter type of solution, on the other hand, involves resorting to data only from markets when all goods are present, which while effective (if potentially inefficient) in some markets, such as the pharmaceutical markets studied by Ellison *et al.* (1997) where generic entry is driven by loss of patent protection so all entry occurs within a very constrained period in the data, would be largely impractical in other arena where product entry and exit occur simultaneously.

In contrast with existing literature, proposition 3.1 establishes that a demand system derived from an indirect utility function that satisfies global consistency and non-satiation *can* be estimated using datasets where significant product entry and exit occurs. Removing a good from the choice set is entirely equivalent to increasing its price to infinity.

We now move on to specify a computationally convenient change of variable. Let, without loss of generality, $V(p, y; \theta, \Im) = H(r, y; \theta, \Im)$ where r denotes the element-by-element inverse $(J+1)$ vector of p_i: $r_i = 1/p_i \in \Re$. As it is easy to verify, under the new indirect utility function, removing a good k from the choice set is entirely equivalent to decreasing the corresponding r_k to zero. Definition 3.1' below provides a restatement of definition 3.1 in terms of the set of (relatively easy to verify) conditions on the function $H(r, y; \theta, \Im)$ that are sufficient to ensure that the resulting indirect utility function is a member of the class of consistent indirect utility functions and may therefore be estimated using pseudo price

[7]While the vast majority of indirect utility function specifications used to generate continuous choice demand models are not members of the set of consistent indirect utility functions, a very few existing demand systems are. These are generally models which have not been empirically popular. For example, the Indirect Addilog model considered by Houthakker, the Translog Reciprocal Indirect Utility Function and Diewert's Reciprocal Indirect Utility Function. See for example Varian (1984) for a discussion of these models and further references. More recently, Chan (2006) develops a continuous hedonic-choice model that is able to cope with entry and exit of products. However, his approach, has the undesirable feature of not modelling unobserved product characteristics.

variation.

Definition 3.1' *A globally consistent indirect utility function* $V(p, y; \theta, \Im) = H(r, y; \theta, \Im)$ *possesses the following properties:*

1. *Continuous in* p *and* y.
2. *Homogeneous of degree zero in* p *and* y.
3. *Strictly increasing in* y *and nonincreasing in* p_i *for any* $i \in \Im$.
4. *Quasiconvex, that is, the set* $\{(p, y) : V(p, y; \theta, \Im) \leq \bar{v}\}$ *is convex for any* $\bar{v}$.
5. $\lim_{p_k \to \infty} V(p, y; \theta, \Im) = \lim_{r_k \to 0} H(r, y; \theta, \Im) = V(p, y; \theta, \Im') = H(r, y; \theta, \Im')$ *for any set of products* $\Im' \subset \Im$ *and for all* $k \notin \Im'$.

Following the vast majority of the continuous-choice literature, we will describe the demand system derived from the globally consistent indirect utility function $H(r, y; \theta, \Im)$ in terms of the Marshallian budget share functions:

$$w_i(r, y; \theta, \Im) = \frac{H_i(r, y; \theta, \Im)\, r_i}{H_y(r, y; \theta, \Im)\, y}, \tag{27}$$

for $i \in \Im$ and where $H_i(r, y; \theta, \Im)$ denotes the marginal utility with respect to r_i, $H_y(r, y; \theta, \Im)$ denotes the marginal utility of income, and finally $w_i(r, y; \theta, \Im) = p_i q_i(r, y; \theta, \Im)/y$ denotes the budget share of good i.

3.1.2 Two Specifications

The actual algebraic functional form for the indirect utility function $H(r, y; \theta, \Im)$ is unknown to the econometrician. Following the continuous-choice literature, we approximate it with a flexible functional form so not to restrict the derived price substitution patterns. In particular, we present two different functional specifications for $H(r, y; \theta, \Im)$, each with its own set of advantages and disadvantages, that can cope with the different needs of various real-world applications.

Specification A

In specification A, we approximate the indirect utility function with a normalized quadratic function, following Berndt *et al.* (1977), McFadden (1978b), and Pinkse and Slade (2004), in which both the representative consumer income y and the inside goods r_i have been normalized by the outside option r_0. The normalized approximation is given by:

$$H^A(r,y;\theta,\Im) = a_0 + \sum_{m=1}^{J} a_m r_m + 1/2 \sum_{m=1}^{J} \sum_{n=1}^{J} b_{mn} r_m r_n + c_0 y + \sum_{m=1}^{J} c_m r_m y, \quad (28)$$

where, with a slight abuse of notation, $H^A(r,y;\theta,\Im)$ denotes the normalized indirect utility function, and both r and y denote the vectors of the corresponding normalized variables. a_0 is a scalar parameter, $a = [a_i]$ is a J vector of parameters that, as we will discuss below, will be used to capture the vector of observed budget shares, $B = [b_{ij}]$ is a $J \times J$ matrix of parameters that will be used to capture the price substitution patterns, and finally $c = [c_i]$ is a J vector of parameters that will be used to capture the income effects.

It is easy to verify that the function $H^A(r,y;\theta,\Im)$ satisfies properties 1, 2 and 5 for a globally consistent indirect utility function. It is an homogeneous of degree zero functional form and continuous in both p and y. Moreover, it can be specialized down in an entirely consistent fashion to generate demand systems over arbitrary subsets of the goods (by setting $r_k = 0$ for those goods not in the current choice set). However, if we allow the parameters to be unrestricted in sign and magnitude, it does not necessarily satisfy properties 3 and 4. We can either impose a set of restrictions on the parameters that, given the vector of prices and income, ensure $H^A(r,y;\theta,\Im)$ satisfies those properties (Appendix I lists the set of those implied restrictions) or, alternatively, we may choose, even in the absence of an underlying model of utility, not to impose such restrictions *a priori*, but rather test whether the data is consistent with them.

If $H^A(r,y;\theta,\Im)$ has the properties of an indirect utility function then standard duality results imply that the demand system is easily obtained via Roy's identity for each inside good $i = 1, \ldots, J$:

$$w_i^A(r,y;\theta,\Im) = \frac{a_i r_i + 1/2 \sum_{m=1}^{J} (b_{im} + b_{mi}) r_i r_m + c_i r_i y}{c_0 y + \sum_{m=1}^{J} c_m r_m y}, \quad (29)$$

where $w_i^A(r,y;\theta,\Im)$ denotes the budget share of good i for specification A. The derived budget share function has three characteristics that are important to discuss. First, it satisfies Proposition 3.1 and hence removing a good from the choice set (equivalent to set

$r_i = 0$) explicitly forces the level of demand to zero. As a consequence, the model is able to match the shares of those goods with zero observed demand and hence can be estimated using datasets where significant product entry and exit occurs. Second, although it is only defined explicitly for the J inside choices (given the normalization with respect to the outside option), the budget constraint implies that a complete model of demand for the $J+1$ budget shares is derived implicitly. Finally, because it is homogeneous of degree zero in the parameters, identification requires a normalization. Without loss of generality, we normalize c_0.

In many policy applications, including merger simulations, the key object of interest is the matrix of own- and cross-price demand elasticities. The analytical expressions for the budget share, price- and income-elasticities implied by the model with respect to any given inside goods i and j, are the following:

$$\begin{aligned} \varepsilon_{ij}^{s}\left(r, y; \theta, \Im\right) &= -1\left(j = i\right) - \frac{1/2\left(b_{ij} + b_{ji}\right) r_j r_i \left(w_i^A y\right)^{-1} - c_j r_j}{c_0 + \sum_{m=1}^{J} c_m r_m} \\ \eta_i^s\left(r, y; \theta, \Im\right) &= -1 + \frac{c_i r_i}{c_0 w_i^A + \sum_{m=1}^{J} c_m r_m w_i^A}, \end{aligned} \tag{30}$$

where ε_{ij}^s and η_i^s denote the price- and income-elasticities, respectively. $[b_{ii}] > 0$ constitutes a sufficient condition (although not necessary) for a downward sloping own demand curve. We may expect $[b_{ij}] < 0$ if goods i and j are substitutes, but this constitutes neither a necessary nor a sufficient condition, since the total price effect depends on the size of the income effect. From the *budget share* elasticities, we can straightforwardly obtain the implied *demand* elasticities for the corresponding inside goods.[8] While the elasticities involving the outside good can not be estimated directly (as, under specification A, the model is defined only over the J inside goods), the budget constraint implies that those elasticities can, nevertheless, still be recovered from equilibrium behavior, given the elasticities for the inside goods.

Specification A has both an advantage and a disadvantage relative to specification B below. The disadvantage is that the use of equilibrium behavior to implicitly derive the elasticities involving the outside good may restrict the estimated substitution patterns. The advantage is that the derived income-elasticities are, as we will show below, Diewert flexible. Specification A is, therefore, particularly suitable for real-world applications where the importance of the outside good is relatively small and where demand exhibits important income effects. An additional advantage of this specification is that, as we show below, it is

[8] For completeness $\varepsilon_{ii}^d = \varepsilon_{ii}^s - 1$, $\varepsilon_{ij}^d = \varepsilon_{ij}^s$ and $\eta_i^d = \eta_i^s + 1$, where ε_{ij}^d and η_i^d denote the demand price- and income-elasticities, respectively.

relatively simple and fast to estimate which can prove a key advantage in competition policy issues, where time and transparency are typically crucial factors.

Specification B

In specification B, we approximate the indirect utility function with a generalized Leontief indirect utility function, following Diewert (1971):

$$H^B(r, y; \theta, \Im) = \sum_{m=0}^{J} a_m r_m y + 1/2 \sum_{m=0}^{J} \sum_{n=0}^{J} b_{mn} r_m r_n y^2, \tag{31}$$

where $a = [a_i]$ is a $J + 1$ vector of parameters that, as we discuss below, will be used to capture the vector of observed budget shares, and $B = [b_{ij}]$ is a $(J + 1) \times (J + 1)$ matrix of parameters that will be used to capture the price substitution patterns.

It is easy to verify that the function $H^B(r, y; \theta, \Im)$ satisfies properties 1, 2 and 5 for a globally consistent indirect utility function. It is an homogeneous of degree zero functional form and continuous in both p and y. Moreover, it can be specialized down in an entirely consistent fashion to generate demand systems over arbitrary subsets of the goods. However, like in specification A, if we allow the parameters to be unrestricted in sign and magnitude, it does not necessarily satisfy properties 3 and 4. We can either impose a set of restrictions on the parameters that, given the vector of prices and income, ensure those properties are satisfied (Appendix I lists the set of those implied restrictions) or, alternatively, we may choose, even in the absence of an underlying model of utility, to test whether the data is consistent with them.

If $H^B(r, y; \theta, \Im)$ has the properties of an indirect utility function then standard duality results imply that the demand system is easily obtained via Roy's identity for $i = 0, \ldots, J$:

$$w_i^B(r, y; \theta, \Im) = \frac{a_i r_i y + 1/2 \sum_{m=0}^{J} (b_{im} + b_{mi}) r_i r_m y^2}{\sum_{m=0}^{J} a_m r_m y + 1/2 \sum_{m=0}^{J} \sum_{n=0}^{J} (b_{mn} + b_{nm}) r_m r_n y^2}, \tag{32}$$

where $w_i^B(r, y; \theta, \Im)$ denotes the budget share of good i under specification B. Similarly to the specification A case, the derived budget share function satisfies Proposition 3.1 and is homogeneous of degree zero in the parameters. The model can hence be estimated using datasets where significant product entry and exit occurs. Identification requires, however, a normalization. Without loss of generality, we normalize b_{00}.

The analytical expressions for the budget share elasticities implied by the model with respect to any given goods i and j (both inside and outside), are the following:

$$\begin{aligned}
\varepsilon_{ij}^{s}\left(r, y; \theta, \Im\right) &= -1\left(j=i\right) - \frac{1/2\left(b_{ij}+b_{ij}\right) r_i r_j y^2 / w_i^B - 1/2 \sum_{m=0}^{J}\left(b_{jm}+b_{mj}\right) r_j r_m y^2}{\sum_{m=0}^{J} a_m r_m y + 1/2 \sum_{m=0}^{J} \sum_{n=0}^{J}\left(b_{mn}+b_{nm}\right) r_m r_n y^2} + w_j^B \\
\eta_i^{s}\left(r, y; \theta, \Im\right) &= \frac{1/2 \sum_{m=0}^{J}\left(b_{im}+b_{mi}\right) r_i r_m y^2 / w_i^B - 1/2 \sum_{m=0}^{J} \sum_{n=0}^{J}\left(b_{mn}+b_{nm}\right) r_m r_n y^2}{\sum_{m=0}^{J} a_m r_m y + 1/2 \sum_{m=0}^{J} \sum_{n=0}^{J}\left(b_{mn}+b_{nm}\right) r_m r_n y^2}.
\end{aligned} \tag{33}$$

We may expect $[b_{ii}] > 0$ and $[b_{ij}] < 0$ if goods are substitutes and face a downward sloping demand curve, but this constitutes neither a necessary nor a sufficient condition. Again, we can straightforwardly obtain the implied demand elasticities from the budget share elasticities. Note that, in contrast to specification A, specification B explicitly models all $(J+1)$ options. The price substitution patterns involving the outside good can, therefore, be estimated *directly* (although under some identification restrictions as we discuss below). Unfortunately, this greater degree in flexibility towards the outside good is traded-off against a lower degree in flexibility towards the income effects as this second specification does impose restrictions on the derived income-elasticities. To sum up, specification B is therefore particularly suitable for real-world applications where the outside good is of relative importance and where income effects are small.

3.1.3 The More General Budget Share Function

Sometimes, it may be interesting to estimate a model which is more general than the model described thus far. The reason is twofold. First, as discussed above, we may choose, even in the absence of an underlying model of utility, to estimate a model without imposing *a priori* restrictions to the parameters (in sign and/or magnitude) and test whether the data is consistent with the properties of a globally consistent indirect utility function.

Second, we may choose to estimate a specification that is asymmetric with reference to the matrix B of parameters. As it is easy to note, the two specifications described above are observationally equivalent to a symmetric model with $b_{ij} = b_{ji} = 1/2\left(b_{ij}+b_{ji}\right)$. Both the budget share and elasticities functions do not depend specifically on the individual $[b_{ij}]$ parameters, but only on their sum. Although this property of the model provides a great advantage in terms of the estimation procedure (as the number of parameters to be estimated decreases substantially), it implicitly restricts the flexibility properties of the model. The price substitution patterns of a demand system with $J+1$ goods can assume up to $(J+1)^2$ arbitrary values (or J^2 if we focus only on the inside goods as in specification A). Under a symmetric specification, the model requires estimates of $(J+1)(J+2)/2$ parameters $[b_{ij}]$ (or $J(J+1)/2$ parameters in specification A), which are clearly insufficient to assume the required arbitrary values. Following the continuous-choice tradition, we may often wish to estimate a model which does not impose such symmetry restrictions.

Specifically, consider the following *more general budget share functions* for the case in which properties 3 and 4, as well as the symmetry assumptions, are not imposed *a priori*:

$$
\begin{aligned}
s_i^A(r,y;\theta,\Im) &= \frac{a_i r_i + \sum_{m=1}^J b_{im} r_m r_i + c_i r_i y}{c_0 y + \sum_{m=1}^J c_m r_m y} \\
s_i^B(r,y;\theta,\Im) &= \frac{a_i r_i y + \sum_{m=0}^J b_{im} r_m r_i y^2}{\sum_{m=0}^J a_m r_m y + \sum_{m=0}^J \sum_{n=0}^J b_{mn} r_m r_n y^2},
\end{aligned} \tag{34}
$$

where $s_i^A(r,y;\theta,\Im)$ and $s_i^B(r,y;\theta,\Im)$ denote the functions for specifications A and B, respectively. In these circumstances, the model is obviously not consistent with consumer utility maximization, but rather it nests a model which is and hence we can test the validity of such assumptions. If these restrictions are consistent with the patterns in the data, we can subsequently impose them on the model.

The analytic expressions for the price- and income-elasticities implied by the more general budget share for specification A are:

$$
\begin{aligned}
\varepsilon_{ij}^s(r,y;\theta,\Im) &= -1(j=i) - \frac{b_{ij} r_j r_i \left(s_i^A y\right)^{-1} - c_j r_j}{c_0 + \sum_{m=1}^J c_m r_m} \\
\eta_i^s(r,y;\theta,\Im) &= -1 + \frac{c_i r_i}{c_0 s_i^A + \sum_{m=1}^J c_m r_m s_i^A},
\end{aligned} \tag{35}
$$

whereas for specification B, we have:

$$
\begin{aligned}
\varepsilon_{ij}^s(r,y;\theta,\Im) &= -1(j=i) - \frac{b_{ij} r_i r_j y^2 / s_i^B - \sum_{m=0}^J b_{mj} r_m r_j y^2}{\sum_{m=0}^J a_m r_m y + \sum_{m=0}^J \sum_{n=0}^J b_{mn} r_m r_n y^2} + s_j^B \\
\eta_i^s(r,y;\theta,\Im) &= \frac{\sum_{m=0}^J b_{mi} r_m r_i y^2 / s_i^B - \sum_{m=0}^J \sum_{n=0}^J b_{mn} r_m r_n y^2}{\sum_{m=0}^J a_m r_m y + \sum_{m=0}^J \sum_{n=0}^J b_{mn} r_m r_n y^2}.
\end{aligned} \tag{36}
$$

3.1.4 Product Characteristics

Lancaster (1971) suggests that consumers are interested in goods because of the characteristics they provide. Classical choice models can be generalized to incorporate such proposal and introduce preferences directly over product characteristics. The model described in this chapter is no exception. This introduction places no additional restrictions or structure on the form of the indirect utility function $H(r, y; \theta, \Im)$ as, in principle, such characteristics may enter through any of the parameters of the model in an arbitrary fashion. Moreover, it carries three important advantages.

First, it allows the introduction of a structural error term addressing the important critique offered by Brown and Walker (1989) which warns about the potential risks of simply 'tagging' on linear error terms to the end of budget share equations.

Let the set of parameters θ be decomposed, for notational purposes, into $\theta = (\theta_1, \theta_2)'$, where θ_1 refers to the $[a_i]$ parameters for $i = 1, \ldots, J$ and θ_2 refers to the remaining ones ($[b_{ij}]$ and $[c_i]$ in specification A, and $[b_{ij}]$ and a_0 - for reasons that will become clear below - in specification B). We introduce the random utility hypothesis by assuming that all product characteristics are observed by the consumer, but not necessarily by the econometrician. In particular, we follow Pinkse *et al.* (2002) and map the θ_1 parameters onto the characteristics space:

$$\theta_1 \equiv a_i(x_i, \xi_i; \beta), \tag{37}$$

where x_i denotes the K-dimensional vector of characteristics associated with good i, observed by both the consumer and the econometrician, ξ_i denotes a one-dimensional vector of characteristics that are observed by the consumer, but not by the econometrician, and finally β refers to the K-dimensional vector of taste parameters associated with the observed characteristics. The precise functional form for $a_i(x_i, \xi_i; \beta)$ is an issue that can be examined using conventional testing procedures. We follow the industrial economics literature and assume a linear specification:

$$a_i(x_i, \xi_i; \beta) = \textstyle\sum_{k=1}^{K} \beta_k x_{ki} + \xi_i, \tag{38}$$

which has the desirable property of being monotonic in the value of a given product's characteristics. The presence of unobserved product characteristics allows for a product-level source of sampling error, giving an explicit structural interpretation to the error term.

The second advantage of introducing product characteristics is related to the fact that it can substantially reduce the number of parameters to be estimated. If the number of goods $J+1$ is large, a dimensionality problem may arise as the model may yield too many parameters to be estimated with the available data. In that case, the number of parameters can be reduced by also mapping the θ_2 parameters onto the characteristics space (whenever the number of characteristics is smaller than the number of goods):

$$\begin{aligned} b_{ii} &= g_1\left(x_{1i};\alpha_1\right) \\ b_{ij} &= g_2\left(d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)\right) \\ c_i &= g_3\left(x_{3i};\alpha_3\right), \end{aligned} \tag{39}$$

where $g_1\left(x_{1i};\alpha_1\right)$ is a function of a set $L1$ of good i's characteristics, $g_2\left(d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)\right)$ is a function of a distance metric between goods i and j in the set of characteristics space $L2$, and finally (for specification A only) $g_3\left(x_{3i};\alpha_3\right)$ is a function of a set $L3$ of good i's characteristics. In theory, all observed characteristics could be mapped onto both θ_1 and θ_2 sets of parameters. In this case, the sets $L1$, $L2$ and $L3$ will all coincide with the set of K-observed characteristics. In real-world applications, however, we hypothesize that if we do so, some high correlation may introduce biases in the estimation procedure. For this reason, we suggest that if all characteristics are allocated to all sets, some transformation should be used (see Pinkse and Slade, 2004).

The precise mapping is, again, a functional form issue that can be examined using conventional testing procedures. The following specifications are among the possible alternatives:

$$\begin{aligned} b_{ii} &= x'_{1i}\alpha_1 \\ b_{ij} &= d_{ij}\left(x_{2i},x_{2j};\alpha_2\right) \\ c_i &= x'_{3i}\alpha_3, \end{aligned}$$

or:

$$\begin{aligned} b_{ii} &= \exp\left(x'_{1i}\alpha_1\right) \\ b_{ij} &= \exp\left(d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)\right) \\ b_{ij} &= \exp\left(x'_{3i}\alpha_3\right), \end{aligned}$$

where the distance metric could be defined as $d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)=\sqrt{\sum_{l=1}^{L2}\alpha_{2l}\left(x_{2li}-x_{2lj}\right)^2}$ or $d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)=\sum_{l=1}^{L2}\alpha_{2l}\left|x_{2li}-x_{2lj}\right|$.[9] Another obvious alternative is to estimate the functions nonparametrically as in Pinkse *et al.* (2002). Independently of the specification chosen

[9] As a technical note, if *a priori* we want to impose $b_{ij}\leq 0$, the following alternatives are possible: $b_{ij}=-\exp\left(d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)\right)$ or $b_{ij}=-d_{ij}\left(x_{2i},x_{2j};\alpha_2\right)$.

though, the important fact is that the θ_2 parameters are mapped onto the characteristics space and hence the number of parameters to be estimated is reduced while allowing the estimates to still be data-driven.

The third and final advantage of introducing product characteristics is related to the new goods problem (Ackerberg *et al.*, 2005). If a demand system is defined over the product space, we can not investigate demand behavior for goods not yet introduced. The introduction of product characteristics solves this problem and makes the analysis of issues related to incentives for entry, possible.

3.1.5 Flexibility

An algebraic functional form for a complete system of consumer budget share functions $s_i(r, y; \theta, \Im)$ is said to be Diewert flexible (see Diewert, 1973, 1974 and Lau, 1986) if, at any given set of non-negative prices and income, the parameters can be chosen so that the complete system of consumer budget share functions, their own- and cross-price demand and income elasticities are capable of assuming arbitrary values at the given set of prices and income (subject only to the requirements of theoretical consistency). Barnett (1983) proved that flexibility in the sense defined by Diewert is necessary and sufficient for a function to satisfy the mathematical definition of a local second order approximation. In this section, we follow Diewert (1973,1974) and show that both our specifications are flexible in that sense.[10]

Matching Predicted to Observed Shares

The first step in establishing flexibility is to show that, for every set of θ_2 parameters, there is a unique value of θ_1 that equates the shares predicted by the model $s_i(r, y; \theta, \Im)$ with the observed shares s_i^{true}, where $s_i(r, y; \theta, \Im)$ denotes the *more general budget share function.* Recall that θ_1 refers to the $[a_i]$ parameters for $i = 1, \dots, J$ and θ_2 refers to the remaining ones ($[b_{ij}]$ and $[c_i]$ in specification A, and $[b_{ij}]$ and a_0 - for reasons that will become clear below - in specification B). This step ensures the model can always match the vector of observed shares, one requirement for a model to be a Diewert flexible functional form. We then proceed by showing that the set of θ_2 parameters is such that the predicted elasticities are capable of assuming arbitrary values.

[10]Our problem differs from Diewert (1971, 1973, 1974) in the sense that we consider flexible functional forms as approximations to indirect utility and demand functions rather than to cost and production functions (Diewert, 1971) or profit and transformation functions (Diewert, 1973) or revenue and factor requirements functions (Diewert, 1974).

Let $\Im_0^+$ denote the subset of $\Im$ that includes the goods that are effectively present in the market and exhibit, as a consequence, strictly positive observed shares: $\Im_0^+ \equiv \{i | s_i^{obs} > 0, i \in \Im\}$. Furthermore, let J_0^+ denote the number of goods in set $\Im_0^+$. In specification A, it will be useful to also define $\Im^+$, the subset of $\Im_0^+$ that includes only the inside goods (in other words, the subset of $\Im_0^+$ that excludes the outside good).

Proposition 3.2 *The more general budget share function* $s_i(r, y; \theta, \Im)$ *can match, under specification A, any vector of observed budget shares. Furthermore, there exists a unique set of* θ_1 *parameters that matches the subset* $\Im_0^+$ *of goods with strictly positive observed demand.*

Proof. Proposition 3.1 establishes that the model explicitly matches predicted and observed demand for all goods that exhibit zero observed shares and hence we can proceed by restricting our analysis to the subset of remaining goods in $\Im$.

We begin by showing that, under specification A, the system of inside goods $J_0^+ - 1$ equations $s_i^{true} = s_i^A(r, y; \theta, \Im)$ for $i \in \Im^+$ has exactly one solution $a(r, y, s^{true}; \theta_2, \Im)$ that equates the shares predicted by the model to the observed shares. In order to see why this is the case, note that the system of equations:

$$s_i^{true} = \frac{a_i r_i + \sum_{m=1}^{J} b_{im} r_m r_i + c_i r_i y}{c_0 y + \sum_{m=1}^{J} c_m r_m y}$$

for $i \in \Im^+$ is linear in the vector $a = (a_1, \ldots, a_{J^+})'$ and can be rewritten as $Da = G$, where D denotes a $(J_0^+ - 1) \times (J_0^+ - 1)$ diagonal matrix with diagonal elements $d_{ii} = r_i$ and G denotes a $J_0^+ - 1$ vector with elements $g_i = s_i^{true}\left(c_0 y + \sum_{m=1}^{J} c_m r_m y\right) - \sum_{m=1}^{J} b_{im} r_m r_n - c_i r_i y$.

It is well known that a sufficient and necessary condition for uniqueness of a system of $J_0^+ - 1$ linear equations with $J_0^+ - 1$ unknowns is that the matrix D is nonsingular. And a square matrix is nonsingular if and only if its determinant is nonzero.

The determinant of the diagonal matrix D is the product of its diagonal elements. In the case above, $\det(D) = \prod_{i \in \Im^+} r_i$. Because the subset $\Im^+$ only includes those goods that exhibit strictly positive observed shares, we know that $r_i \neq 0$ for all $i \in \Im^+$. As a consequence, matrix D is nonsingular. The nonsingularity of D establishes that there is a unique vector of a's associated with each good that solves the system of equations $s_i^{true} = s_i^A(r, y; \theta, \Im)$ for $i \in \Im^+$. Once the shares of the $J_0^+ - 1$ inside goods are matched, the share of the outside good will automatically be matched as a consequence. The model is therefore able to equate predicted and actual budget shares for all goods (both inside and outside) with strictly positive observed shares.

We have already noted that Proposition 3.1 establishes that the model is able to match the shares of those goods with zero observed demand. Thus, as the proposition claims, the model can match, under specification A, any vector of observed budget shares. ■

Proposition 3.3 *The more general budget share function* $s_i(r, y; \theta, \Im)$ *can match, under specification B, any vector of observed budget shares. Furthermore, there exists a unique set of* θ_1 *parameters that matches the subset* $\Im_0^+$ *of goods with strictly positive observed demand.*

Proof. The proof for specification B follows similar lines to the one for specification A. Proposition 3.1 establishes that the model explicitly matches predicted and observed demand for all goods that exhibit zero observed shares and hence we can proceed by restricting our analysis to the subset of remaining goods in $\Im$.

We begin by showing that the system of J_0^+ equations $s_i^{true} = s_i^B(r, y; \theta, \Im)$ for $i \in \Im_0^+$ has, under specification B, exactly one solution $a(r, y, s^{true}; \theta_2, \Im)$ that equates the shares predicted by the model to the observed shares. Note that in contrast to the number of equations in specification A, under specification B the number of equations is J_0^+ as the outside option is modelled explicitly. In order to show the above proposition, we define the element-by element function $\delta_i = f_i(r, y; \theta, \Im)$. Conditional on the vectors r, y and on the set of θ_2 parameters, it is entirely equivalent to equate the shares predicted by the model and the observed shares using directly the vector $a = (a_0, a_1, \ldots, a_{J^+})'$ or indirectly the vector $\delta = (\delta_0, \delta_1, \ldots, \delta_{J^+})'$ as long as the inverse function $a_i = f_i^{-1}(r, y; \delta, \theta_2, \Im)$ exists and is unique.

Let $\delta_i = a_i r_i y + \sum_{m=0}^{J} b_{im} r_m r_n y^2$. It is clear that given this functional form for δ_i, it is always possible to recover a_i uniquely. Furthermore, we can now rewrite the share function for specification B as follows:

$$s_i^B(r, y; \theta, \Im) = \frac{a_i r_i y + \sum_{m=0}^{J} b_{im} r_m r_i y^2}{\sum_m a_m r_m y + 1/2 \sum_{m=0}^{J} \sum_{n=0}^{J} b_{mn} r_m r_n y^2} = \frac{\delta_i}{\sum_{m=0}^{J} \delta_m} = s_i^B(\delta).$$

The share function is homogenous of degree zero in the δ's. However, for specification B, θ_2 includes, in addition to the $[b_{ij}]$ parameters, a_0 for identification purposes. Note that being conditional on a_0 (together with r, y and the remaining θ_2 parameters) corresponds to being conditional on δ_0. Our problem reduces itself to a system of $J_0^+ - 1$ inside goods equations $s_i^{true} = s_i^B(\delta)$ for $i \in \Im^+$. We begin by noting that this system is linear in the vector $\delta = (\delta_1, \ldots, \delta_{J^+})'$ and can be rewritten as $D\delta = G$, where D denotes a $(J_0^+ - 1) \times (J_0^+ - 1)$ matrix with diagonal elements $d_{ii} = 1 - s_i^{true}$ and cross-diagonal elements $d_{ij} = -s_i^{true}$, and finally G denotes a $J_0^+ - 1$ vector with elements $g_i = s_i^{true} \delta_0$.

It is well know that a sufficient and necessary condition for uniqueness of a system of $J_0^+ - 1$ linear equations with $J_0^+ - 1$ unknowns is that the matrix D is nonsingular. A square matrix is nonsingular if and only if its determinant is nonzero. A property of the determinant is that adding a multiple of one row to another row, or a multiple of one column to another column, does not affect its value. In particular, consider the following sequence of linear operations: (1) subtract the first column to all remaining columns, and (2) add to the first row all other rows. This sequence yields a triangular matrix with the first diagonal element given by $d_{11} = 1 - \sum_{i \in \Im^+} s_i^{true}$ and the remaining diagonal elements given by $d_{ii} = 1$. The determinant of a triangular matrix equals the product of its diagonal elements: $\det(D) = \prod_{i \in \Im^+} d_{ii}$, which is nonzero because $\sum_{i \in \Im^+} s_i^{true} < 1$ for an outside good with a strictly positive share. If this was not the case, the argument still follows if we renormalize the parameters in terms of an inside good with a strictly positive market share.

The nonsingularity of D establishes that there is a unique vector valued function δ that solves the system of equations $s_i^{true} = s_i^B(\delta)$ for $i \in \Im^+$. We can then, conditional on r, y and θ_2, invert each δ_i and recover the unique vector a. Once the shares of the $J_0^+ - 1$ inside goods are matched, the share of the outside good will automatically be matched as a consequence. The model is therefore able to equate predicted and actual budget shares for all goods (both inside and outside) with strictly positive observed shares.

We have already noted that Proposition 3.1 establishes that the model is able to match the shares of those goods with zero observed demand. Thus, as the proposition claims the model can match, under specification B, any vector of observed budget shares. ■

Arbitrary Elasticities

Given that the model is capable of matching observed with predicted budget shares, we proceed by investigating if the set of θ_2 parameters are such that the model is able to also assume arbitrary values for the predicted elasticities - which concludes the flexibility result.

The flexibility properties will depend on the specification under consideration as there is a trade-off involving the choice of specification. We will see that specification A has the advantage of the derived income-elasticities being Diewert flexible. In contrast, specification B has the advantage of *all* the derived price-elasticities being Diewert flexible - not only those that refer to the inside goods, but also the ones that refer to the outside option. As a consequence, the choice of specification will typically be application-driven. Specification A is particularly suitable for real-world applications where the importance of the outside good is relatively small and where demand exhibits important income effects, whereas specification B is particularly suitable for applications where a premium is placed on the flexibility of the estimated price substitution patterns for *all* options. Naturally, the development of a specification that is able to aggregate the good properties of the two specifications presented here is a very interesting potential area for future research.

Proposition 3.4 *There exists a set of θ parameters such that the* ***more general budget share function*** $s_i(r, y; \theta, \Im)$ *can match, under specification A,* *any vector of budget shares, any vector of income elasticities for the inside goods*, and finally *any matrix of own- and cross-price elasticities for the inside goods.*

Proof. In order to establish Proposition 3.4, we want to show that we can choose the set of parameters θ so that the model satisfies simultaneously the following equations:

$$\begin{aligned} s_i^{true} &= s_i(r, y; \theta, \Im) \\ \eta_i^{true} &= \eta_i^s(r, y; \theta, \Im) \\ \varepsilon_{ij}^{true} &= \varepsilon_{ij}^s(r, y; \theta, \Im), \end{aligned}$$

where s_i^{true}, η_i^{true} and ε_{ij}^{true} denote any true vector of budget shares, any true vector of income elasticities for the inside goods, and finally any true matrix of own- and cross-price elasticities for the inside goods.

Proposition 3.2 establishes that the more general budget share function $s_i(r, y; \theta, \Im)$ can match, under specification A, any vector of observed budget shares. It establishes also that, for any set of $[b_{ij}]$ and $[c_i]$ parameters (θ_2), there exists a unique set of $[a_i]$ parameters (θ_1) that matches the subset $\Im_0^+$ of goods with strictly positive observed demand. Given such solution for θ_1, we want to show that the θ_2 parameters can be chosen simultaneously to equate the inside goods' *i*) own- and cross-price elasticities as well as *ii*) income elasticities predicted by the model to the observed ones.

We begin by investigating the ability of the model to match predicted with true income-elasticities for inside goods. Since the model can match any vector of budget shares, the income-elasticities predicted by the model are given by:

$$\eta_i^s(r, y; \theta, \Im) = -1 + \frac{c_i r_i}{c_0 s_i^{true} + \sum_{m=1}^{J} c_m r_m s_i^{true}},$$

for $i \in \Im^+$. It is easy to show that the system of $J_0^+ - 1$ inside goods equations $\eta_i^{true} = \eta_i^s(r, y; \theta, \Im)$ is linear in the $[c_i]$ parameters and can be rewritten as follows: $Dc = G$, where D denotes a $\left(J_0^+ - 1\right) \times \left(J_0^+ - 1\right)$ matrix with diagonal elements $d_{ii} = r_i s_i^{true}\left(\eta_i^{true} + 1\right) - r_i$ and cross-diagonal elements $d_{ij} = r_j s_i^{true}\left(\eta_i^{true} + 1\right)$, and finally G denotes a $J_0^+ - 1$ vector with elements $g_i = c_0 s_i^{true}\left(\eta_i^{true} + 1\right)$.

It is well know that a sufficient and necessary condition for uniqueness of a system of linear $J_0^+ - 1$ equations with $J_0^+ - 1$ unknowns is that the matrix D is nonsingular. A square matrix is nonsingular if and only if its determinant is nonzero. We begin by notting that the matrix D can rewritten as the product of two square matrices: $D = D_1 D_2$, where D_1 denotes a $\left(J_0^+ - 1\right) \times \left(J_0^+ - 1\right)$ matrix with diagonal elements $d_{ii}^1 = s_i^{true}\left(\eta_i^{true} + 1\right) - 1$ and cross-diagonal elements $d_{ij} = s_i^{true}\left(\eta_i^{true} + 1\right)$, while D_2 denotes a $\left(J_0^+ - 1\right) \times \left(J_0^+ - 1\right)$ diagonal matrix with diagonal entries $d_{ii}^2 = r_i$. The determinant of the product of two square matrices is the product of the individual determinants.

In order to compute the determinant of D_1, we note that a property of the determinant is that adding a multiple of one row to another row, or a multiple of one column to another column, does not affect its value. In particular, consider the following sequence of linear operations: (1) subtract the first column to all remaining columns, and (2) add to the first row all other rows. This sequence yields a triangular matrix with the first diagonal element given by $d_{11} = -1 + \sum_{i \in \Im^+} s_i^{true}\left(\eta_i^{true} + 1\right)$ and the remaining diagonal elements given by $d_{ii} = -1$. The determinant of a triangular matrix equals the product of its diagonal elements: $\det\left(D_1\right) = \prod_{i \in \Im^+} d_{ii}^1 = (-1)^{J_0^+ - 1} s_0^{true}\left(\eta_0^{true} + 1\right)$, which is nonzero for an outside good with a strictly positive share and a *demand* income elasticity different from zero.[11] If this was not the case, the argument still follows if we renormalize the parameters in terms of an inside good with a strictly positive market share and a *demand* income elasticity different from zero. The determinant of the diagonal matrix D_2 is the product of its diagonal elements: $\det\left(D_2\right) = \prod_{i \in \Im^+} d_{ii}^2 = \prod_{i \in \Im^+} r_i$. Because the subset $\Im^+$ only includes those goods that exhibit strictly positive observed shares, we know that $r_i \neq 0$ for all $i \in \Im^+$. As a consequence, matrix D_2 is nonsingular. The determinant of matrix D is therefore the product of the individual determinants: $\det(D) = (-1)^{J_0^+ - 1} s_0^{true}\left(\eta_0^{true} + 1\right) \prod_{i \in \Im^+} r_i$, which is nonzero for an outside good with a strictly positive share and a *demand* income elasticity different from zero. Again, if this was not the case, the argument still follows if we renormalize the parameters in terms of an inside good with a strictly positive market share and a *demand* income elasticity different from zero.

The nonsingularity of D establishes that there is a unique vector of the $[c_i]$ parameters, independent of the set of $[b_{ij}]$ parameters, that solves the system of equations $\eta_i^{true} = \eta_i^s\left(r, y; \theta, \Im\right)$ and matches the $J_0^+ - 1$ vector of predicted to true income-elasticities associated with the inside goods.

We proceed by considering the price-elasticities. Under the asymmetry assumption of the more general budget share function, we show that the model predicts elasticities that

[11] Recall the relationship between the budget share and the demand income-elasticities. $\eta_i^d = \eta_i^s + 1$. A budget share income-elasticity $\eta_i^{s,true} = -1$ is equivalent to a demand income-elasticity $\eta_i^{d,true} = 0$.

are able to assume arbitrary $\left(J_0^+ - 1\right)^2$ values for the inside goods. We start by noting that since the model can match any vector of budget shares, the own- and cross-price elasticities predicted by the model are given by:

$$\varepsilon_{ij}^{s}\left(r, y; \theta, \Im\right) = -1\left(j = i\right) - \frac{b_{ij} r_j r_i \left(s_i^{true} y\right)^{-1} - c_j r_j}{c_0 + \sum_{m=1}^{J} c_m r_m}.$$

It remains to show that there is a vector of $[b_{ij}]$ parameters that solve all the inside goods $\varepsilon_{ij}^{true} = \varepsilon_{ij}^{s}\left(r, y; \theta, \Im\right)$ equations simultaneously for any $i, j \in \Im^+$. The problem is that the system of price-elasticities equations depends on the $[c_i]$ parameters. However, we have shown that the vector of $[c_i]$ parameters that matches the $J_0^+ - 1$ vector of predicted to true income-elasticities associated with the inside goods is independent of the set of $[b_{ij}]$ parameters and depends only on the vectors r, s_i^{true} and η_i^{true}. As a consequence, we can solve the joint system of equations by recursion. In particular, we have that setting:

$$\begin{aligned} b_{ii} &= \left(s_i^{true} \eta_i^{true} + s_i^{true} - 1 - \varepsilon_{ij}^{true}\right) \frac{s_i^{true} y}{r_i r_i} f\left(r, s^{true}, \eta^{true}\right) \\ b_{ij} &= \left(s_j^{true} \eta_j^{true} + s_j^{true} - \varepsilon_{ij}^{true}\right) \frac{s_i^{true} y}{r_j r_i} f\left(r, s^{true}, \eta^{true}\right), \end{aligned}$$

for any $i, j \in \Im^+$ matches the $\left(J_0^+ - 1\right)^2$ vector of predicted to true price-elasticities for the inside goods, where $f\left(r, s^{true}, \eta^{true}\right)$ denotes the expression $\left(c_0 + \sum_{m=1}^{J} c_m r_m\right)$ after substitution the $[c_i]$ parameters for the corresponding implicit solution that matches the vector of true income-elasticities for the inside goods. ■

Proposition 3.5 *There exists a set of θ parameters such that the* ***more general budget share function*** $s_i\left(r, y; \theta, \Im\right)$ *can match, under specification B,* <u>*any vector of budget shares*</u> *and* <u>*any matrix of own- and cross-price elasticities*</u>.

Proof. In order to establish Proposition 3.5, we want to show that we can choose the set of parameters θ so that the model satisfies simultaneously the following equations:

$$\begin{aligned} s_i^{true} &= s_i\left(r, y; \theta, \Im\right) \\ \varepsilon_{ij}^{true} &= \varepsilon_{ij}^{s}\left(r, y; \theta, \Im\right), \end{aligned}$$

where s_i^{true} and ε_{ij}^{true} denote any true vector of budget shares and any true matrix of own- and cross-price elasticities, respectively. In other words, we want to show that at an arbitrary point (r_i, y_i), for any true s_i^{true} and ε_{ij}^{true} where $i,j \in \Im_0^+$ and $j \neq i$, we can always choose the $[a_i]$ and $[b_{ij}]$ parameters that matches the predicted budget shares and price elasticities to the observed ones. The point of approximation is taken, without loss of generality, to be $r_i = y_i = 1$ for $i \in \Im_0^+$. At this point of approximation, we want to show that the model satisfies the following equations:

$$\begin{aligned} s_i^{true} &= \frac{a_i + \sum_{m\in\Im_0^+} b_{im}}{\sum_{m\in\Im_0^+} a_m + \sum_{m\in\Im_0^+}\sum_{n\in\Im_0^+} b_{mn}} \\ \varepsilon_{ij}^{true} &= -1\,(j=i) - \frac{(b_{ij}/s_i^{true}) - \sum_{m\in\Im_0^+} b_{mj}}{\sum_{m\in\Im_0^+} a_m + \sum_{m\in\Im_0^+}\sum_{n\in\Im_0^+} b_{mn}} + s_j^{true}. \end{aligned}$$

In the spirit of Proposition 3.3, we begin the proof of Proposition 3.5 by defining $\delta_i = a_i + \sum_{m\in\Im_0^+} b_{im}$. We know the budget share equations can be solved by setting $\delta_i = s_i^{true}/s_0^{true}$. In order to see why this is the case, note that $\sum_{m\in\Im_0^+} \delta_m = \sum_{m\in\Im_0^+} s_m^{true}/s_0^{true} = 1/s_0^{true}$ and, as a consequence, we have that:

$$s_i^{true} = \frac{\delta_i}{\sum_{m\in\Im_0^+} \delta_m} = \delta_i s_0^{true},$$

which rearranging implies that $\delta_i = s_i^{true}/s_0^{true}$ as required.

Having matched the observed budget shares by choosing the $[\delta_i]$ parameters appropriately, we know focus on matching the true price elasticities, which can be rewritten as follows:

$$\varepsilon_{ij}^{true} = -1\,(j=i) - \frac{(b_{ij}/s_i^{true}) - \sum_{m\in\Im_0^+} b_{mj}}{(1/s_0^{true})} + s_j^{true},$$

since $\sum_{m\in\Im_0^+} a_m + \sum_{m\in\Im_0^+}\sum_{n\in\Im_0^+} b_{mn} = \sum_{m\in\Im_0^+} \delta_m = 1/s_0^{true}$. This constitutes a system of equations in the unknown $[b_{ij}]$ parameters. In order to see that we can always choose the $[b_{ij}]$ parameters that matches the predicted price elasticities to the true ones, impose the J_0^+ restrictions $\sum_{m\in\Im_0^+} b_{mj} = 0$ which imply $\sum_{m\in\Im_0^+}\sum_{n\in\Im_0^+} b_{mn} = 0$. After imposing these restrictions, we can write:

$$\varepsilon_{ij}^{true} = -1\,(j=i) - \frac{(b_{ij}/s_i^{true})}{(1/s_0^{true})} + s_j^{true},$$

which rearranging implies that setting:

$$\begin{aligned} b_{ii} &= \frac{s_i^{true}}{s_0^{true}}\left(-1 - \varepsilon_{ii}^{true} + + s_i^{true}\right) \\ b_{ij} &= \frac{s_i^{true}}{s_0^{true}}\left(-\varepsilon_{ij}^{true} + s_j^{true}\right), \end{aligned}$$

for any $i, j \in \Im_0^+$ solves the $\left(J_0^+\right)^2$ equations associated with all price elasticities and matches the model predictions to the observed ones. Note the solution does not depend on the $[a_i]$ parameters. The result is that the model is able to assume arbitrary values for the predicted elasticities even imposing these normalizations.

In the last step of the proof, we note that the $[a_i]$ parameters can always be choosen to match the observed budget shares as follows:

$$a_i = \frac{s_i^{true}}{s_0^{true}} - \sum_{m \in \Im_0^+} b_{im} = \frac{s_i^{true}}{s_0^{true}} \left(1 + \sum_{m \in \Im_0^+} \varepsilon_{im}^{true}\right),$$

where the last equality is derived after substituting the expressions for b_{im} in terms of the true budget shares and price elasticities. ■

3.2 Identification and Estimation

The identification of the different parameters requires a set of normalizations we now describe. We have already noted that the budget share functions are homogeneous of degree zero in the parameters and hence are identified up to a scalar. Without loss of generality, we normalize c_0 for specification A and b_{00} for specification B.

The full identification of specification B requires, however, some additional normalizations motivated by data issues. Because the outside option is modelled explicitly under this specification, we need to discuss the identification of the price-substitution patterns from and to the outside good. We typically observe no variation in the price of the latter option. As a consequence, we can not expect to be able to identify b_{00}. This constitutes the reason why we choose to normalize b_{00} in order to fix the scale of the parameters in specification B. Moreover, and for exactly the same reason, we can not also expect to be able to identify the $[b_{i0}]$ parameters, i.e. the parameters that define the degree of substitution towards product i when the price of the outside option varies. We choose to impose the $[b_{i0} = b_{0i}]$ symmetry assumption for all $i \neq 0$. The $[b_{0i}]$ parameters define the degree of substitution towards the outside option when the price of inside product i varies. Given the typical price variation of an inside good, we can expect to identify the later parameters and, therefore, the imposition of symmetry seems a natural restriction.

Given these restrictions, the identification of the remaining parameters is standard given a large enough sample. The $[a_i]$ parameters in θ_1 are identified from variation in the budget shares across the different goods. Having identified the $[a_i]$ parameters, the taste parameters in vector β are identified from variations in the observed product characteristics. The $[b_{ii}]$ and $[b_{ij}]$ parameters in θ_2 are identified from variation in prices, with the identification of the former relying in variation from own-prices, and the latter in variation from competitors prices. Finally, the $[c_i]$ parameters, in specification A, are identified from variations in income. Of course, in many instances it will be appropriate to use instruments rather than (say) the variation in the actual prices to empirically identify the model's parameters, an approach we discus further below.

The estimation algorithm that we propose is based in Berry (1994) and BLP, and includes four steps that we now describe.[12] Although the sample dimension will not impact the general setup of the algorithm, it will have an affect on the set of θ_2 parameters. If the number of goods is relatively small so that the dimensionality problem does not constitute an issue, θ_2 will include the set of $[b_{ij}]$ and $[c_i]$ parameters in specification A, and a_0 and $[b_{ij}]$ in specification B, whereas if, on the other hand, the number of goods yields a too great number of parameters to be estimated, a mapping needs to be defined and θ_2 will include instead the vector α in specification A (plus a_0 in specification B). Although we already hinted the reason why a_0 is included in the set of θ_2 parameters under specification B, we will discuss the issue further below.

Step 1 *Set initial values for the parameters in θ_2.*

The choice of the initial values for θ_2 is always arbitrary and ideally would not have an impact on the final parameter estimates. Unfortunately, this may fail to happen on the class of highly nonlinear demand models for differentiated products that involve simulation - of which BLP constitutes the most extensively used example in empirical exercises. Despite the work on identification by Berry *et al.* (2004), Berry and Haile (2008), and Fox and Gandhi (2008), numerical problems difficulties can arise. Knittel and Metaxoglou (2008) explore the issue of potential multiple local minima using BLP's estimator and find that convergence may occur at a number of local extrema, at saddles and in regions of the objective function where the first-order conditions are not satisfied.

Our estimation procedure, although motivated by and closely related to Berry (1994) and BLP, will prove to be globally convex and hence avoid some of the numerical issues related to local extrema - as our monte carlo simulations below demonstrate. Furthermore,

[12] The marginal utility of income under specification A is given by the price index $c_0 + \sum_{m=1}^{J} c_m r_m$. If, following Pinkse and Slade (2004), the researcher is willing to assume this price index to be constant, then the specification can be estimated by a simple linear instrumental regression approach.

we propose an alternative algorithm in the spirit of Davis(2006a) and Dubé *et al.* (2009) that improves the rate of convergence substantially.

Step 2 *Computation of the budget shares conditional on the θ_2 parameters*

Propositions 3.2 and 3.3 establish that, conditional on the θ_2 parameters, the more general budget share function can match any vector of observed budget shares. In Step 2, we solve for the unique θ_1 parameters that, given each guess for the set of parameters in θ_2, match the observed s_i^{true} with the predicted $s_i(r, y; \theta, \Im)$ budget shares in *each* market or time period. For specification A, the solution to this problem is analytical:

$$a_i = \left(s_i^{true}/r_i\right)\left(\gamma_0 y + \sum_{m=1}^{J} \gamma_m r_m y\right) - \sum_{m=1}^{J} b_{im} r_m - c_i y. \tag{40}$$

For specification B, we can not solve for the unique θ_1 parameters analytically and hence have to solve it numerically. Proposition 3.3 establishes that, even in the most general of the cases in which no restrictions are imposed on the parameters, there is a unique vector valued function a that solves the system of equations $s_i^{true} = s_i^B(r, y; \theta, \Im)$ for $i \in \Im_0^+$. From the budget share function it is clear that this solution must be such that $N_i^B(r, y; \theta, \Im) = a_i r_i y + \sum_{m=0}^{J} b_{im} r_m r_i y^2$ is either positive or negative for all $i \in \Im_0^+$. If $N_i^B(r, y; \theta, \Im)$ was to be positive for some i and negative for others, the model would not be able to equate the predicted shares to strictly positive observed shares. As a consequence, we can without any loss of generality work with the element-by element function:

$$N_i^B(r, y; \theta, \Im) = a_i r_i y + \sum_{m=0}^{J} b_{im} r_m r_i y^2 = \exp(\delta_i) = N_i^B(\delta, \Im). \tag{41}$$

We have already noted that, conditional on the vectors r, y and θ_2, it is entirely equivalent to equate the predicted shares by the model and the observed shares using the vector $a = (a_0, a_1, \ldots, a_{J+})'$ directly or the vector $\delta = (\delta_0, \delta_1, \ldots, \delta_{J+})'$, as long as the inverse function $a_i = f_i^{-1}(r, y; \delta, \theta_2, \Im)$ exists and is unique. We can then rewrite the share function for specification B as follows:

$$s_i(r, y; \theta, \Im) = \frac{N_i^B(r, y; \theta, \Im)}{\sum_{m=0}^{J} N_m^B(r, y; \theta, \Im)} = \frac{\exp(\delta_i)}{\sum_{m=0}^{J} \exp(\delta_m)} = s_i(\delta, \Im). \tag{42}$$

This budget share function has two characteristics that are important to discuss. First, it is independent of the sign of $N_i^B(r, y; \theta, \Im)$. Suppose the marginal utility function $N_i^B(r, y; \theta, \Im)$ is negative: in that case, we can work instead with $-\exp(\delta_i)$, for all $i \in \Im_0^+$, which does not impact the ratio in the budget share function. Second, conditional on the θ_2 parameters, it is homogeneous of degree zero in the vector δ and hence identification requires a normalization. Without loss of generality, we choose to normalize a_0 (which, conditional on the θ_2 parameters translates into a normalization of δ_0) by including it on the set of θ_2 parameters. Note that this normalization is only for Step 2 purposes. Of course, we could have, alternatively, normalized a_0 to start with (and not b_{00} as we did). However, given the typical data variation we observe for the outside option in real-world applications, we can expect to be able to identify a_0 although not b_{00}.

We now address the numerical alternatives we propose to solve for the unique vector valued function δ that solves the system of equations $s_i^{true} = s_i^B(\delta, \Im)$ for $i \in \Im_0^+$. The first alternative involves using BLP's contraction algorithm.

Proposition 3.6 *The solution to the problem of equating observed and predicted budget shares for the subset $\Im_0^+$ of goods with strictly positive shares may be found recursively (for each time period or market) using the operator $g(\delta_i)$ defined pointwise by:*

$$g(\delta_i) = \delta_i + \ln\left(s_i^{true}\right) - \ln\left[s_i(\delta, \Im)\right],$$

for every $i \in \Im^+$. In other words, the operator $g(\delta_i)$ is a contraction mapping with modulus less than one.

Proof. In order to establish the above proposition, we apply the contraction mapping theorem in BLP to specification B. They prove that the operator $g(\delta_i)$ is a contraction mapping with modulus less than one if it exhibits the following properties: $g(\delta_i)$ is continuously differentiable, with $\forall i$ and j, $\partial g(\delta_i)/\partial\delta_j \geq 0$ and $\sum_{j\in\Im^+} \partial g(\delta_i)/\partial\delta_j < 1$.

We need to show that operator above satisfies the properties of their theorem. We begin by noting that the differentiability of the budget share function ensures the function $g(\delta_i)$ is differentiable. In what the monotonicity properties is concerned, we have that $g(\delta_i)/\partial\delta_j = s_j(\delta, \Im) > 0$ $\forall i$ and j. Finally, it is clear that $\sum_{j\in\Im^+} \partial g(\delta_i)/\partial\delta_j = \sum_{j\in\Im^+} s_j(\delta, \Im) = 1 - s_0(\delta, \Im) < 1$ if the outside good exhibits a share strictly greater than zero.

This implies that we can solve for the vector δ recursively. The initial guess for δ_i is used to evaluate $g(\delta_i)$ and obtain a new estimate δ_i. The process is then repeated until convergence is achieved for δ_i, yielding a match between predicted and observed budget shares. We can then, conditional on the vectors r, y and θ_2, invert each δ_i and recover the unique vector a. ■

The numerical properties of BLP's contraction mapping approach is explored by Dubé *et* *al.* (2009). They use the contraction mapping theorem to show that its rate of convergence is in general slow. In order to address such problem, we propose an alternative, based on Quasi-Newton methods, that provides a faster rate of convergence in a manner which ensures global convergence.

Proposition 3.7 *The solution to the problem of equating observed and predicted budget shares for the subset $\Im_0^+$ of goods with strictly positive shares may be found as the unique solution to the following strictly convex optimization problem (for each time period or market):*

$$\min_{\delta_i \in \Im^+} \log\left(\sum\nolimits_{m\in\Im^+} N_m^B(\delta, \Im)\right) - \sum\nolimits_{m\in\Im^+} \delta_m s_m^{true}.$$

Moreover, the solution to this minimization problem can be found numerically using a Quasi-Newton method such as Davidson-Fletcher-Powell (DFP) with exact line search. Such an iterative algorithm provides a super-linear rate of convergence.

Proof. The proof of the above proposition follows directly the first-order conditions to the optimization problem, which yield $s_i(\delta, \Im) = s_i^{true}$. It remains to show that the problem is convex. It is well known that the problem is convex if the second partial derivatives matrix of the problem's objective function is positive semidefinite. For the problem above, the second partial derivatives matrix has diagonal entries given by $s_i(\delta, \Im)(1 - s_i(\delta, \Im))$ and non-diagonal entries given by $-s_i(\delta, \Im)\, s_j(\delta, \Im)$ for all $i, j \in \Im^+$.

A sufficient condition for positive definiteness (and automatically positive semidefiniteness) of a symmetric matrix is that all the diagonal entries are positive and there is a dominant diagonal (each diagonal entry is greater than the sum of the absolute values of all other entries in the same row). The properties of the budget share function ensure that both conditions are satisfied for the second partial derivatives matrix of the problem's objective function. The former condition, that the diagonal entries are positive, is satisfied if each good exhibits a share strictly between zero and one. In order to see why the latter condition, that each diagonal entry is greater than the sum of the absolute values of all other entries in the same row, is also satisfied, note that $s_i(\delta, \Im)(1 - s_i(\delta, \Im)) > \sum_{j\neq i} s_i(\delta, \Im)\, s_j(\delta, \Im)$ for all $i, j \in \Im^+$ simplifies to $s_i(\delta, \Im)(1 - s_i(\delta, \Im)) > s_i(\delta, \Im)(1 - s_i(\delta, \Im) - s_0(\delta, \Im))$, which is true since $s_i(\delta, \Im)\, s_0(\delta, \Im) > 0$. Thus, the objective function is strictly convex.

Given the strictly convexity of the problem, any local minimum will also be a global minimum of the problem and hence the solution will be unique. In addition, it is well known that suitably chosen Quasi-Newton methods will be globally convergent for convex problems. For example, Powell (1971, 1972) establishes that if an objective function is convex, then the DFP method with exact line search converges globally, with super-linear convergence. ■

Let, in both specifications, $a\left(r, y, s^{obs}; \theta_2, \Im\right)$ denote the solution vector of the a's that ensure that the observed s_i^{obs} and the predicted $s_i\left(r, y; \theta, \Im\right)$ budget shares are equated.

Step 3 *Computation of the structural error*

Having solved for the unique θ_1 parameters that match, in *each* market or time period, the observed budget shares s_i^{true} with the predicted ones $s_i\left(r, y; \theta, \Im\right)$, we proceed by running a Berry (1994) style regression on the following relationship:

$$a_{it}\left(r, y, w^{obs}; \theta_2, \Im\right) = \textstyle\sum_{i=1}^{K} \beta_k x_{kit} + \xi_{it}, \tag{43}$$

and obtain estimates for the vector β of parameters and for the $[\xi_{it}]$ unobserved characteristics. The latter estimates will be a function of both β and θ_2 vectors of parameters and will be used to compute the objective function of a Generalized Method of Moments (GMM) procedure.

Step 4 *Estimation of the θ_2 parameters*

Estimate the θ_2 parameters by GMM. The approach relies on an identifying restriction on the distribution of the true unobserved characteristics and is based on the sample analogue to the population condition.

The standard identifying restriction states that, at the true values of the parameters, $\theta^{true} = \left(\theta_1^{true}, \theta_2^{true}\right)'$, the true unobserved characteristics are mean independent of a set of M instruments $Z_{it} = [z_{1it}, \ldots, z_{Mit}]$:

$$E\left[\xi_{it}\left(\theta^{true}\right)|Z_{it}\right] = 0. \tag{44}$$

Please note that other identifying restrictions would also enable the estimation of the model. In particular, given the typical panel structure of the data, an alternative assumption could incorporate the likelihood of the econometric error and the set of instruments to be more similar for a given product across time, than for those of different products. Please see BLP and Davis (2006a) for a more detailed analysis on this subject.

The above population moment condition can be used, akin to Hansen (1982), to render a method of moments estimator of θ^* by interacting the estimated unobserved characteristics with the set of instruments, and then search for the value of the θ parameters that set the sample analogues of the moment conditions as closed as possible to zero. Let $G_n(\theta)$ denote the sample analogues of the moment conditions:

$$G_n(\theta) = \frac{1}{n} \sum_{i=1}^{J} \sum_{t=1}^{T} \tilde{\xi}_{it}(\theta) \tilde{Z}'_{it}, \tag{45}$$

where for notational purposes $\tilde{\xi}_{it}(\theta) = \xi_{it}(\theta)\chi_{it}$, $\tilde{Z}_{it} = [z_{1it}\chi_{it}, \ldots, z_{Mit}\chi_{it}]$, and $\chi_{it} = 1$ if good i is sold in market t and zero otherwise. χ_{it} provides, thereby, a missing value indicator used to compute $n = \sum_{i=1}^{J} \sum_{t=1}^{T} \chi_{it}$.

Formally, the method of moments estimator for $\hat{\theta}$ is the argument that minimizes the weighted norm criterion of $G_n(\theta)$, for some weighting matrix A_n with rank at least equal to the dimension of θ:

$$\hat{\theta} = \arg\min_{\theta} \quad \|G_n(\theta)\|_{A_n} = G_n(\theta)' A_n G_n(\theta). \tag{46}$$

The strong non-linearity of the objective function requires a minimization routine. The non-linear search over θ can be simplified by making use of the fact that the first order conditions for a minimum of $\|G_n(\theta)\|_{A_n}$ are linear for the subset β of the θ_1 parameters of estimation in $\theta = (\theta_1, \theta_2)$. In particular, it is possible, given the standard instrumental variables results, to express the vector β as a function of θ_2, and limit the non-linear search over θ_2:

$$\hat{\beta} = \left(\tilde{X}'\tilde{Z}A_n^{-1}\tilde{Z}'\tilde{X}\right)^{-1} \tilde{X}'\tilde{Z}A_n^{-1}\tilde{Z}'a\left(r, y, s^{obs}; \theta_2, \Im\right), \tag{47}$$

where $\tilde{X}$ denotes the $n \times K$ matrix of $[x_{kjt}\chi_{it}]$ observed characteristics, and $\tilde{Z}$ denotes the $n \times M$ matrix of $[z_{mit}\chi_{it}]$ instruments.

3.2.1 Standard Errors

In contrast to a model based on simulation, the GMM estimator for this model does not need to be corrected for simulation error and hence the standard formulae for a GMM estimator apply. Hansen (1982) establishes the formal conditions under which $\hat{\theta}$, the method of moments estimator, is consistent and asymptotically normal with bounded variance, consistently estimated as follows:

$$\sqrt{n}\left(\hat{\theta} - \theta^*\right) \sim N\left[0, \left(\widehat{\Gamma}'A_n\widehat{\Gamma}\right)^{-1} \widehat{\Gamma}'A_n\hat{\Phi}A_n\widehat{\Gamma}\left(\widehat{\Gamma}'A_n\widehat{\Gamma}\right)^{-1}\right], \tag{48}$$

where $\widehat{\Gamma}$ denotes a consistent estimator of the gradient of the objective function:

$$\widehat{\Gamma}' = \widehat{\Gamma_n\left(\hat{\theta}\right)} = \frac{1}{n}\sum_{i=1}^{J}\sum_{t=1}^{T}\left[\frac{\partial\tilde{\xi}_{it}\left(\hat{\theta}\right)\tilde{Z}'_{it}}{\partial\theta'}\right],$$

and $\hat{\Phi}$ denotes a consistent estimator of the variance-covariance matrix of the moment conditions $\hat{\Phi} = Var\left[G_n\left(\hat{\theta}\right)\right]$. The optimal weighting matrix is proportional to Φ^{-1}, giving less weigh to those moments with a higher variance.

We may follow the traditional literature on the continuous demand for differentiated products, wherein asymptotic arguments are assumed to work in the number of markets (or time periods) or, alternatively, we may follow the BLP assumption that asymptotic arguments work in the number of products. Furthermore, the structure of the matrix $Var\left[G_n\left(\hat{\theta}\right)\right]$ will depend on our assumptions about the covariance structure of the unobserved product characteristics. We may, for example, assume that they are independent across markets, but allow for correlation across products within a given market or, alternatively, we may assume that they are correlated across markets and independent across products. Naturally, it is also possible - and probably desirable - to assume richer error variance structures. See, for example, Davis (2002).

3.3 Consumer Heterogeneity and Welfare Analysis

An important advantage of a structural model is that it can be used for welfare analysis. In this section, we extend the model to account for consumer heterogeneity in order for the indirect utility function to constitute a *social* indirect utility function. The model can easily account for such heterogeneity by allowing the set of parameters θ_2 to be consumer-specific. In theory, consumer heterogeneity may extend any specification of the model. In practice, however, since this extension increases the vector of parameters of interest, it may make sense to introduce consumer heterogeneity into a model where the θ_2 parameters are initially mapped onto the characteristics space.

Consider a population of I households, where each household h is described by income y_h and a set of observed and unobserved consumer characteristics, d_h and u_h. Let the preferences of household h be given by $H_h\left(r, y_h; \theta_h, \Im\right)$, where θ_h denotes the set of individual-specific parameters. Following the random-coefficients discrete-choice literature, we may assume that θ_h can be decomposed into a set θ_1 that will capture the mean aggregate budget shares and is common to all households, and a set θ_{2h} that will capture the price and income substitution patterns and is individual-specific. In particular, let the θ_{2h} parameters be defined as functions of the observed and unobserved consumer characteristics, d_h and u_h:

$$\theta_{2h} = \theta_2 + \theta_2^d d_h + \theta_2^u u_h, \tag{49}$$

where θ_2^d and θ_2^u represent coefficients on the interactions with observed and unobserved individual attributes, respectively.

The social indirect utility function $H(r, y_1, \ldots, y_I; \theta, \Im) = (1/I) \sum_{h=1}^{I} H_h(r, y_h; \theta, \Im)$ will be, in general, a function of the vector of prices and the complete distribution of household income. If $H_h(r, y_h; \theta_h, \Im)$ has the properties of an indirect utility function then standard duality results imply that, under non-satiation, the demand system is easily obtained via Roy's identity. Let $w_{ih}(r, y_h; \theta_h, \Im)$ denote the budget share of good i for household h.

The introduction of consumer heterogeneity is not obtained costlessly, since there is no closed form expression for the market-level budget share, $w_i(r, y_h; \theta_h, \Im)$, that aggregates the I individual budget shares:

$$w_i(r, y_h; \theta_h, \Im) = \int w_{ih}(r, y_h; \theta, \Im)\, dP^*(y, d, u) = \int w_{ih}(r, y_h; \theta, \Im)\, dP_{yd}^*(y, d)\, dP_u^*(u), \tag{50}$$

where $P^*(y, d, u)$ denotes the population distribution of the consumer types (y_h, d_h, u_h) and the last equality is a direct consequence of the independence assumptions that can be made on (y, d) and u. Following Pakes (1986), Pakes and Pollard (1989), and McFadden (1989), we can approximate the above integral by simulation using, for example, a smooth estimator. Such computation requires drawing ns pseudo-random vectors of consumer characteristics from $P_{yd}^*(y, d)$ and $P_u^*(u)$ and simulate the aggregate budget shares by:

$$w_i(r, y_h; \theta_h, \Im, P^{ns}) = (1/ns) \sum_{h=1}^{ns} w_{ih}(r, y_h; \theta_h, \Im), \tag{51}$$

where P^{ns} denotes the empirical distribution of the simulation draws. A drawback of this type of solution is that it introduces simulation error, which, as Berry *et al.* (2004) point out, influences the asymptotic distribution of the GMM estimator and, therefore, needs to be explicitly taken it account. However, apart from these modifications, the introduction of consumer heterogeneity would not impact significantly the general setup of the estimation algorithm.

If the aggregate demand functions are generated, via Roy's identity, from a social indirect utility function, they have welfare significance and can be used to make welfare judgments via the standard welfare measurement techniques. By estimating the parameters of the demand functions, we have the required parameters of the social indirect utility function, which we can then easily invert - either algebraically or numerically - to derive the expenditure function - and compute *compensating* and *equivalent variations.* Of course this approach only makes sense if the estimated parameters satisfy the various restrictions that ensure an underlying model of utility.

Although the social indirect utility function will be, in general, a function of the vector of prices and the complete distribution of household income, Gorman (1953)'s seminal article on exact aggregation establishes, however, that when consumers have indirect utility of the Gorman form, aggregate demand can always be thought of as being generated by a normative representative consumer with indirect utility function $H(r, y; \theta, \Im) = (1/I) \sum_{h=1}^{I} H_h(r, y_h; \theta, \Im)$ defined in terms of y, the average income, regardless of the form of the social welfare function.

An indirect utility function for consumer h is said to be of the Gorman form if it can be written in terms of functions $d_h(r)$, which may depend on the specific consumer, and $k(r)$, which is common to all consumers:

$$H_h(r, y_h; \theta, \Im) = d_h(r) + k(r)\, y_h.$$

We now show that the specification A for $H(r, y; \theta, \Im)$ satisfies the Gorman polar form and, as a consequence, the preferences of our *fictional* representative consumer constitutes a measure of aggregate social welfare. As a consequence, for this particular specification, welfare judgments can be made without incorporating consumer heterogeneity.

Proposition 3.8 *The indirect utility function under specification A, $H^A(r, y; \theta, \Im)$, constitutes an admissible social indirect utility function for the normative representative consumer.*

Proof. Let the indirect utility function of each household h, $H_h^A(r, y_h; \theta, \Im)$, be given by a generalization (to the household-level) of specification A's indirect utility function:

$$H_h^A(r, y_h; \theta, \Im) = a_0 + \sum_{m=1}^{J} a_m r_m + 1/2 \sum_{m=1}^{J} \sum_{n=1}^{J} b_{mn} r_m r_n + c_0 y_h + \sum_{m=1}^{J} c_m r_m y_h.$$

It is easy to verify that $H_h^A(r, y_h; \theta, \Im)$ satisfies the Gorman form with $d_h(r) = a_0 + \sum_{m=1}^{J} a_m r_m + 1/2 \sum_{m=1}^{J} \sum_{n=1}^{J} b_{mn} r_m r_n$ and $k(r) = c_0 + \sum_{m=1}^{J} c_m r_m$. ■

3.4 Monte Carlo Experiment

In this section, we describe the data-generating process for a Monte Carlo experiment designed to analyze the convergence properties of the demand model for the different specifications proposed. We consider a setting with J goods in T markets and allow the econometrician to observe, in addition to price, $K = 5$ characteristics for each good-market combination. The observed characteristic k for good i in market t, x_{kit}, is drawn from a $(0, 1)$ uniform distribution. Because not all characteristics of a good are observed by the econometrician, we allow also for unobserved characteristics, ξ_{it}, which we draw from a $(-l_\xi, l_\xi)$ uniform distribution. The set of $[a_{it}]$ parameters are defined as $a_{it} = \beta_0 + \sum_{k=1}^{5} \beta_k x_{kit} + \xi_{it}$, with $\beta_0 = 50$ and $\beta_k = 1$ for all $k \geq 1$. With the variance of the observed characteristics fixed, l_ξ controls the 'noise-to-signal' ratio in the model. If l_ξ is *small*, we expect to require relatively smaller samples to consistently estimate the parameters than when compared with the *large* l_ξ case. The baseline case focus on $l_\xi = 0.50$, but we investigate the sensitivity of the estimates to other assumptions.

In order to simulate the endogeneity that arises from profit-maximizing price setting, we define prices to follow $p_{it} = \left|2.5 + \sum_{k=1}^{5} \beta_k x_{kit} + \xi_{it} + e_{it}\right|$, where e_{it} denotes a $(0, 1)$ uniform innovation. In order to deal with this endogeneity problem, we construct a number M instruments correlated with price, but not with the unobserved characteristics. The instruments are derived following $z_{mit} = \sum_{k=1}^{5} \beta_k x_{kit} + e_{it} + 0.5 g_{mit}$, with g_{mit} being drawn from a $(0, 1)$ uniform distribution. As we discuss below, we consider several sample designs in our Monte Carlo experiment in order to investigate the estimation properties of the model. However, for comparison purposes, we maintain the number of instruments constant across the sample designs. In particular, we consider a conservative number of intruments following a combination of Hausman and Bresnahan approaches, where observed product characteristics of a good in other markets become instruments for its price in a given market: $M = K[\min(T) - 1]$, where $\min(T)$ denotes the minimum of markets across the different sample designs.

The market-level income y_t is assumed to follow a $(200, 209)$ uniform distribution, which despite the narrow range exhibits a variance comparable to the price variable. The reason for such narrow range is simulation related. The model can be used to estimate substitution patterns from *any* real data. However, when we simulate data for a given set of parameters and repeatedly draw random prices and income (as in our Monte Carlo experiment), we can encounter numerical problems (e.g., simulated market shares being negative). A problem that would never face when working with data. The bounds on income were therefore chosen to ensure positive and adding-up consistent budget shares (the latter of particular importance for specification A). Again, this is not a problem working with real data - in that case we are estimating the parameters and real budget shares are never negative.

The variation in the set of choices available to consumers can provide important information about substitution patterns. We allowed goods to be missing at random from a given market in order to mimic the entry and exit behavior characteristic of typical real-world datasets. In particular, a good was assumed missing in market t if the realization of a standard continuous-uniform random variable was less than 0.2 subject to the constraint of having at least one inside good.

Finally, the numerical optimization over the structural parameters was performed using GAUSS's *sqpsolvemt* solver with user-supplied objective function exact derivatives.

3.4.1 Specification A

We consider two broad sample designs: one where the number of different goods in a market is small, but a reasonably large number of markets exist, and another with a large number of goods marketed on a small number of markets. We begin by addressing the small number of goods case and hence assume that the asymptotic arguments work in the number of independent markets. The true values of the parameters of the indirect utility function are assumed to be $b_{ii} = 40$, $b_{ij} = -10$ and $c_i = 2$ for all i, j. For expositional convenience, the symmetry assumption was imposed on the estimation of the $[b_{ij}]$ parameters.

Table 3.1 presents the means and standard errors (in parenthesis) of the GMM estimates for the θ_2 parameters across 50 sample experiments. All the results are conditional on $J = 3$ goods. Panel A explores the sensitivity of the estimates to the number of markets, with columns (i)-(iii) considering the following three cases: $T = 100$, $T = 200$ and $T = 400$. It seems that the algorithm performs reasonably well, with both price and income parameters converging to the true values even at small sample sizes, although in what the price parameters are concerned, it seems to be easier to identify own-price than cross-prices parameters (as the former have lower proportional standard errors).

Panel B explores the sensitivity of the algorithm to the variance of the unobserved characteristics ξ_{it} that controls the noise-to-signal ratio. We consider the following three cases: $l_\xi = 1.50$, $l_\xi = 0.50$ and $l_\xi = 0.00$, with column (iii) just reproducing the baseline case from panel A for comparison easiness. As expected, the results suggest that an increase in the variance of ξ_{it} (and consequent increase in the noise-to-signal ratio) deteriorates the performance of the algorithm - which implies that we require additional data in order to

TABLE 3.1

Monte-Carlo Results: Specification A - Small J Case

Panel A: Sensitivity to Number of Markets T*				
Parameter	True	(i): 100	(ii): 200	(iii): 400
b_{11}	40.00	40.13 (2.94)	39.60 (2.80)	39.89 (2.20)
b_{21}	-10.00	-10.11 (0.96)	-10.04 (0.72)	-10.11 (0.85)
b_{22}	40.00	39.92 (3.28)	39.66 (3.05)	40.06 (2.09)
b_{31}	-10.00	-10.00 (1.06)	-9.95 (0.89)	-9.95 (0.97)
b_{32}	-10.00	-9.69 (1.10)	-9.74 (0.99)	-9.83 (1.10)
b_{33}	40.00	40.72 (2.48)	40.52 (2.85)	40.39 (2.05)
c_1	2.00	2.00 (0.04)	1.99 (0.03)	1.99 (0.03)
c_2	2.00	2.00 (0.04)	1.99 (0.03)	2.00 (0.04)
c_3	2.00	2.00 (0.03)	1.99 (0.03)	2.00 (0.03)

* 50 Monte Carlo simulations for J=3, l_ξ=0.50 and starting values at θ_2^{true}.

Panel B: Sensitivity to Noise-to-Signal Ratio l_ξ*				
Parameter	True	(iv): 1.50	(iii): 0.50	(v): 0.00
b_{11}	40.00	39.69 (7.54)	39.89 (2.20)	40.00 (0.00)
b_{21}	-10.00	-10.33 (2.59)	-10.11 (0.85)	-10.00 (0.00)
b_{22}	40.00	39.93 (7.21)	40.06 (2.09)	40.00 (0.00)
b_{31}	-10.00	-9.84 (2.94)	-9.95 (0.97)	-10.00 (0.00)
b_{32}	-10.00	-9.52 (3.29)	-9.83 (1.10)	-10.00 (0.00)
b_{33}	40.00	41.35 (6.63)	40.39 (2.05)	40.00 (0.00)
c_1	2.00	1.98 (0.09)	1.99 (0.03)	2.00 (0.00)
c_2	2.00	2.00 (0.12)	2.00 (0.04)	2.00 (0.00)
c_3	2.00	1.99 (0.08)	2.00 (0.03)	2.00 (0.00)

* 50 Monte Carlo simulations for J=3, T=400 and starting values at θ_2^{true}.

Panel C: Sensitivity to Starting Values*				
Parameter	True	(vi): $(1/2)\theta_2^{true}$	(iii): θ_2^{true}	(vii): $2\theta_2^{true}$
b_{11}	40.00	39.83 (2.36)	39.89 (2.20)	40.02 (2.29)
b_{21}	-10.00	-10.11 (0.85)	-10.11 (0.85)	-10.11 (0.87)
b_{22}	40.00	39.82 (2.19)	40.06 (2.09)	40.20 (2.15)
b_{31}	-10.00	-9.95 (0.98)	-9.95 (0.97)	-9.94 (0.99)
b_{32}	-10.00	-9.83 (1.13)	-9.83 (1.10)	-9.83 (1.13)
b_{33}	40.00	40.37 (2.10)	40.39 (2.05)	40.39 (2.06)
c_1	2.00	1.99 (0.03)	1.99 (0.03)	2.00 (0.03)
c_2	2.00	2.00 (0.04)	2.00 (0.04)	2.00 (0.04)
c_3	2.00	2.00 (0.03)	2.00 (0.03)	2.00 (0.03)

* 50 Monte Carlo simulations for J=3, l_ξ=0.50 and T=400.

obtain a given level of statistical precision. We should note also that, for the case where no unobserved characteristics exist ($l_\xi = 0$), the model identifies the exact true values of the parameters at every single replication of the Monte Carlo experiment. The reason is that on specification A, the $[a_{it}]$ parameters are retrieved analytically and, as a result, no numerical error is introduced in the contraction inner loop (please see Dubé *et al.* 2009 for more details on numerical contraction induced errors).

The results presented on panels A and B of Table 3.1 are computed using the true parameters as starting points of the algorithm. Panel C explores the potential multiple local minima property of the GMM objective function in the lines of Dubé *et al.* (2009) and Knittel and Metaxoglou (2008) by considering multiple starting points. In particular, we consider two alternative starting points: $(1/2)\theta_2^{true}$ and $2\theta_2^{true}$, with column (iii) again just reproducing the baseline case from panel A. The results point to the robustness of the algorithm to the different starting points since it converges to very similar parameter values each time.

To sum up, the results suggest three key features of the estimation procedure: *i*) the estimators seem to be consistent, *ii*) the biases are typically non-increasing with the sample size and non-decreasing with the magnitude of the noise-to-signal ratio l_ξ, and finally *iii*) the GMM objective function seems to be have an apparent global minimum.

We now move to sample experiments with a number of goods that yield a too great number of parameters to be estimated. As a consequence, the $[b_{ij}]$ and $[c_i]$ parameters need to be mapped onto the characteristics space. In theory, all observed characteristics can be mapped onto both θ_1 and θ_2 sets of parameters. We choose to consider the case where the set of characteristics that affect the parameters in θ_1 is disjoint from the set of characteristics that affect the parameters in θ_2. In particular, we consider that the initial K observed characteristics affect the former and draw an additional set of $KL = 3$ observed characteristics that are assumed to affect the latter. A possible alternative to this specification would be to consider that the observed characteristics that affect the parameters in θ_2 are some transformation of the ones that affect the parameters in θ_1 in order to avoid the introduction of eventual biases (see Pinkse *et al.*, 2002). The setup is identical to the one for the small number of goods case above with the exception that:

$$\begin{aligned} (b_{ii})_t &= \alpha_{10} + \alpha_{11} x_{1it} \\ (b_{ij})_t &= \alpha_{20} + \alpha_{21} \left| x_{2it} - x_{2jt} \right| \qquad (52) \\ (c_i)_t &= \alpha_{30} + \alpha_{31} x_{3it}, \end{aligned}$$

TABLE 3.2

Monte-Carlo Results: Specification A - Large J Case

Panel A: Sensitivity to Number of Goods J*

Parameter	True	(i): 30	(ii): 60	(iii): 120
α_{10}	50.00	49.32 (2.12)	49.97 (1.74)	49.98 (1.26)
α_{11}	50.00	49.88 (0.95)	49.93 (0.83)	50.20 (0.86)
α_{20}	-2.00	-2.05 (0.38)	-2.03 (0.27)	-1.97 (0.16)
α_{21}	-2.00	-2.01 (0.13)	-2.00 (0.07)	-2.00 (0.04)
α_{30}	30.00	30.00 (0.03)	30.00 (0.04)	30.01 (0.05)
α_{31}	30.00	30.00 (0.01)	30.00 (0.02)	30.01 (0.03)

* 50 Monte Carlo simulations for T=10, l_{ξ}=0.50 and starting values at θ_2^{true}.

Panel B: Sensitivity to Noise-to-Signal Ratio l_{ξ}*

Parameter	True	(iv): 1.50	(iii): 0.50	(v): 0.00
α_{10}	50.00	49.88 (3.79)	49.98 (1.26)	50.00 (0.00)
α_{11}	50.00	50.60 (2.58)	50.20 (0.86)	50.00 (0.00)
α_{20}	-2.00	-1.93 (0.46)	-1.97 (0.16)	-2.00 (0.00)
α_{21}	-2.00	-1.99 (0.11)	-2.00 (0.04)	-2.00 (0.00)
α_{30}	30.00	30.02 (0.14)	30.01 (0.05)	30.00 (0.00)
α_{31}	30.00	30.01 (0.09)	30.01 (0.03)	30.00 (0.00)

* 50 Monte Carlo simulations for J=120, T=10 and starting values at θ_2^{true}.

Panel C: Sensitivity to Starting Values*

Parameter	True	(vi): $(1/2)\theta_2^{true}$	(iii): θ_2^{true}	(vii): $2\theta_2^{true}$
α_{10}	50.00	49.97 (1.27)	49.98 (1.26)	49.98 (1.27)
α_{11}	50.00	50.20 (0.86)	50.20 (0.86)	50.20 (0.86)
α_{20}	-2.00	-1.97 (0.16)	-1.97 (0.16)	-1.97 (0.16)
α_{21}	-2.00	-2.00 (0.04)	-2.00 (0.04)	-2.00 (0.04)
α_{30}	30.00	30.01 (0.05)	30.01 (0.05)	30.01 (0.05)
α_{31}	30.00	30.01 (0.03)	30.01 (0.03)	30.01 (0.03)

* 50 Monte Carlo simulations for J=120, T=10 and l_{ξ}=0.50.

with $\alpha_{10} = \alpha_{11} = 50$, $\alpha_{20} = \alpha_{21} = -2$ and $\alpha_{30} = \alpha_{31} = 30$.

Table 3.2 presents the results for the large number of goods case and hence assumes that the asymptotic arguments work in the number of goods. All the results are conditional on $T = 10$ markets. The first panel explores the sensitivity of the estimates to the number of goods. We consider three cases: $J = 30$, $J = 60$ and $J = 120$. Similarly to the previous case, the results point to the convergence of the algorithm, with both price and income parameters converging to the true values even at small sample sizes. The second panel again explores the sensitivity of the algorithm to noise-to-signal ratio. We consider three cases: $l_\xi = 1.50$, $l_\xi = 0.50$ (the level under which the results in panel A are derived) and $l_\xi = 0.00$. The results suggest that, as before, an increase in the noise-to-signal ratio deteriorates the performance of the algorithm (even though in a slightly smaller scale when compared with the previous case). Again, when there are no unobserved characteristics ($l_\xi = 0$), the algorithm identifies the exact true values of the parameters at every single replication of the Monte Carlo experiment, since the $[a_{it}]$ parameters are retrieved analytically and no numerical error is introduced in the contraction inner loop. The third panel explores the potential multiple local minima property of the GMM objective function by considering two multiple starting points: $(1/2)\theta_2^{true}$ and $2\theta_2^{true}$, and the algorithm, again, seems robust to those alternative starting points.

The results for the large number of goods case under specification A point to same three key features of the estimation procedure outlined for the previous case: *i*) the estimators seem to be consistent, *ii*) the biases are typically non-increasing with the sample size and non-decreasing with the magnitude of the noise-to-signal ratio l_ξ, and finally *iii*) the GMM objective function seems to be have an apparent global minimum.

3.4.2 Specification B

We consider now the algorithm's properties under specification B. As before, we investigate both the small and large number of goods cases. Table 3.3 presents the results for the small number of goods case: $J = 3$. The true values of the parameters of the indirect utility function to be estimated are $b_{ii} = 20$ and $b_{ij} = -2$ for all i, j. The symmetry assumption on the estimation of the $[b_{ij}]$ parameters was again imposed for expositional convenience. The first panel explores the sensitivity of the estimates to the number of markets. We consider the same cases as in specification A: $T = 100$, $T = 200$ and $T = 400$. The parameters seem to converge to the true values with those parameters that drive own-price effects being (again) easier to identify than those that impact cross-prices effects. The second panel presents the

TABLE 3.3

Monte-Carlo Results: Specification B - Small J Case

Panel A: Sensitivity to Number of Markets T*				
Parameter	True	(i): 100	(ii): 200	(iii): 400
b_{01}	-2.00	-2.00 (0.01)	-2.00 (0.01)	-2.00 (0.01)
b_{02}	-2.00	-2.00 (0.01)	-2.00 (0.01)	-2.00 (0.01)
b_{11}	20.00	19.99 (0.10)	19.99 (0.07)	19.99 (0.09)
b_{21}	-2.00	-2.00 (0.02)	-2.00 (0.02)	-2.00 (0.02)
b_{22}	20.00	19.99 (0.10)	19.99 (0.08)	19.99 (0.09)
a_0	50.00	49.79 (2.28)	49.79 (1.90)	49.63 (1.84)
* 50 Monte Carlo simulations for J=3, l_ξ=0.50 and starting values at θ_2^{true}.				
Panel B: Sensitivity to Noise-to-Signal Ratio l_ξ*				
Parameter	True	(iv): 1.50	(iii): 0.50	(v): 0.00
b_{01}	-2.00	-1.99 (0.03)	-2.00 (0.01)	-2.00 (0.00)
b_{02}	-2.00	-1.99 (0.03)	-2.00 (0.01)	-2.00 (0.00)
b_{11}	20.00	19.89 (0.27)	19.99 (0.09)	20.00 (0.00)
b_{21}	-2.00	-1.98 (0.05)	-2.00 (0.02)	-2.00 (0.00)
b_{22}	20.00	19.89 (0.26)	19.99 (0.09)	20.00 (0.00)
a_0	50.00	47.16 (5.42)	49.63 (1.84)	50.00 (0.00)
* 50 Monte Carlo simulations for J=3, T=400 and starting values at θ_2^{true}.				
Panel C: Sensitivity to Starting Values*				
Parameter	True	(vi): $(1/2)\theta_2^{true}$	(iii): θ_2^{true}	(vii): $2\theta_2^{true}$
b_{01}	-2.00	-2.00 (0.01)	-2.00 (0.01)	-2.00 (0.01)
b_{02}	-2.00	-2.00 (0.01)	-2.00 (0.01)	-2.00 (0.01)
b_{11}	20.00	19.99 (0.09)	19.99 (0.09)	19.99 (0.09)
b_{21}	-2.00	-2.00 (0.02)	-2.00 (0.02)	-2.00 (0.02)
b_{22}	20.00	19.99 (0.09)	19.99 (0.09)	19.99 (0.09)
a_0	50.00	49.58 (1.91)	49.63 (1.84)	49.65 (1.88)
* 50 Monte Carlo simulations for J=3, l_ξ=0.50 and T=400.				

results of experiments in which we varied the noise-to-signal ratio: $l_\xi = 1.50$, $l_\xi = 0.50$ and $l_\xi = 0.00$. The conclusions are in line to those for specification A, with an increase in the ratio being correlated to a deterioration in the algorithm's performance - implying that we require additional data in order to obtain a given level of statistical precision. For specification B, a numerical error may be introduced in the contraction inner loop that retrieves the $[a_{it}]$ parameters. However, the results seem to suggest that, at least for the stopping criteria assumed, this numerical error is not significant. Finally, we address the robustness of the algorithm to different starting points. The last panel considers the same two alternative starting points as before: $(1/2)\theta_2^{true}$ and $2\theta_2^{true}$. The results seem to suggest the algorithm is robust to those different starting points.

To sum up, the results point to the same three key features of the estimation procedure outline for specification A: *i*) the estimators seem to be consistent given enough data, *ii*) the biases are typically non-increasing with the sample size and non-decreasing with the magnitude of the noise-to-signal ratio l_ξ, and finally *iii*) the GMM objective function seems to have an apparent global minimum.

We now turn to the large number of goods case. The setup is identical to the one for specification A above with the exception that:

$$\begin{aligned} (b_{ii})_t &= \alpha_{10} + \alpha_{11} x_{1it} \\ (b_{ij})_t &= -1/\left(\alpha_{20} + \alpha_{21} \left| x_{2it} - x_{2jt} \right|\right), \end{aligned} \tag{53}$$

with $\alpha_{10} = \alpha_{11} = 40$ and $\alpha_{20} = \alpha_{21} = 30$. Tables 4 presents the results conditional on the constant number of independent markets $T = 10$. The first panel explores the sensitivity of the estimates to the number of goods. We consider the same cases as in specification A: $J = 30$, $J = 60$ and $J = 120$. The results suggest that, given enough data, the algorithm converges to the true parameter values. The second panel evaluates the sensitivity of the algorithm to the variance of the unobserved characteristics ξ_i. We consider $l_\xi = 1.50$, $l_\xi = 0.50$ and $l_\xi = 0.00$. The results are similar to the previous case with an increase in the noise-to-signal ratio deteriorating the performance of the algorithm. Furthermore, the results seem to suggest that, at least for the stopping criteria assumed, the numerical error introduced in the contraction inner loop that retrieves the $[a_{it}]$ parameters is not significant. The last panel addresses the robustness of the algorithm to alternative starting points, and again the algorithm seems robust.

The results of the large number of goods case for specification B seem to point to the same three key features of the estimation procedure: *i*) the consistence of estimators given enough

TABLE 3.4

Monte-Carlo Results: Specification B - Large J Case

Panel A: Sensitivity to Number of Goods J*				
Parameter	True	(i): 30	(ii): 60	(iii): 120
α_{10}	40.00	40.01 (0.08)	39.99 (0.08)	40.00 (0.06)
α_{11}	40.00	40.00 (0.08)	39.99 (0.08)	40.00 (0.06)
α_{20}	30.00	30.07 (0.36)	30.04 (0.26)	30.00 (0.19)
α_{21}	30.00	30.26 (1.08)	30.09 (0.70)	29.94 (0.55)
a_0	50.00	50.01 (1.70)	49.73 (1.74)	50.12 (1.50)

* 50 Monte Carlo simulations for T=10, l_ξ=0.50 and starting values at θ_2^{true}.

Panel B: Sensitivity to Noise-to-Signal Ratio l_ξ*				
Parameter	True	(iv): 1.50	(iii): 0.50	(v): 0.00
α_{10}	40.00	39.98 (0.19)	40.00 (0.06)	40.00 (0.00)
α_{11}	40.00	39.98 (0.19)	40.00 (0.06)	40.00 (0.00)
α_{20}	30.00	30.00 (0.56)	30.00 (0.19)	30.00 (0.00)
α_{21}	30.00	29.84 (1.67)	29.94 (0.55)	30.00 (0.00)
a_0	50.00	49.78 (4.45)	50.12 (1.50)	50.00 (0.00)

* 50 Monte Carlo simulations for J=120 T=10 and starting values at θ_2^{true}.

Panel C: Sensitivity to Starting Values*				
Parameter	True	(vi): $(1/2)\theta_2^{true}$	(iii): θ_2^{true}	(vii): $2\theta_2^{true}$
α_{10}	40.00	40.00 (0.06)	40.00 (0.06)	40.00 (0.06)
α_{11}	40.00	40.00 (0.06)	40.00 (0.06)	40.00 (0.06)
α_{20}	30.00	30.00 (0.19)	30.00 (0.19)	30.00 (0.19)
α_{21}	30.00	29.94 (0.55)	29.94 (0.55)	29.94 (0.55)
a_0	50.00	50.12 (1.50)	50.12 (1.50)	50.12 (1.50)

* 50 Monte Carlo simulations for J=120, T=10 and l_ξ=0.50.

data, *ii*) the biases are typically non-increasing with the sample size and non-decreasing with the magnitude of the noise-to-signal ratio l_ξ, and finally *iii*) the GMM objective function seems to have an apparent global minimum.

3.4.3 Substitution Patterns

Real-world applications are typically concerned with the estimation of own- and cross-price elasticities. Table 3.5 presents the results and denotes the mean biases and standard errors (in parenthesis) of the predicted substitution patterns across a selection of the above experiments. The labels on the columns correspond to the Monte Carlo specification from the previous tables: (iii) denotes the results for the baseline case with $l_\xi = 0.50$ and starting values at θ_2^{true}, (iv) denotes the results for the case with $l_\xi = 1.50$ and starting values at θ_2^{true}, and finally (vii) denotes the results for the baseline case with starting values at $2\theta_2^{true}$. All the specifications seem to capture well both the general pattern and level of substitution across goods.

3.5 Dynamic Demand

This chapter leaves many estimation issues yet to be explored. In this section we briefly discuss how could we extend the model to account for dynamic behaviour as some real-world applications may involve forward-looking behavior by consumers. A more extensive study of the properties of this extension seems more appropriately considered as a separate one and hence is left for future research.

Let us consider a demand setting in the lines of Gorman (1971)'s multi-stage budgeting approach as an example. Consider that our representative consumer is faced with J different brands of a storable good and has to decide, in each period, how much of each brand to purchase. Following Gorman (1971)'s approach (see Aguirregabiria (2002) and Hendel and Nevo (2006a) for similar dynamic applications), we can separate the quantity decision from the brand decision. Assume that the purchased amount, denoted by x_t, is simply a choice of size, with $x_t = 0$ standing for no purchase. Let $d_{xt} = 1$ denote a purchase of size x in period t, and assume $\sum_x d_{xt} = 1$. Because the good is storable, the consumer does not need to consume all the purchased quantity in a given period. As a consequence, the consumer has also to decide how much to consume in each period. Quantity not consumed is stored as inventory.

TABLE 3.5

Monte-Carlo Results: Predicted Elasticity Bias

Panel A: Specification A

Elasticity	Small J Case* (iii)	Small J Case* (iv)	Small J Case* (vii)	Large J Case** (iii)	Large J Case** (iv)	Large J Case** (vii)
ε_{11}	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ε_{21}	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ε_{22}	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ε_{31}	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ε_{32}	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ε_{33}	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)

* 50 Monte Carlo simulations for J=3 and T=400. ** 50 Monte Carlo simulations for J=120 and T=10.

Panel B: Specification B

Elasticity	Small J Case* (iii)	Small J Case* (iv)	Small J Case* (vii)	Large J Case** (iii)	Large J Case** (iv)	Large J Case** (vii)
ε_{01}	0.00 (0.00)	-0.01 (0.01)	-0.10 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ε_{02}	0.00 (0.00)	-0.01 (0.01)	-0.10 (0.01)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ε_{11}	0.00 (0.00)	0.01 (0.01)	0.09 (0.00)	0.00 (0.00)	0.00 (0.01)	0.06 (0.00)
ε_{21}	0.00 (0.00)	-0.01 (0.01)	-0.07 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
ε_{22}	0.00 (0.00)	0.01 (0.01)	0.09 (0.01)	0.00 (0.00)	0.00 (0.01)	0.06 (0.00)

* 50 Monte Carlo simulations for J=3 and T=400. ** 50 Monte Carlo simulations for J=120 and T=10.

In each period, the consumer's problem can be represented by:

$$
\begin{aligned}
V(s_1) &= \max_{\{c(s_t),x(s_t)\}} \sum_{t=1}^{\infty} \delta^{t-1} E\left[u(c_t, v_t;\theta) - C(I_{t+1};\theta) + \textstyle\sum_x d_{xt} H_{xt}(r, y; \theta, \Im_x)\,|s_1\right], \\
&\quad \text{s.t.} \\
0 &\leq I_t, \quad 0 \leq c_t, \quad 0 \leq x_t, \quad \textstyle\sum_x d_{xt} = 1, \quad I_{t+1} = I_t + x_t - c_t
\end{aligned}
$$

where s_t denotes the state at time t, $\delta > 0$ denotes the discount factor, $u(c_t + v_t; \theta)$ denotes the utility from consumption, v_t denotes a shock to utility that impacts the marginal utility from consumption, $C(I_{t+1}; \theta)$ denotes the cost of storage, and $H_{xt}(r, y; \theta, \Im_x)$ denotes the utility from a size x purchase ($\Im_x$ denotes the set of brands available with size x).

The state vector s_t consists, at time t, of the current (or beginning-of-period) inventory I_t, current prices, and the shock v_t. We can, following Hendel and Nevo (2006a), make the simplifying assumptions that v_t is independently distributed over time, and prices follow an exogenous Markov process.

The estimation algorithm could be based in the procedure outlined in Hendel and Nevo (2006a). The first step would consist of estimating the parameters in $H_{xt}(r, y; \theta, \Im_x)$. This could be achieved by estimating a static continuous choice model conditional on the choice set $\Im_x$. This procedure yields consistent, although potentially inefficient, parameters for the indirect utility function. The second step would consist of computing the indirect utility associated with each size and their transition probabilities from period to period. Finally, the third step would consist of solving a simplified version of the full dynamic problem - restricted to the remaining parameters only - by maximizing the likelihood of the observed sequence of sizes purchased.

3.6 Conclusions

In this chapter, we consider a new method of uncovering demand information from market level data on differentiated products. We follow the continuous-choice literature and develop a globally consistent continuous-choice demand model that combines desirable properties of both the discrete- and continuous-choice literatures: *i*) it is flexible in the sense of Diewert (1974), *ii*) it is globally consistent in the sense it can deal with entry and exit of products over time, and *iii*) incorporates a structural error term. In order to encompass different possible real-world applications, we propose two alternative specifications of our baseline model depending on the degree of flexibility the researcher is willing to accept for the substitution patterns between inside and outside goods.

The estimation procedure follows an analog to the algorithm derived in Berry (1994) and BLP. Depending on the specification considered, the contraction mapping for matching observed and predicted budget shares may be analytical or not. The case for which the contraction is analytical is relatively simple and fast to estimate which can prove a key advantage in competition policy issues, where time and transparency are typically crucial factors. For the case it is not, we propose an alternative to BLP's contraction mapping with super-linear rate of convergence.

We provide a series of Monte Carlo experiments to illustrate the estimation properties of the model and discuss how it can be extended to cope with consumer dynamic behaviour. A more extensive study of the properties of this extension seems more appropriately considered as a separate one and hence is left for future research.

Chapter 4

Market Dominance and Barriers to Competition in Financial Trading Venues

The interaction between competition and economic growth is a well established fact in the literature (Porter (1990), Aghion and Howitt (1992), Blundell *et al.* (1995), Aghion *et al.* (1999)). Competition impacts economic growth via a more efficient allocation of market resources that contributes to "better economic performance, better prices and better services for consumers and businesses" (Kroes (2007)).

Recent years have witnessed a strong and ferocious promotion of competition in a large spectrum of markets and industries and a clear example of this trend is the new Market in Financial Instruments Directive (MiFID) that fosters a fair, competitive, transparent, efficient and integrated European financial market. MiFID aims - among other objectives - to harmonize the trading structures across the Member States by abolishing the requirement to concentrate the execution of trading orders by financial intermediaries in a single venue. This challenges the market power of existing venues and fosters entry by new players. I argue that, although positive, the impact of this increased *potential* competition on the degree of *actual* competition may be limited due to two barriers to competition: *i*) direct network effects together with economies of scale and *ii*) post-trading constraints (since venues typically bundle trading and post-trading services).

Direct network effects and economies of scale impose a barrier to competition because they provide a first-mover advantage to the incumbent venue. In order to understand why this is the case, I examine the details of the trading decision. This decision can be decomposed in two stages. First, investors decide the order characteristics and send it to a financial intermediary to be executed. Second, after receiving the order, the chosen intermediary decides the trading venue where to execute it, conditional on the order characteristics received, with financial intermediaries choosing the venue that achieves the best price at a lower cost.[13]

[13] Currently, this constitutes a legal obligation for financial intermediaries. MiFID determines that the choice of trading venue by financial intermediaries must achieve best execution to their clients, where best

This suggests that the decision of financial intermediaries must therefore take into account different dimensions, which the literature typically divide in two broad categories: explicit and implicit costs. Explicit trading costs denote the transaction costs of a venue and include the costs of executing the order (trading fees) and the cost of post-trading (clearing and settlement fees). Implicit trading costs relate to the liquidity of a venue and typically include the bid-ask spread, the potential price impact of a trade, and the opportunity cost of missed trades.

Implicit trading costs relate to the direct network effects feature of the industry. The valuation of a venue by financial intermediaries is increasing in the number of other agents that choose the same venue since it reduces the costs of finding a counterpart. In other words, a more liquid venue translates into lower implicit trading costs as it *i*) stabilizes the market price of a financial instrument, and *ii*) reduces the extent to which placing an order has an adverse effect on the corresponding price. Incumbent stock exchanges typically have higher liquidity than smaller trading venues and therefore lower implicit trading costs. In order for competitors to succeed, they need to trade-off the disadvantage of having higher implicit trading costs with lower explicit trading costs. However, economies of scale in trading (and post-trading) prevent those smaller trading venues to compete on explicit trading costs, giving incumbent stock exchanges a competitive advantage. Direct network effects and economies of scale constitute then a barrier to competition.

The second barrier to competition is induced by post-trading constraints that increase venue differentiation. Different trading venues can not be considered as effective substitutes if they imply different post-trading arrangements - with different clearing and settlement costs. The competition for trading venues is, therefore, limited by the fact that financial intermediaries can not freely choose post-trading arrangements.

This chapter proposes to empirically address the following questions: *i*) evaluate the actual degree of competition between alternative trading venues, *ii*) measure the impact of network effects on competition, and finally *iii*) assess the barriers to competition induced by post-trading constraints. To this end, I specify a structural discrete-choice multinomial random-coefficients logit demand model for trading in the lines of Berry, Levinsohn, and Pakes (1995) (henceforth BLP) that takes into account the trade-off between explicit and implicit trading costs following Pagano (1989). The model is flexible in the sense that the implied substitution patterns do not suffer from the problem of the Independence of Irrelevant Alternatives (IIA) property characteristic of more standard demand models. Furthermore, following the literature, the error term is structurally embedded in the model and, thereby,

execution denotes choosing the venue that achieves the best price at a lower cost.

circumvents the critique provided by Brown and Walker (1989) related to the addition of add-hoc errors and their induced correlations.

I apply the model to the set of 16 most traded securities in the FTSE 100 following the list of liquid securities published (and updated regularly) by CESR after the implementation of MiFID. The results imply that financial intermediaries tend to value liquidity more than total fees when deciding to which venue to route a given order for execution. For this reason, the incumbent venue has a clear advantage relatively to its competitors and can, as a result, exert market power when setting its fees level. After estimating the degree of substitutability between the different trading venues, I examine the impact of the mentioned barriers to competition. First, I study the role of direct network effects by computing the counterfactual market shares that would arise if there were no liquidity differences across venues. Then, I evaluate the effect of the post-trading constraints induced by the typical bundle of trading and post-trading services. I simulate the equilibrium market shares that would arise if the different trading services were fungible. In both cases, the results suggest that eliminating the corresponding barrier to competition is associated with a significant decrease (of a similar magnitude) in the asymmetry of the industry. Finally, I draw some economic policy implications.

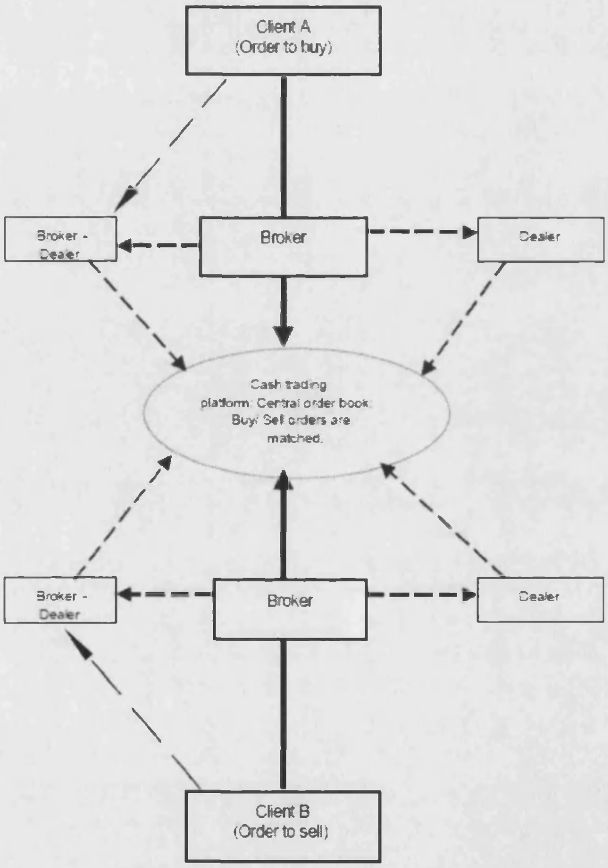
4.1 The Economics of Trading and MiFID

The process of trade begins with investors sending their buying or selling orders to a broker or a broker-dealer. If investors choose the former, the broker receives the order and can either *i*) place it directly on a trading venue order book or *ii*) decide to go indirectly via a dealer. If the broker chooses option *ii*) or the investors send their orders directly to a broker-dealer, then the dealer (or broker-dealer depending on the case) can match the order from its own inventory, place the order on a trading venue or go to another dealer. Figure 4.1 illustrates the process of trading in an electronic platform. I do not attempt to model the clients choice for one of these financial intermediaries and, therefore, do not distinguish between brokers, dealers and broker-dealers (henceforth denoted financial intermediaries). Rather, I focus on studying the subsequent choice of venue where to execute the clients order.

The trading market in Europe does not seem, at least at a first glance, extremely concentrated for an industry with strong network effects and economies of scale. If we consider the set of all European securities, the volume market share of the leading trading venue is roughly 30% with the top-3 venues capturing approximately 60% of the market. However,

FIGURE 4.1

*Trading Mechanism**



Source: Pagano and Padilla (2005).

TABLE 4.1
*Market Concentration**

Concentration Ratios	C_1	C_3	C_5
European Equities	31%	58%	75%
UK Equities	63%	87%	89%
FTSE 100 Equities	70%	98%	99%

* Source: Reuters Market Share Reports, November 2007.

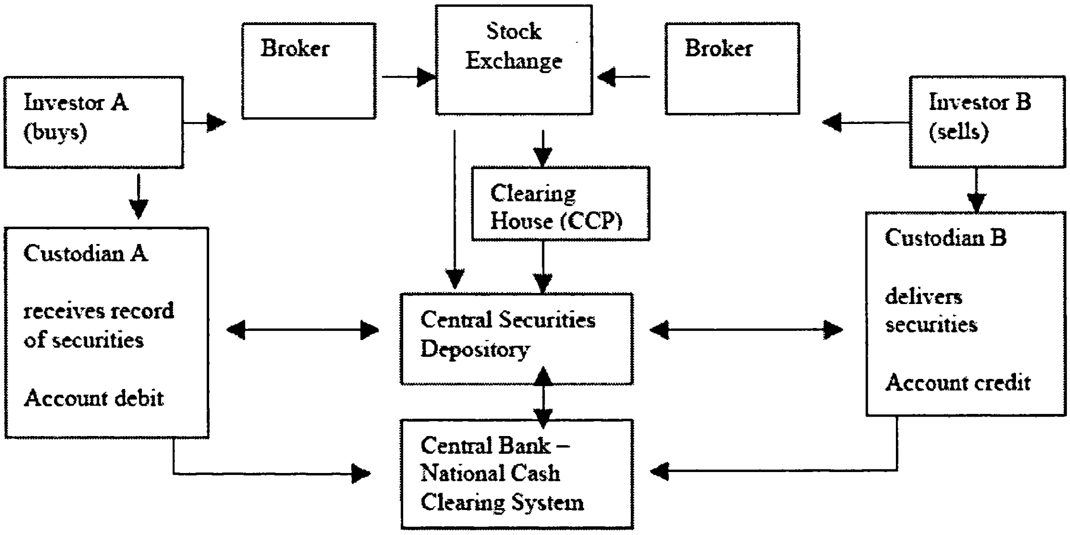
these statistics are somehow misleading. When we examine concentration for narrower market definitions, we conclude that trading for a particular security is concentrated on a smaller set of trading venues. If you consider, for example, the set of the FTSE 100 securities, the market share for the leading trading venue is now roughly 70% with the top 3 venues capturing approximately 98% of the market! Table 4.1 displays concentration ratios for different sets of European equities (and therefore market definitions).

MiFID tries to promote a significant change in the shape of the industry. It aims to increase competition by creating a common harmonized European market for financial products and to foster client protection through improved transparency, suitability requirements and best execution principles. In particular, it abolishes the so-called "concentration rule" that allowed, in the past, Member States to impose that securities admitted to trading on a regulated market would have to be traded only on regulated markets. In contrast, MiFID allows trading services to be provided by a variety of venues, namely Regulated Markets (RM), Multilateral Trading Facilities (MTF) and Systematic Internalizers (SI). RM or MTF are entities that offer multilateral trading for financial instruments (such as an order book), with slightly different standards applying to each, whereas SI refer to financial firms which, on an organized, frequent and systematic basis, deal on own account by executing client orders outside a RM or an MTF.

A financial intermediary desiring to trade a given security is therefore faced with a venue choice. It can choose a RM like the London Stock Exchange, Euronext or Frankfurt Stock Exchange, a MTF like Chi-X, or a SI like ABN AMRO, Goldman Sachs or UBS. The only requirement is that the chosen venue achieves best execution, taking into account a number of factors that include transaction costs, price and liquidity, speed of execution, likelihood of execution, clearing and settlement arrangements, etc.

Transaction costs refer, as mentioned above, to the explicit trading costs of each venue. These costs can be decomposed into costs of executing an order (trading fees) and costs of post-trading (clearing and settlement fees). Clearance refers to the validation of a trade and the subsequent establishment of the obligations of the parties to the trade (what each owes

FIGURE 4.2
*Clearing and Settlement Flows**



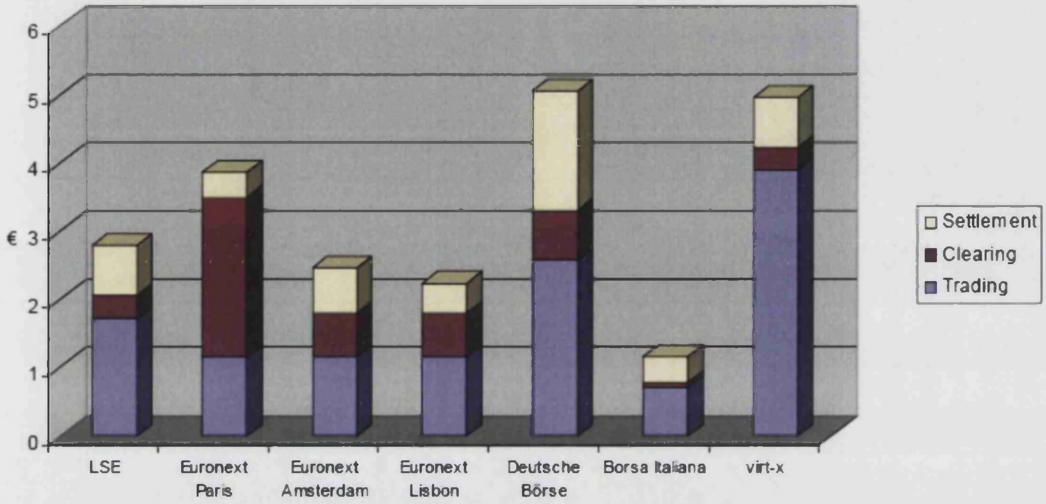
Source: Carvalho (2004).

and is entitled to receive), while settlement is the process during which buyer and seller details are matched and the security changes ownership against the appropriate payment. Clearing and settlement services are typically performed by specializing institutions: the transfer of ownership is carried out by a central securities depository or an international central securities depository, whereas the banking/payment system handles the payment of funds. Figure 4.2 displays the flows involved in the clearing and settlement of a trade.

Transaction costs vary substantially across trading venues. Not only in absolute terms but also in their decomposition. Figure 4.3 compares transaction costs for a series of trading venues. This comparison raises an important question. What prevents trades to concentrate on the venue which offers the lowest fees? As suggested above, explicit trading costs are not the only drivers of best execution. Price and liquidity are other important factors in achieving best execution. And these relate to the implicit trading costs of each venue, which typically include the bid-ask spread, the potential price impact of a trade, and the opportunity cost of missed trades. Implicit trading costs are important since cash trading exhibits direct network effects. Financial intermediaries valuation of a venue is increasing in the number of other agents that choose the same venue since it reduces the costs of finding a counterpart. In other words, a more liquid venue translates into lower implicit trading costs as it *i*) stabilizes the market price of a financial instrument, and *ii*) reduces the extent to which placing an order has an adverse effect on the corresponding price. Pagano (1989) shows that if the explicit trading costs are identical across venues, the direct network effects

FIGURE 4.3

*Europe: Decomposition of Explicit Costs per Trade**



Source: European Commission (2006). Data refers to 2004.

promote the concentration of trade on only one venue. However, if low explicit trading costs of a venue are traded-off against high implicit trading costs (or vice-versa), multiple trading venues can coexist in equilibrium.

Underestimating the importance of network effects can lead to a dismal failure. As an illustration consider the case of Jiway, a pan-European trading platform for retail investors launched in the last quarter of 2000 by Morgan Stanley and the Swedish company OM. The two companies invested \$100 million on the project that promised access to 6,000 European securities, but that was incapable to attract liquidity: in January 2001 it executed 1,996 trades, in February 474 trades, and in March 577 trades. By the end of 2002, Jiway was shut down.

Incumbent stock exchanges typically have higher liquidity than smaller trading venues and therefore lower implicit trading costs. In order for competitors to succeed, they need to trade-off the disadvantage of having higher implicit trading costs with lower explicit trading costs. This has been exactly the strategy of Chi-X, a multilateral trading facility set up in the first quarter of 2007, which offers a fee schedule that reverses the standard in the industry and includes, in certain cases, a negative execution fee - corresponding to a payment from the venue to the intermediary. A solution with very optimistic results up to this moment. However, long-term low explicit trading costs are difficult to sustain in these type of industries because the strong economies of scale give incumbent stock exchanges the competitive advantage of being able to set lower fees levels.

TABLE 4.2
*Average Volume per Order**

London Stock Exchange	3,509
Chi-X	1,302
Systematic Internalizers	42,386

* Source: Author's calculations, November 2007 - March 2008.

Another alternative strategy to trade-off the disadvantage of having higher implicit trading costs is to avoid direct competition with the incumbent and specialize in attracting intermediaries with niche trading profiles. Table 4.2 presents the average volume per order involving the 20 most traded FTSE 100 securities for the top 3 trading venues. The data suggests that segmentation may be in fact an issue in this market and, as a result, the concentration ratios presented above may be even be higher if certain characteristics of the orders - like size - are taken into consideration. In order to evaluate the actual degree of competition between trading venues, the empirical framework must be able to deal with eventual segmentation of the market.

Direct network effects and economies of scale do not constitute however the only barrier to competition. The bundling of trading and of post-trading services constitutes another. The reason is that even though financial intermediaries can *a priori* choose between a set of competing trading venues to execute an order, the services offered by the different venues can not actually be considered real substitutes or fungible because different trading venues may imply different clearing and settlement arrangements. In order to understand why this is the case, consider, as an illustration, a financial intermediary with an order to trade Royal Dutch securities. The intermediary can execute the order on a set of alternative venues from Euronext Amsterdam to Deutsche Borse. However because post-trading services are typically bundled with trading services, when the intermediary chooses a venue, she is implicit choosing also the corresponding post-trading provider. Table 4.3 presents the trading venues and associated central securities depositories for Royal Dutch securities. In this illustration, only the securities trading in Euronext Amsterdam, London Stock Exchange and Chi-X are fully fungible as they settle in the same CSD - Euroclear Amsterdam. Trading Royal Dutch in Virt-X or Deutsche Borse may imply settlements across different CSD with associated higher costs. Carvalho (2004) shows that the costs of clearing and settlement across different CSD within Europe are 42% higher than if using the same CSD. As a consequence, venues that settle in the same CSD have a competitive advantage when compared with those that settle in different CSD. This advantage may induce intermediaries to choose a venue that does not *a priori* offer the best execution fee. In sum, there can not be real competition between

TABLE 4.3
*Royal Dutch Trading and post-Trading (Venue/CSD)**

Venue	Central Securities Depository
Euronext Amsterdam	Euroclear Amsterdam
London Stock Exchange	Euroclear Amsterdam
Chi-X	Euroclear Amsterdam
Virt-X	Euroclear Bank
Deutsche Borse	Clearstream Banking Frankfurt

* Source: Misra (2007).

trading venues if financial intermediaries can not freely choose post-trading arrangements.

I argue that the impact of the increased *potential* competition on the degree of *actual* competition is positive. In general, multi-venue trading promotes lower explicit trading costs via higher competition. However it also has a fragmentation effect. When different trading venues coexist, markets become fragmented and the liquidity available in any one setting is reduced, thereby potentially limiting any market's ability to provide stable prices. The bid-ask spreads may increase and daily securities returns may have a larger variance. Moreover, as liquidity facilitates the crucial price discovery role of markets, as order flow fragments, the ability of prices to aggregate information can be reduced, and with it the efficiency of the market. MiFID addresses this point by requiring every venue not only to publish the price, volume and time of a transaction as close to real-time as possible, but also to do it in a way that is easily accessible to other market participants. Furthermore, it also consolidates the hitherto fragmented market of European over-the-counter (OTC) securities. For these reasons, the fragmentation issues of increased trading venue competition may be less significant for MiFID.

The main contribution of the chapter is that, although positive, the impact of this increased *potential* competition on the degree of *actual* competition may be limited due to the two barriers to competition discussed above: *i*) direct network effects together with economies of scale and *ii*) post-trading constraints since venues typically bundle trading and post-trading services.

4.2 Literature Review

The demand model specified in this chapter is indirectly related to the literature on market dominance. This literature typically focuses on the source of such dominance. Gilbert and Newbery (1982) and Reinganum (1983) show that a monopolist can maintain her dominance due to stronger incentives for preemptive innovation. Budd, Harris and Vickers (1993) examine the dynamics of market structure in a duopoly and, in particular, in what circumstances we may see a process of increasing dominance sourced on higher levels of technology. Cabral and Riordan (1984) investigate another source of eventual market dominance, the hypothesis that due to a learning curve, unit costs may decline with cumulative production. Athey and Schmutzler (2001) model an oligopolistic setting to examine conditions under which dominance sourced in ongoing investment may emerge. Cabral (2002) considers a similar setting, but where firms choose the amount of resources to invest and how to allocate those resources.

This chapter examines market dominance sourced on *i*) network effects and *ii*) trading and post-trading bundling. The literature on network effects begins with Katz and Shapiro (1985) and from then on it has developed along two different strands. While one of the strands tries to empirically measure the effect of network effects, the other studies its implications. Katz and Shapiro (1994), Economides (1996), Shy (2001), and Farrell and Klemperer (2006) provide an excellent overview of this literature. The literature on trading and post-trading bundling is less numerous. Tapking and Yang (2004) and Koppl and Monnet (2003) provide some excellent examples. They model competition between trading and post-trading services, with the former studying different forms of industry structures between venues and post-trading firms and the latter examining the impact of integrating the two services.

The demand model specified in this chapter is also indirectly related to the literature on venue competition. The seminal work is from Hamilton (1979), who establishes the two opposite effects of multi-venue trading and reports empirical estimates of the effect of off-boarding trading on liquidity and volatility of NYSE stocks. Multi-venue trading promotes lower explicit trading costs via higher competition but also has a fragmentation effect. When different trading venues coexist, markets become fragmented and the liquidity available in any one setting is reduced, thereby potentially limiting any market's ability to provide stable prices. The bid-ask spreads may increase and daily securities returns may have a larger variance. Moreover, since liquidity facilitates the crucial price discovery role of markets, as order flow fragments, the ability of prices to aggregate information can be reduced and, with it, the efficiency of the market. Hamilton finds that the competitive effect exceeds the fragmentation effect, and that both effects are small.

The literature on venue competition has evolved along two different strands. The first follows the lines of Hamilton (1979) and typically uses a reduced-form strategy that regress spreads and liquidity on stock and market characteristics that include a competition variable. More recent examples include Weston (2002) and Gresse (2006). Weston (2002) investigates whether the shift towards electronic communication networks leads to tighter bid-ask spreads and greater depths. He finds that this particular competition has a significant negative impact on bid-ask spreads, but no significant impact on quoted depth. Gresse (2006) studies the impact of crossing networks on the liquidity of the dealer market segment of the London Stock Exchange (SEAQ). She finds that spreads decrease due to competition but no fragmentation effect is detected. In parallel to the above approach, the second strand evolved towards more structural and micro-founded strategies of modelling financial markets. Hortaçsu and Syverson (2004) and Cantillon and Ying (2007) constitute some recent examples. Hortaçsu and Syverson (2004) investigate the role that nonportfolio fund differentiation and information/search frictions play in creating two salient features of the mutual fund industry: the large number of funds and the sizable dispersion in fund fees. Cantillon and Ying (2007) study the determinants of the dynamics of the market for the future on the Bund.

I propose to estimate a structural discrete-choice demand model for trading following BLP that tries to reconcile the advantages of Hamilton (1979)'s approach with the desirable features of a micro-founded model, taking into account two eventual barriers to competition, network effects as well as the bundle of trading and post-trading services.

4.3 Demand for Trading

The trading decision can be decomposed in two stages. First, investors decide the order characteristics and send it to a financial intermediary to be executed. Second, after receiving the order, the intermediary decides the trading venue where to execute it, conditional on the order characteristics received. In this chapter, I take the first stage as given and propose to model the second stage choice by financial intermediaries. A very interesting and natural extension will be to incorporate the first-stage into the model's framework.

Consider that in period $t = 1, \dots, T$ an investor sends an order with characteristics k to financial intermediary $i = 1, \dots, I$ for her to execute. The characteristics of the order include the code of the security, which I denote by j, the direction of the trade and the volume to be traded. After receiving the order, the financial intermediary has to choose, subject to her internal best execution policy, the trading venue where to execute the order.

The best execution policy, which needs under MiFID to be previously accepted by the investor, defines the intermediaries commitment with regard to the different dimensions that govern the venue decision: price, costs, speed, likelihood of execution and settlement, size, nature and any other consideration relevant to the execution of the order. An alternative view for the intermediaries best execution policy is to think of it as an auction where the intermediary allocates the order across the alternative trading venues according to an allocation rule known to the investor (although unknown to the econometrician). In order to estimate this allocation rule, I specify a structural multinomial random-coefficients logit discrete-choice demand model for trading in the lines of BLP. Conditional on order characteristics k, heterogeneous financial intermediaries decide on the venue $v = 0, 1, \dots, V$, with $v = 0$ denoting the outside option of executing the order over-the-counter or in an alternative venue. Financial intermediaries are assumed to be myopic and therefore decide for the venue that maximizes the per-period expected utility. This constitutes a reasonable assumption since best execution policies have to be applied on a trade by trade case.

The conditional indirect utility that financial intermediary i obtains from executing an order of characteristics k at venue v in period t can be specified as:

$$u_{ikvt}\left(p_{ikvt}, w_{jvt}; \theta_i\right) = u^*\left(p_{ikvt}, w_{jvt}; \theta_i\right) + \varepsilon_{ikvt}, \tag{54}$$

where w_{jvt} represents a vector of attributes with regard to the order, venue and time period, and p_{ikvt} denotes the all-in explicit trading costs incurred by the financial intermediary (which include execution, clearing and settlement fees). Because the fees schedules are typically a function of *both* intermediary i's trading profile[14] and order characteristics, the explicit trading costs p_{ikvt} are indexed by i and k. In order to explicitly illustrate the non-linearity of the fees schedules, I denote $p_{ikvt} = p_{vt}\left(z_i, k\right)$, where z_i represents intermediary i's trading profile. Allocation rule heterogeneity across financial intermediaries is modelled into the conditional indirect utility by allowing intermediary-specific valuations θ_i for the different elements included in the best execution policy. Finally, ε_{ikvt} denotes an additive preference shock.

The attributes of a trading venue, w_{jvt}, that impact the choice of intermediaries include naturally the implicit trading costs, which I denote by b_{jvt}. As discussed above, cash trading exhibits network effects and participants value liquidity. Although there is no uncontroversial definition of liquidity, the negative correlation between liquidity and implicit trading costs is generally accepted. A large installed base of intermediaries trading at venue v promotes lower implicit trading costs as it *i*) stabilizes the market price of a security and *ii*) reduces the extent to which placing an order has an adverse effect on the corresponding price. Note that these network effects can be artificially reinforced by fees schedules that are decreasing in trade volume.

[14] Volume discounts can reflect venue economies of scale that are passed to agents.

I assume $u^*(\cdot)$ can be specified as follows:

$$u^*\left(p_{ikvt}, w_{jvt}; \theta_i\right) = -\alpha_i p_{vt}\left(z_i, k\right) - \gamma_i b_{jvt} + x'_{jvt}\beta_i + \xi_{jvt}, \tag{55}$$

where:

i) the vector of characteristics w_{jvt} is split between the implicit trading costs of trading financial instrument j, b_{jvt}, a K-dimensional vector of observables, x_{jvt}, and a vector of unobserved characteristics (to the econometrician), whose mean valuation for orders referring to financial instrument j executed in venue v in period t across financial intermediaries is given by ξ_{jvt};

ii) the increasing function $\gamma_i b_{jvt}$ captures the network effects, where γ_i is the parameter that controls the strength of those network effects;

iii) and θ_i denotes the parameters of estimation: $\theta_i = (\alpha_i, \gamma_i, \beta_i)'$.

For completeness, I note that financial intermediaries can also choose to execute the order on an outside venue. The conditional indirect utility from the outside option is assumed to be $u_{ik0t} = \xi_{j0t} + \varepsilon_{ik0t}$. Following the literature, I normalize $\xi_{j0t} = 0$ without loss of generality since due to the ordinality of utility, only $\xi_{jvt} - \xi_{j0t}$ matters for the venue decision.

The parameters of estimation α_i, γ_i and β_i are indexed by intermediary in order to capture, as discussed above, possible heterogeneous allocation rules across intermediaries. In particular, I allow those parameters to be a function of the intermediaries trading profiles, z_i:

$$\begin{pmatrix} \alpha_i \\ \gamma_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} 0 \\ \gamma \\ \beta \end{pmatrix} + \theta^o z_i, \tag{56}$$

where θ^o denotes the vector of coefficients that govern the heterogeneity of intermediaries with regard to their trading profile. As a consequence, the parameters to be estimated reduce to $\theta = (\gamma, \beta, \theta^o)'$. After substituting equation (56) into the indirect utility in equation (54), it is possible to summarize the financial intermediaries conditional indirect utility as a sum of two terms: a first term that is common across intermediaries, $\delta_{jvt} = -\gamma b_{jvt} + x'_{jvt}\beta + \xi_{jvt}$, and a second term, $\mu_{ikvt} + \varepsilon_{ikvt}$, that introduces intermediary heterogeneity:

$$u_{ikvt} = \delta_{jvt} + \mu_{ikvt} + \varepsilon_{ikvt}, \tag{57}$$

where:

$$\mu_{ikvt} = \left[p_{vt}(z_i, k), b_{jvt}, x_{jvt}^{'}\right] \theta^o z_i.$$

As p_{ikvt} will typically vary with the financial intermediaries trading profile, so will μ_{ikvt}. Following European Commission (2006), I consider the following types of intermediaries with regard to their trading profiles: 1) typical volume and value trades, 2) large volume of low value trades, 3) large volume of high value trades, and 4) small volume of low value trades.

Given the heterogeneity of the financial intermediaries specified in the model, the solution to the maximization problem of the indirect conditional utility over all the different venues will vary from one intermediary to another, depending on their specific attributes (z_i, ε_{ikt}) where $\varepsilon_{ikt} = (\varepsilon_{ik0t}, \ldots, \varepsilon_{ikVt})$. The set of financial intermediaries that execute an order of characteristics k at venue v in period t is given by:

$$A_{kvt}(x_t, p_t, \delta_t; \theta) = \{(z_i, \varepsilon_{ik0t}, \ldots, \varepsilon_{ikVt}) \,|u_{ikvt} > u_{ikgt} \forall g \text{ s.t. } v \neq g\}, \tag{58}$$

where x_t, p_t and δ_t are the vectors of observed characteristics, explicit trading costs and deltas. If the preference shock, ε_{ikvt}, follows an independent and identical extreme value distribution, the probability that intermediary i opts for venue v to execute order with characteristics k in period t is then given by the following multinomial logit type expression:

$$\text{Prob}_{ikvt}(x_t, p_t, \delta_t; \theta, k) = \frac{e^{\delta_{jvt} + \mu_{ikvt}}}{1 + \sum_q e^{\delta_{jqt} + \mu_{ikqt}}}. \tag{59}$$

The predicted market-level share of venue v for instrument j in each period t is obtained by integrating over the distribution of intermediaries trading profiles and order characteristics (z_i, k):

$$s_{jvt}(x_t, p_t, \delta_t; \theta) = \int_{A_{vt}} \frac{e^{\delta_{jvt} + \mu_{ikvt}}}{1 + \sum_q e^{\delta_{jqt} + \mu_{ikqt}}} dP^*(z, k), \tag{60}$$

where $P^*(z, k)$ denotes the population joint distribution function of the intermediary types and order characteristics (z_i, k), not necessarily independent.

4.3.1 Identification and Estimation Procedure

I move on to specify the estimation procedure. In what follows, I assume that the joint distribution of the intermediary types and order characteristics is known. However, the procedure can easily be modified for the case where that distribution is unknown by assuming a distribution and estimating its unknown parameters jointly with the remaining parameters of the model. The estimation algorithm encompasses four steps that I now describe.

Step 1 *Set initial values for the mean utilities, δ_t, and for the parameters of estimation, θ*

Step 2 *Approximate the predicted market-level shares*

The predicted market-level shares have no closed form expression. I follow Pakes (1986), Pakes and Pollard (1989), and McFadden (1989) and approximate that intractable integral using a (smooth) simulation estimator. I therefore draw ns pseudo-random vectors of unobserved intermediary attributes $(z_1^r, \ldots, z_{ns}^r)$ and order characteristics $(k_1^r, \ldots, k_{ns}^r)$ from $dP^*(z, k)$, which I use to compute $\delta_{jvt} + \mu_{ikvt}^r$, given the initial values for δ_t and θ:

$$\mu_{ikvt}^r = -p_{vt}(z_i^r, k^r) + \left[b_{jvt}, x_{jvt}^{'}\right] \theta^o z_i^r. \tag{61}$$

The smooth estimator that simulates the aggregate market-level shares is, then, given by:

$$s_{jvt}(x_t, p_t, \delta_t; \theta, P^{ns}) = \frac{1}{ns} \sum_{i=1}^{ns} \frac{e^{\delta_{jvt}+\mu_{ikvt}^r}}{1 + \sum_q e^{\delta_{jqt}+\mu_{ikqt}^r}}, \tag{62}$$

where P^{ns} denotes the empirical distribution of the simulation draws. This estimator has two advantages relatively to other simulation estimators.[15] First, it integrates over the ε's analytically and therefore limits the simulation error in the sampling process. Second, it is instrumental in obtaining simulated market-level shares that are smooth functions, positive and sum to one. Nevertheless, as Berry, Linton, and Pakes (2004) point out, the introduction of simulation error influences the asymptotic distribution of the estimator and, therefore, needs to be explicitly taken it account. On this subject please see Step 4 below.

Step 3 *Estimate the econometric error, ξ_{jvt}, as a function of the parameters of estimation θ*

The mean utility δ_{jvt} can not be solved for analytically. However, BLP showed that, for a given θ, it is possible to solve recursively for the unique δ_{jvt} that matches the simulated market-level shares, $s_{jvt}(x, p_t, \delta_t; \theta, P^{ns})$ with the observed ones, s_{jvt}^n, for all j and t. In

[15] Please see Berry, Levinsohn, and Pakes (1995) for a detailed survey on the optimal importance sampling simulator, and the appendix to Nevo (2000) for an analysis on the naive frequency estimator.

particular they show that the operator $\delta^k_{jvt}(\theta)$ is a contraction mapping with modulus less than one:

$$\delta^k_{jvt}(\theta) = \delta^{k-1}_{jvt}(\theta) + \ln\left[s^n_{jvt}\right] - \ln\left[s_{jvt}\left(x_t, p_t, \delta^{k-1}_t; \theta, P^{ns}\right)\right], \tag{63}$$

and therefore its iteration converges geometrically to an unique fixed point. Denote this fixed point by $\delta_{jvt}(s^n_t, \theta, P^{ns})$ where s^n_t represents the vector of observed aggregate market-level shares. Given the unique fixed point, it is relatively straigthforward to obtain an estimate of the econometric error as a function of the data, (x, p_t, s_t), the parameters of estimation, θ, and the simulation process, P^{ns}:

$$\xi_{jvt}(s^n_t, \theta, P^{ns}) = \delta_{jvt}(s^n_t, \theta, P^{ns}) + \gamma b_{jvt} - x'_{jvt}\beta. \tag{64}$$

Step 4 *Estimate the parameters θ*

Estimate the θ parameters by a Generalized Method of Moments (GMM). The approach relies on an identifying restriction on the distribution of the true unobserved characteristics and is based on the sample analogue to the population condition. The standard identifying restriction states that, at the true values of the parameters, the true econometric error, $\xi_{jvt}(s^\infty_t, \theta, P^\infty)$ for $n = ns = \infty$ is mean independent of a set of M instruments $Z_{it} = [z_{1it}, \ldots, z_{Mit}]$:

$$E\left[\xi_{jvt}(s^\infty_t, \theta, P^\infty) \,|Z_{jt}\right] = 0, \tag{65}$$

where ξ_{jvt} denotes the unobserved (to the econometrician) valuation of instrument j at venue v in period t. Instrumental variables techniques are required because of the possible correlation between trading costs and the econometric error term. This correlation is to be expected since venues set trading costs based on information that the econometrician does not possess and, therefore, is compelled to include in the econometric error term. Note that other identifying restrictions would also enable the estimation of the model. In particular, given the typical panel structure of the data, an alternative assumption could incorporate the likelihood of the econometric error and the set of instruments to be more similar for a given venue across time, than for those of different venues. Please see BLP and Davis (2006a) for a more detailed analysis on this subject.

The above population moment conditions can be used, akin to Hansen (1982), to render a method of moments estimator of θ^*, by interacting the estimated econometric error with the set of instruments, and search for the value of the parameters, θ, that set the sample analogues of the moment conditions as closed as possible to zero. Let $G_{n,ns}(\theta)$ denote the sample analogues of the moment conditions:

$$G_{n,ns}(\theta) = \frac{1}{n}\sum_{t=1}^{T}\sum_{v=1}^{V}\sum_{j=1}^{J^{vt}} \xi_{jvt}\left(s_t^n, \theta, P^{ns}\right) Z_{jvt} = \frac{1}{n}\sum_{t=1}^{T}\sum_{v=1}^{V}\sum_{j=1}^{J^{vt}} \psi(\theta). \tag{66}$$

Formally, the method of moments estimator, $\hat{\theta}$, is the argument that minimizes the weighted norm criterion of $G_{n,ns}(\theta)$, for some weighting matrix A_n with rank at least equal to the dimension of θ:

$$\hat{\theta} = \arg\min_{\theta} \quad \|G_{n,ns}(\theta)\|_{A_n} = G_{n,ns}(\theta)' A_n G_{n,ns}(\theta). \tag{67}$$

The strong non-linearity of the objective function requires a minimization routine. The standard practice in the literature has been to use either the Nelder-Mead (1965) nonderivative "simplex" search method or a quasi-Newton method with an analytic gradient (see Press *et al.*, 1994). The latter has the important (computational) advantage of being two orders of magnitude faster than the former. However, because the first method is more robust and less sensitive to starting values, I perform the search using the Nelder-Mead (1965) nonderivative "simplex" search.

The non-linear search over θ can be simplified by making use of the fact that the first order conditions for a minimum of $\|G_{n,ns}(\theta)\|_{A_n}$ are linear for the subset $\theta_1 = (\gamma, \beta)$ of the parameters of estimation, $\theta = (\theta_1, \theta^u)$. Consequently, it is possible, given the standard instrumental variables results, to express θ_1 as function of θ^u, and limit the non-linear search over θ^u:

$$\theta_1 = \left(Q'ZA_n^{-1}Z'Q\right)^{-1} Q'ZA_n^{-1}Z'\delta(\theta^u). \tag{68}$$

where Q denotes the matrix of trading costs and observed characteristics, Z denotes the matrix of instruments, and, finally, δ denotes the matrix of mean utilities, expressed only in terms of θ^u after concentrating out θ_1.

4.4 Empirical Analysis

4.4.1 Data Description

I apply the model to the set of 16 most traded securities in the FTSE 100 following the list of liquid securities published (and updated regularly) by CESR after the implementation of MiFID. REUTERS Market Shares Reports provided information on the top trading venues. I follow Pinkse and Slade (2004) with regard to the criterion of which venues to include in the sample and include those that accounted for at least one percent of the market in volume: the London Stock Exchange, Chi-X and the systematic internalizers aggregated in Markit Boat.

For each security and trading venue, I collected daily information from DATASTREAM on the official price, ask and bid prices, the number of trades, and the number of shares traded. For both Chi-X and the systematic internalizers aggregated in Markit Boat information on the number of trades and the number of shares traded was obtained directly.

Market size was assumed to be the total number of shares traded per security across all possible trading venues and was collected via DATASTREAM. Trading venue market shares were then computed as the ratio of the corresponding number of shares traded over market size.

Information on execution, settlement and clearing fees was obtained directly via the published fee schedules. In what concerns the systematic internalizers in Markit Boat, these information was obtained from JP Morgan MiFID Report II that discriminates the average execution, settlement and clearing costs of a systematic internalizer. Given that typically (although not always) those fee schedules are a function of each financial intermediary trading profile, I considered the fees that would arise for the four types of intermediaries specified in European Commission (2006): 1) typical volume and value trades, 2) large volume of low value trades, 3) large volume of high value trades, and 4) small volume of low value trades.

I measured the implicit trading costs by effective spreads. The effective spread is defined as the difference between the transaction price and the current mid-quote for time period t:

$$es_{jt} = |P_{jt} - M_{jt}|, \tag{69}$$

where M_{jt} is the quote mid-point, i.e. $(A_{jd} + B_{jd})/2$, A_{jt} denotes the ask price, B_{jt} the bid price, and P_{jt} the effective transaction price of instrument j in period t. This measure takes into account the fact that trades can occur either inside or outside the quoted spread.

TABLE 4.4
Summary Statistics

Variable	Venue	Mean	Std	Min	Max
	CHX	0.03	0.03	0.00	0.14
Market Share	LSE	0.70	0.06	0.47	0.92
	SI	0.04	0.03	0.00	0.19
	CHX	13.14	10.05	1.50	40.87
Price (£)	LSE	13.14	10.05	1.49	40.90
	SI	13.06	9.97	1.50	40.90
	CHX	0.09	0.56	0.00	6.87
Effective Spread (£)	LSE	0.05	0.02	0.00	0.30
	SI	0.06	0.11	0.00	0.77
	CHX	1.36	1.22	0.20	5.30
Volume per trade ('000 shares)	LSE	3.75	4.24	0.39	26.70
	SI	42.67	78.75	0.25	852.04

Therefore, it incorporates both the impacts of market spreads and market impact on trading costs, even if it does not allow the separation of the two effects. Microstructure literature has shown that the effective spread reflects expected losses to informed traders (Glosten and Milgrom(1985), Copeland and Galai (1983)), inventory costs (Stoll (1978), Amihud and Mendelson (1980), Ho and Stoll (1981)) and order processing costs (Stoll (1985)).

Finally, I follow Stoll (2000) and Jain (2001) and median the different variables at a weekly frequency to reduce measurement errors due to random daily fluctuations.

Table 4.4 presents some general statistics for the resulting dataset ranging from the first week of November 2007 to the last week of March 2008. Several interesting trends are noteworthy. The incumbent venue - LSE - clearly dominates the industry with an average market share of 70% - against 3%-4% for each of the competing venues. There is no significant difference in the price at which securities are traded, but the bid-ask spread is lower at LSE with an average effective spread of £0.05 against £0.06-£0.09 on the competing venues. The statistics on the volume per trade suggest a clear segmentation of the industry, with the different venues attracting distinct type of orders. Chi-X attracts the lowest average volume per trade, the SI attract the highest average volume per trade, and LSE positions itself between those two. As there is no significant difference in the price securities are traded, the heterogeneity in volume per trade carries to the turnover per trade.

Total fees are a function of each financial intermediary trading profile in terms of volume and value. For this reason they are not presented in the summary statistics table. For

FIGURE 4.4

Data: Decomposition of Total Fees per Trade

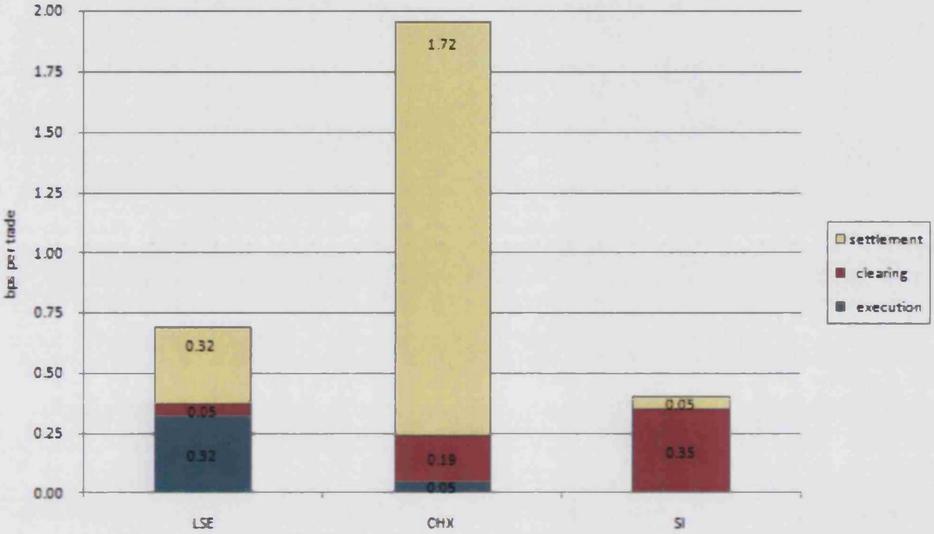


illustration purposes, Figure 4.4 plots the total fees (and corresponding decomposition) that would arise for the typical financial intermediary following the European Commission (2006) classification. It is clear that Chi-X offers the lowest execution fee, but its competitiveness is penalized due to high clearing and settlement fees.

4.4.2 Demand Identification

Total fees are typically set taking into account information that the researcher does not possess and, therefore, has to include in the econometric error term. Moreover, effective spreads are the outcome of unobserved information to the researcher. As a consequence, total fees and effective spreads are expected to be correlated with the error term. As discussed above, instrumental variables techniques are therefore required. The use of securities- and venue-specific fixed effects decreases the requirements on the instruments needed for a consistent estimation. However, it does not eliminate completely their need since both fees and spreads are likely to still be correlated with unobserved time-specific deviations from the overall mean valuations.

In the lines of Arellano and Bond (1991), and Arellano and Bover (1995), I use lag liquidity values as instruments for both total fees and effective spreads under the assumption that those lags are uncorrelated with the error term and, at the same time, correlated with the endogenous variables that needs instrumenting. Please see the demand estimation section

TABLE 4.5
*Results from Full Model**

Variable	Means (γ,β's)	Standard Deviations Σ	Interactions with log (Order Size) Π
Constant	0.666	0.000	-0.096
	(0.165)	(0.006)	(0.001)
log (total fees)	—	0.000	-0.017
	—	(0.000)	(0.021)
Effective Spread	1.552	0.357	-1.109
	(0.128)	(0.001)	(0.000)
Chi-X dummy	-1.429	—	—
	(0.098)	—	—
SI dummy	-1.320	—	—
	(0.099)	—	—

* Regression based on 1008 observations. Security, venue and week dummy variables are included as controls. Asymptotically standard errors in parentheses.

for more details.

4.4.3 Demand Estimation

I now move on to discuss the random-coefficients multinomial logit demand model estimation. The estimated specification includes total fees and effective spreads covariates as observed attributes, while controlling (at least in part) for unobserved attributes by allowing security, venue and week fixed effects. The log transformation of the total fees variable was used to reduce skewness.

The coefficients on fees and liquidity are allowed to be intermediary specific in order to capture the fact that the valuations of the different elements in the allocation rule can depend on the characteristics of intermediaries. In particular, I model the intermediaries trading profiles z_i as follows:

$$\begin{pmatrix} \alpha_i \\ \gamma_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} 0 \\ \gamma \\ \beta \end{pmatrix} + \theta^o z_i^e = \begin{pmatrix} 0 \\ \gamma \\ \beta \end{pmatrix} + \Pi o_i + \Sigma v_i, \tag{70}$$

where o_i is a denotes the log transformation of order size from intermediary i - drawn from the Chi-X, LSE and SI order books,[16] v_i is a 3×1 vector of random-variables drawn from

[16] I sampled 500 intermediaries per week and security.

TABLE 4.6
*Median Estimated Demand Elasticities**

	CHX	LSE	SI
Market share with respect to Total Fees			
CHX	-0.069	0.067	0.003
LSE	0.001	-0.021	0.002
SI	0.001	0.037	-0.067
Market share with respect to Effective Spread			
CHX	-2.877	0.031	0.001
LSE	0.000	-0.891	0.000
SI	0.002	0.042	-2.812

* The elasticity in a cell gives the percent change in market share of the row's venue with a one percent change in the variable of the column's venue.

a normalized multivariate normal distribution, Π is a 3×1 matrix of order size coefficients, and Σ is a 3×3 diagonal matrix that scales the effect of v_i.

Table 4.5 reports the estimated GMM results. The first column reports the estimates of the different coefficients means, whereas the other columns present estimates with regard to the associated heterogeneity. The coefficients are of the expected sign, suggesting that market shares react negatively to both total fees and liquidity. Figures 4.5 and 4.6 display the predicted distribution for each of those coefficients. Most of the heterogeneity is due to order size, with the magnitude of the coefficients on the unobserved intermediaries characteristics (v_i) being small. As expected, intermediaries with higher order sizes tend to be more sensitive to both fees and liquidity.

In order to evaluate the impact of both fees and spreads on market shares, I examine the own- and cross-price elasticities for both variables. Table 4.6 reports the estimated elasticities computed using the estimates in Table 4.5. The top panel of the table displays elasticities with respect to total fees. The results suggest that all venues enjoy a certain degree of market power. Conditional on the venues bid-ask spreads, intermediaries are estimated to have a low price sensitivity. A one percent decrease in the total fees charged by each venue is estimated to impact only marginally the corresponding market shares. A possible justification may lie on the network effects that characterize the industry. As intermediaries value both low cost and high liquidity, a decrease in the total fees charged by a given venue may not be sufficient to induce a substantial demand change. The bottom panel of the table displays elasticities with respect to effective spread. The results support the important role of liquidity on the choice of venue. Conditional on the venues total fees, a one percent increase in the effective spread is estimated to decrease the market-share of LSE in almost 1% and of CHX or SI in almost 3%. Contrasting the elasticities in the two panels suggests that liquidity may play a

Figure 4.5
Frequency Distribution for Total Fees Coefficient

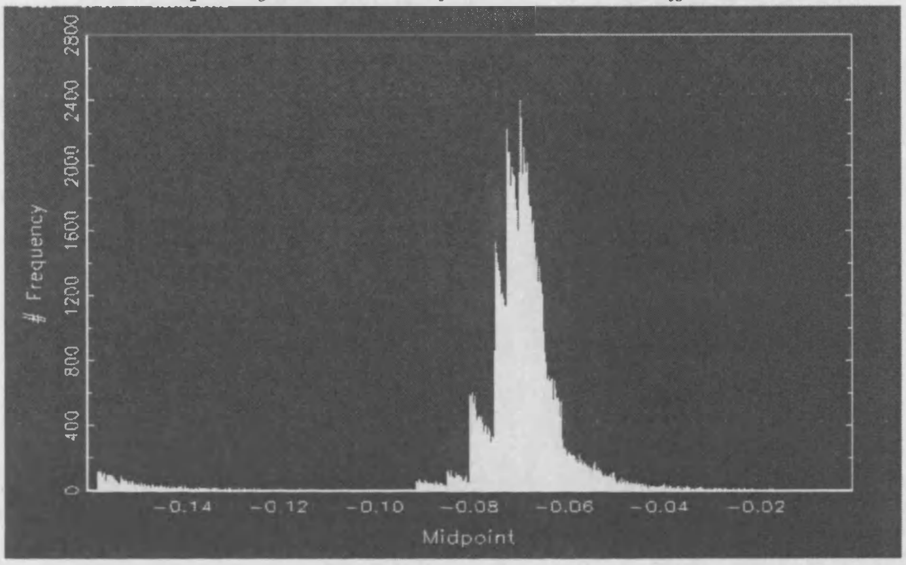


Figure 4.6
Frequency Distribution for Liquidity Coefficient

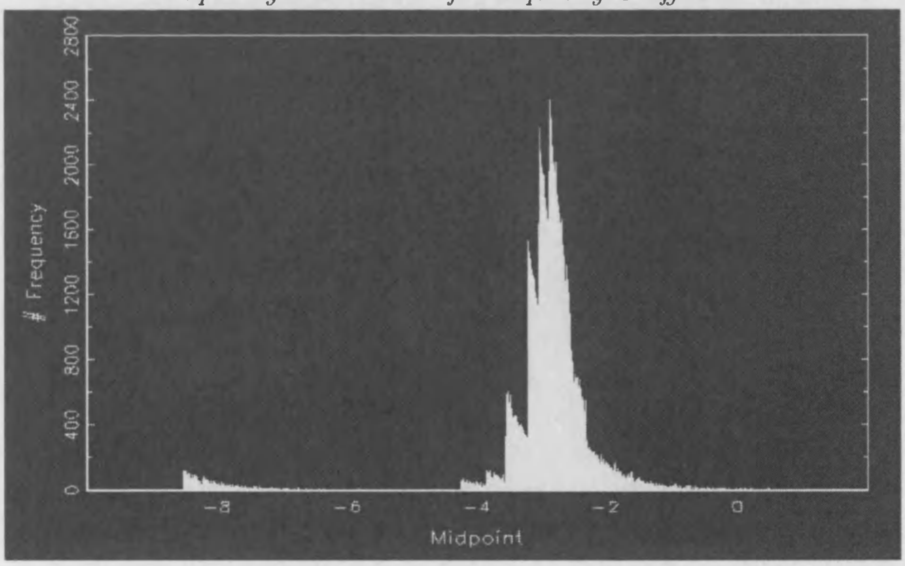


TABLE 4.7

*Liquidity Equation**

Explanatory Variable	Effective Spread	
Market Share		
CHX	-3.645	(1.092)
LSE	-1.271	(0.378)
SI	-4.512	(1.142)
log (volume)	0.019	(0.005)
Price	0.006	(0.001)
Volatility	1.886	(0.165)
Unobserved Characteristics		
CHX	0.274	(0.069)
LSE	0.688	(0.340)
SI	0.187	(0.063)
R^2	0.350	

* Regression based on 1008 observations. Standard errors in parentheses.

more important role than total fees in the choice of the venue where to route a given order.

However, liquidity and fees are clearly not exogenous relatively to each other - one would expect venues to take into consideration liquidity when setting fees, as well as liquidity to be a function of the fees schedules. Given the highly endogenous nature of liquidity, the above elasticities have to be cautiously interpreted. The elasticities with regard to total fees are *conditional* on the liquidity levels and the elasticities with regard to liquidity are *conditional* on the fees levels. I now move on to examine the endogeneity between liquidity and fees.

Micro-finance theory implies that liquidity may be potentially a non-linear function of a series of factors that affect both the demand and supply for trading. I follow Goolsbee and Petrin (2004) and introduce a reduce-form approach that estimates a liquidity equation as a function of those factors. I include, in line with Stoll (2000), Wahal (1997) and Weston (2002), venue market share, share volume, price, and share volatility[17]. However, in contrast

[17] Share volatility is defined, following Ding and Charoenwong (2003), as the standard deviation over the average of the quoted mid-point within each time period,

$$SV_{jt} = \frac{sd\,[M_{jd}]}{mean\,[M_{jd}]},$$

where $sd\,[\cdot]$ represents the standard deviation taken over the days included in period t.

TABLE 4.8

*Median Estimated Liquidity Elasticities**

	CHX	LSE	SI
Effective Spread with respect to Total Fees			
CHX	0.351	-0.342	-0.018
LSE	-0.221	3.766	-0.391
SI	-0.009	-0.304	0.558

* The elasticity in a cell gives the percent change in the liquidity of the row's venue with a one percent change in the total fees of the column's venue.

with those studies, I control for unobserved venue- and security-specific characteristics:

$$\xi_{jvt} = \delta_{jvt} + \gamma b_{jvt} - x_{jvt}^{'}\beta, \tag{71}$$

where δ_{jvt},γ and β are demand side estimates.

Table 4.7 presents the instrumental variables results of the liquidity equation regression.[18] Most of the coefficients are of the expected sign, with an increase in the market share lowering the effective spread and an increase in price and volatility being associated with increases in spreads. The liquidity equation is instrumental in understanding the impact of total fees on effective spreads. Total fees influence relative market shares which in turn determine venue liquidity. In order to evaluate the total impact of fees on spreads, I computed *unconditional* own- and cross-price median elasticities as follows:

$$\frac{\partial b_{jvt}}{\partial p_{qt}}\frac{p_{qt}}{b_{jvt}} = \frac{\partial b_{jvt}}{\partial s_{jvt}}\varepsilon_{vq}^{p}\frac{s_{jvt}}{b_{jvt}}, \tag{72}$$

where ε_{vq}^{p} denotes the cross-total fees elasticity between venues v and q. Table 4.8 reports the estimated elasticities.

4.5 Barriers to Competition

After estimating the degree of substitutability between the different trading venues, I move on to evaluate the barriers to competition induced by network effects and post-trading constraints. In order to examine the impact of network effects as a barrier to competition, I propose to compute the counterfactual market shares that would arise if there were no liquidity differences across venues (although still allowing for heterogeneity across the securities traded). In particular, I consider the case where the effective spread for each security-week

[18] Log transformation of the volume variable was used to reduce skewness.

TABLE 4.9

Barriers to Competition: Counterfactual (Median) Results

Panel A: Liquidity as a Barrier					
	Current Market Shares		Counterfactual Market Shares		
CHX	0.021		0.094		
LSE	0.703		0.426		
SI	0.037		0.118		

Panel B: Post-Trading as a Barrier					
	Current		Counterfactual		
	Market Shares	Eff. Spreads	Direct Impact Market Shares	Direct Impact Eff. Spreads	Total Impact Market Shares
CHX	0.021	0.015	0.022	0.009	0.094
LSE	0.703	0.005	0.702	0.006	0.445
SI	0.037	0.020	0.037	0.020	0.107

pair is the same across venues and equal to the median of the actual observed spreads. The results - presented in Table 4.9, Panel A - suggest that eliminating the liquidity advantage of the incumbent venue contributes to a less asymmetric industry. Chi-X would benefit less than the SI because of the disadvantage from a post-trading perspective - a point I address below.

The competitiveness of a given venue can also be penalized by higher post-trading costs. I propose to evaluate the barriers to competition induced by post-trading constraints by simulating the equilibrium market shares that would arise if the securities traded in the different trading venues were fungible and intermediaries could choose the post-trading arrangements with the lowest clearing and settlement fees. Allowing intermediaries to freely choose post-trading arrangements is equivalent to an effective decrease in the total fees paid by some of them (those that switch from current arrangements). A decrease in the total fees has a direct impact on relative markets shares and consequently on effective spreads which in turn also influence market shares. Table 4.9, Panel B presents the counterfactual results, discriminating the different effects that would arise.

I should note that I do not compute market equilibrium fees, which is beyond the scope of this chapter (although providing such a framework constitutes a very interesting potential area for future research). I eliminate post-trading constraints, but maintain the same level of clearing and settlement fees. The results suggest that eliminating the post-trading constraints and allowing intermediaries to choose the most competitive post-trading arrangements would also induce a less asymmetric industry - of the same order of magnitude as of eliminating the network effect.

4.6 Conclusions

As of 1 November 2007, the Market in Financial Instruments Directive (MiFID) aims to increase competition and to foster client protection in the European financial market. Among other provisions, it abolishes the concentration rule and challenges the market power of existing trading venues. The directive introduces venue competition in order to achieve better execution and ultimately lower trading costs. I argue that, although positive, the impact on the degree of actual competition may be limited due to two barriers to competition: *i*) direct network effects together with increasing returns to scale and *ii*) post-trading constraints (since venues typically bundle trading and post-trading services).

I empirically examine market dominance and barriers to competition in financial trading venues by addressing the following questions: *i*) evaluate the actual degree of competition between trading venues, *ii*) measure the impact of network effects on competition, and finally *iii*) assess the barriers to competition induced by the bundle of trading and post-trading services.

The results imply that financial intermediaries tend to value liquidity more than total fees when deciding where to route a given order for execution. For this reason, the incumbent venue has a clear advantage relatively to its competitors and can, as a result, exert market power when setting total fees. Furthermore, eliminating the mentioned barriers to competition seems to be associated with a significant decrease (of a similar magnitude) in the asymmetry of the industry. As a consequence, policies that promote competition on the post-trading after market are instrumental in boosting the effectiveness of MiFID.

Appendix A

Appendix to Chapter 3

A.1 Parameter Restrictions

Specification A

This appendix presents the set of parameter restrictions that ensure specification A function $H^A(r, y; \theta, \Im)$, satisfies properties 3 and 4 for a globally consistent indirect utility function. The set of restrictions resemble conditions 3.15 and 3.16 in Diewert (1971) that ensure the generalized Leontief function can be interpreted as the cost function corresponding to some underlying production possibilities set.

Assume that the vectors r and y are strictly positive. We begin by addressing property 3. In order for $H^A(r, y; \theta, \Im)$ to be strictly increasing in y and nonincreasing in p_i (or in other words, nondecreasing in r_i) for any $i \in \Im$, the following $J + 1$ inequalities need to be satisfied:

$$\begin{aligned} H_y^A(r, y; \theta, \Im) &= c_0 + \textstyle\sum_{j=1}^J c_j r_j > 0 \\ H_i^A(r, y; \theta, \Im) &= a_i + 1/2 \textstyle\sum_{j=1}^J (b_{ij} + b_{ji}) r_j + c_i y \geq 0, \end{aligned}$$

for any $i \in \Im$.

We now turn to property 4 and present the set of restrictions that ensure $H^A(r, y; \theta, \Im)$ is quasiconvex. It is well known that every convex function is quasiconvex. Furthermore, a twice differentiable function is convex over a convex set if and only if its matrix of second partial derivatives is positive semidefinite for every point in the set.

If symmetry is imposed on matrix B (please see the symmetry subsection for more details), the second partial derivatives matrix of $H^A(r, y; \theta, \Im)$ will be symmetric. A symmetric matrix is positive definite (and automatically positive semidefinite) if and only if all its leading principal minors are strictly positive. If, on the other hand, no symmetry on matrix B is assumed *a priori*, the second partial derivatives matrix of $H^A(r, y; \theta, \Im)$ will not necessarily be symmetric. An arbitrary matrix is positive definite (and automatically positive semidefinite) if and only if its Hermitian part is positive definite. In other words, an arbitrary matrix is positive definite if and only if all the leading principal minors of its Hermitian part are strictly positive.

Let D_k denote the determinant of the kth order principal submatrix of the second partial derivatives matrix of $H^A(r, y; \theta, \Im)$. For specification A algebraic functional form to be quasiconvex, the following $J + 1$ inequalities (the formulation nests both the case where symmetry is imposed on matrix B and the case where it is not) need to be satisfied:

$$\begin{aligned} D_1 &= b_{11} > 0 \\ D_2 &= b_{11}b_{22} - (1/4)(b_{12} + b_{21})^2 > 0 \\ &\vdots \\ D_{J+1} &> 0. \end{aligned}$$

Specification B

This appendix presents the set of restrictions on the parameters that ensure that specification B function, $H^B(r, y; \theta, \Im)$, satisfies properties 3 and 4 for a globally consistent indirect utility function. Again, the set of restrictions resemble conditions 3.15 and 3.16 in Diewert (1971) that ensure the generalized Leontief function can be interpreted as the cost function corresponding to some underlying production possibilities set.

Assume that the vectors r and y are strictly positive. We begin by addressing property 3. In order for $H^B(r, y; \theta, \Im)$ to be strictly increasing in y and nonincreasing in p_i (or in other words, nondecreasing in r_i) for any $i \in \Im$, the following J inequalities need to be satisfied:

$$H_i^B(r, y; \theta, \Im) = a_i + 1/2 \sum_{j=0}^{J} (b_{ij} + b_{ji}) r_j y \geq 0,$$

for any $i \in \Im$.

We now turn to property 4 and present the set of restrictions that ensure $H^B(r, y; \theta, \Im)$ is a quasiconvex. It is well known that every convex function is quasiconvex. Furthermore, a twice differentiable function is convex over a convex set if and only if its matrix of second partial derivatives is positive semidefinite for every point in the set.

If symmetry is imposed on matrix B (please see the symmetry subsection for more details), the second partial derivatives matrix of $H^B(r, y; \theta, \Im)$ will be symmetric. A symmetric matrix is positive definite (and automatically positive semidefinite) if and only if all its leading principal minors are strictly positive. If, on the other hand, no symmetry on matrix B is assumed *a priori*, the second partial derivatives matrix of $H^B(r, y; \theta, \Im)$ will not necessarily be symmetric. An arbitrary matrix is positive definite (and automatically positive semidefinite) if and only if its Hermitian part is positive definite. In other words, an arbitrary matrix is positive definite if and only if all the leading principal minors of its Hermitian part are strictly positive.

Let D_k denote the determinant of the kth order principal submatrix of the second partial derivatives matrix of $H^B(r, y; \theta, \Im)$. For specification B algebraic functional form to be quasiconvex, the following $J + 1$ inequalities (the formulation nests both the case where symmetry is imposed on matrix B and the case where it is not) need to be satisfied:

$$\begin{aligned}
D_1 &= b_{11} > 0 \\
D_2 &= b_{11}b_{22} - (1/4)(b_{12} + b_{21})^2 > 0 \\
&\vdots \\
D_{J+1} &> 0.
\end{aligned}$$

Appendix B

Appendix to Chapter 4

B.1 Data

TABLE B.1

List of Securities used in the Demand Estimation

Anglo American
Barclays
BG
BHP Billiton
BP
Glaxosmithkline
HBOS
HSBC
Lloyds TSB
Prudential
Reckitt Benckiser
Royal Dutch Shell B
Standard Chartered
Tesco
Vodafone
Xstrata

Bibliography

ACKERBERG, D., L. BENKARD, S. BERRY and A. PAKES, 2005, "Econometric tools for analyzing market outcomes." Mimeo.

AGUIRREGABIRIA, V., 2002, "Sales promotions in supermarkets: Estimating their effects on profits and consumer welfare." Mimeo.

ANAND, B. N. and R. SHACHAR, 2004, "Advertising, the matchmaker." Mimeo.

ANOLLI, M., 2007, "The impact of MiFID on the European securities industry: a simulation of the internalizer behavior on the Italian stock market." Mimeo.

ARELLANO, M. and S. R. BOND, 1991, "Some specifications tests for panel data: Monte Carlo evidence and an application to employment equations." *Review of Economic Studies*, 58, 277-298.

ARELLANO, M. and O. BOVER, 1995, "Another look at the instrumental variables estimation of error-component models." *Journal of Econometrics*, 68, 29-51.

AVGOULEAS, E. and S. DEGIANNAKIS, 2005, "The impact of the EC financial instruments markets Directive on the trading volume of EU equity markets." Mimeo.

BARNETT, W. A., 1983, "Definitions of 'second order approximation' and of 'flexible functional form'." *Economic Letters*, 12: 31-35.

BENITEZ-SILVA, H., G. HALL, G. J. HITSCH, G. PAULETTO, and J. RUST, 2000, "A comparison of discrete and parametric approximation methods for solving dynamic programming problems in economics." Mimeo.

BERNDT, E. R., M. A. FUSS and L. WAVERMAN, 1977, "Dynamic models of the industrial demand for energy." Electric Power Institute, EA-580, Palo Alto.

BERRY, S. and P. A. HAILE, 2008, "Nonparametric identification of multinomial choice models with heterogeneous consumers and endogeneity." Mimeo.

BERRY, S. and A. PAKES, 2005, "The pure characteristics demand model." Mimeo.

BERRY, S., O. B. LINTON and A. PAKES, 2004, "Limit theorems for estimating the parameters of differentiated product demand systems." *Review of Economic Studies*, 71: 613-654.

BERRY, S., J. LEVINSOHN, and A. PAKES, 1995, "Automobile prices in market equilibrium." *Econometrica*, 63, 841-890.

BERRY, S., 1994, "Estimating discrete-choice models of product differentiation." *Rand Journal of Economics*, 25, 242-262.

BLATTBERG, R. C., G. D. EPPEN, and J. LIEBERMAN, 1981, "A theoretical and empirical evaluation of price deals for consumer nondurables." *Journal of Marketing*, 45, 116-129.

BROWN, B. W. and M. B. WALKER, 1989, "The random utility hypothesis and inference in demand systems." *Econometrica*, 59, 815-829.

CANTILLON, E. and P. YIN, 2007, "How and when do markets tip? Lessons from the battle of the Bund." *ECB Working Paper Series*, 766.

CARVALHO, C. H., 2004, "Cross-Border Securities Clearing and Settlement Infrastructure in the European Union as a Prerequisite to Financial Markets Integration: Challenges and Perspectives." *HWWA Discussion Paper* 287.

CHAMBERLAIN, 1985, "Heterogeneity, omitted variable bias, and duration dependence." In J. J. Heckman and B. Singer (Eds.), *Longitudinal analysis of labor market data* (pp. 3-38). N^o 10 in Econometric Society Monographs series, Cambridge, New York and Sidney: Cambridge University Press.

CHAN, T. Y., 2006, "Estimating a continuous hedonic-choice model with an application to demand for soft drinks." *RAND Journal of Economics*, 37: 466-482.

CHEN, J, U. DORASZELSKI, and J. HARRINGTON Jr., 2007, "Avoiding market dominance: product compatibility in markets with network effects." Mimeo.

CHINTAGUNTA, P. K., D. C. JAIN, and N. J. VILCASSIM, 1991, "Investigation heterogeneity in brand preferences in logit models for panel data." *Journal of Marketing Research*, 28, 417-428.

CHRISTENSEN, L. R., D. W. JORGENSON and L. J. LAU, 1975, "Transcendantal logarithm utility functions." *American Economic Review*, 65: 367-383.

CONSLIK, J., E. GERSTNER and J. SOBEL, 1984, "Cyclic pricing by a durable goods monopolist." *Quarterly Journal of Economics*, 99, 489-505.

DAVIS, P., 2006a, "The discrete choice analytically flexible (DCAF) model of demand for differentiated products." Mimeo.

—, 2006b, "Spatial competition in retail markets: movie theaters." *RAND Journal of Economics*, 37, 964-982.

—, 2005, "System estimation of multi-way error components models with unbalanced data structures." Mimeo.

—, 2002, "Estimating multi-way error components models with unbalanced data structures." *Journal of Econometrics*, 106: 67-95.

DEATON, A. and J. MUELBAUER, 1980, "An almost ideal demand system." *American Economic Review*, 312-326.

DING, D. and C. CHAROENWONG, 2003, "Bid-ask spreads, volatility, quote revisions, and trades of thinly traded futures contracts." *The Journal of Futures Markets*, 23, 455–486.

DIEWERT, E., 1974, "Functional forms for revenue and factor requirements functions." *International Economic Review*, 15: 119-130.

—, 1973, "Functional forms for profit and transformation functions." *Journal of Economic Theory*, 6: 284-316.

—, 1971, "An application of the Shephard duality theorem: a generalized Leontief production function." *Journal of Political Economy*, 79: 481-507.

DUBE, J-P., J. T. FOX and C-L. SU, 2009, "Improving the numerical performance of BLP static and dynamic discrete choice random coefficients demand estimation." Mimeo.

DUBE, J. H., G. J. HITSCH and P. E. ROSSI, 2006, "Do switching costs make markets less competitive?" Mimeo.

DUBOIS, P., A. HERNANDEZ-PEREZ, and M. IVALDI, 2006, "Academic journals as two-sided platforms: empirical evidence from data on French libraries." Mimeo.

ELLISON, S. F., I. COCKBURN, Z. GRILICHES and J. HAUSMAN, 1997, "Characteristics of demand for pharmaceutical products: an examination of four cephalosporins." *RAND Journal of Economics*, 28, 426-446.

ERDEM, T., S. IMAI and M. P. KEANE, 2003, "Brand and quantity choice dynamics under price uncertainty." *Quantitative Marketing and Economics*, 1(1), 5-64.

EUROPEAN COMMISSION, 2006, "Competition in EU securities trading and post-trading." *Working Paper of the Commission Financial Services.*

FAISON , E. W. J., 1977, "The neglected variety drive." *Journal of Consumer Research*, 4, 172-175.

FOX, J. T. and A. GANDHI, 2008, "Identifying heterogeneity in economic choice and selection models using mixtures." Mimeo.

GARVEY, R. and A. MURPHY, 2006, "Crossed markets: arbitrage opportunities in Nasdaq ttocks." *The Journal of Alternative Investments*, 23, 46–58.

GOLDFARB, A., 2006, "State dependence at internet portals". *Journal of Economics & Management Strategy*, 15(2), 317-352.

GONUL, F. and K. SRINIVASAN, 1993, "Modeling unobserved heterogeneity in multinomial to logit models: methodological and managerial implications," *Marketing Science*, 12, 213-229.

GOOLSBEE, A. and A. PETRIN, 2004, "The consumer gains from direct broadcast satellites and the competition with cable TV." *Econometrica*, 72, 351–381.

GORMAN, W., 1971, "Two stage budgeting." Mimeo.

—, 1953, "Community preference fields," *Econometrica*, 21(1), 63-80.

GOURVILLE, J. T. and D. SOMAN, 1998, "Payment depreciation: the behavioral effects of temporally separating payments from consumption." *Journal of Consumer Research*, 25(2), 160-174.

GRESSE, C. , 2006, "The effect of crossing-network trading on dealer market's bid-ask spreads." *European Financial Management*, 12, 143-160.

HAMILTON, J. L., 1979, "Market fragmentation, competition, and the efficiency of the stock exchange." *The Journal of Finance*, 34, 171-187.

HANSEN, L. P., 1982, "Large sample properties of generalized method of moments estimators." *Econometrica*, 50, 1029-1054.

HARTMANN, W. R. and V. B. Viard, 2008, "Do frequency reward programs create switching costs? A dynamic structural analysis of demand in a reward

program." *Quantitative Marketing and Economics*, 6, 109-137.

HARTMANN, W. R., 2006, "Intertemporal effects of consumption and their implications for demand elasticity estimates." *Quantitative Marketing and Economics*, 4, 325-349.

HAUSMAN, J., 1994, "Valuation of new goods under perfect and imperfect competition." *NBER Working Paper 4970.*

—, G. LEONARD and D. ZONA, 1994, "Competitive analysis with differentiated products," *Annales d'Economie et de Statistique*, 34: 159-180.

— and D. WISE, 1978, "A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences," *Econometrica*, 46:

HECKMAN, J., 1981, "Heterogeneity and state dependence." In S. Rosen (Ed.), *Studies in labor markets* (pp. 91-139). Chicago: University of Chicago Press.

HENDEL, I. and A. NEVO, 2006, "Measuring the implications of sales and consumer inventory behavior." *Econometrica*, 74(6), 1637-1673.

—, 2006, "Sales and consumer inventory."*RAND Journal of Economics*, 37(3), 543-561.

HOCH, S. J., X. DREZE and M. E. PURK, 1994, "EDLP, hi-low, and margin arithmetic." *The Journal of Marketing*, 58(4), 16-27.

HORTACSU, A. and C. SYVERSON, 2004, "Product differentiation, search costs, and competition in the mutual fund industry: a case study of S&P 500 index funds." *The Quarterly Journal of Economics*, 119, 403-456.

INFORMATION RESOURCES, INC., 1993, *Managing your Business in an EDLP Environment.* Chicago: IRI.

JAIN, P., 2001, "Institutional design and liquidity at stock exchanges around the World." Mimeo.

JAIN, D. C., N. J. VILCASSIM and P. K. CHINTAGUNTA, 1994, "A random-coefficients logit brand-choice model applied to panel data." *Journal of Business and Economic Statistics*, 12, 317–328.

JEULAND, A. P. and C. NARASIMHAN, 1985, "Dealing - temporary price cuts - by seller as a buyer discrimination mechanism." *Journal of Business*, 58(3), 295-308.

JEULAND, A. P., 1978, "Brand preference over time: a partially deterministic operationalization of the notion of variety seeking." In S. Jain (Ed.), *Research frontiers in marketing: dialogues and directions.* N^o 42, AMA 1978 Educator's Proceedings, Chicago: American Marketing Association.

JPMORGAN, 2006, "MiFID Report II." *European Equity Research.*

KAHN, B. E., 1995, "Consumer variety seeking among goods and services: an integrative review," *Journal of Retailing and Consumer Services*, 2(3), 139-148.

KAMAKURA, W. A. and G. J. RUSSELL, 1989, "A probabilistic choice model for market segmentation and elasticity structure," *Journal of Marketing Research*, 26, 379-390.

KEANE, 1997, "Modelling heterogeneity and state dependence in consumer choice behavior." *Journal of Business and Economic Statistics*, 15(3). 310-327.

KLEMPERER, P., 1995, "Competition when consumers have switching costs: an overview with applications to industrial organization, macroeconomics, and international trade." *Review of Economic Studies*, 62, 515-539.

KNITTEL, C. R. and K. METAXOGLOU, 2008, "Estimation of random coefficient demand models: challenges, difficulties and warnings." Mimeo.

LAL, R. and R. RAO, 1997, "Supermarket competition: the case of every day low pricing." *Marketing Science*, 16(1), 60-80.

LANCASTER, K., 1971, *Consumer Demand: A New Approach.* Columbia University Press.

LAU, L., 1986, "Functional forms in econometric model building." In *Handbook of Econometrics.* Vol. 3.

MCALISTER, L., 1982, "A dynamic attribute satiation model of variety seeking behavior." *Journal of Consumer Research*, 9, 141-150.

— and E. PESSEMIER, 1982, "Variety-seeking behaviour: an interdisciplinary review." *The Journal of Consumer Research*, 9(3), 311-322.

MCFADDEN, 1989, "A method of simulated moments for estimation of discrete response models without numerical integration". *Econometrica*, 57: 995-1026.

—, 1981a, "Econometric models of probabilistic choice." In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data and Econometric Applications* (pp, 198-272). MA: MIT Press.

—, 1981b, "Structural discrete probability models derived from theories of choice," In C. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data and Econometric Applications*, MA: MIT Press.

—, 1978a, "Modelling the choice of residential location." In A. Karlqvist, L. Lundqvist, F. Snickars and J. Weibull (Eds.), *Spatial Interaction Theory and Planning Models.* North Holland.

—, 1978b, "The general linear profit function." In D. McFadden and M. Fuss (Eds.), *Production Economics: A Dual Approach to Theory and Applications*, North Holland.

—, 1974, "Conditional logit analysis of qualitative choice behavior." In P. Zarembka (Ed.), *Frontiers in Econometrics.* Academic Press.

MENON, S. and B. E. KAHN, 1995, "The impact of context on variety seeking in product choices." *Journal of Consumer Research*, 22, 285-295.

NARASIMHAN, C., 1988, "Competitive promotional strategies." *Journal of Business*, 61(4), 427-429.

NELDER, J.A., and R. MEAD, 1965, "A simplex method for function minimization." *Computer Journal*, 7, 308-313.

NEVO, A., 2000, "A practioner's guide to estimation of random-coefficients logit models of demand." *Journal of Economics & Management Strategy*, 9, 513-548.

OSBORNE, M., 2007, "Consumer learning, switching costs, and heterogeneity: a structural examination." Mimeo.

PAGANO, M., 1989, "Trading volume and asset liquidity." *The Quarterly Journal of Economics*, 104, 255-274.

— and A.J. PADILLA, 2005, "The economics of cash trading: an overview." *LECG Report for Euronext.*

PAKES, A., 1986, "Patents as options: some estimates of the value of holding European patent stocks". *Econometrica*, 54, 755-784.

— and D. POLLARD, 1989, "Simulation and the asymptotics of optimization estimators". *Econometrica*, 57, 1027-1057.

PESENDORFER, M., 2002, "Retail sales: a study of pricing behavior in supermarkets". *Journal of Business*, 75(1). 33-66.

PESSEMIER, E. A., 1985, "Varied individual behavior: some theories, measurement methods and models." *Multivariate Behavioral Research*, 20, 69-94.

PIGOU, A. C., 1920, *The Economics of Welfare.* London: MacMillan.

PINKSE, J. and M. SLADE, 2004, "Mergers, brand competition, and the price of a pint." *European Economic Review*, 48: 617-643.

— and C. BRETT, 2002, "Spatial price competition: a semiparametric approach." *Econometrica*, 70: 1111-1153.

POLLACK, R., 1970, "Habit formation and dynamic demand functions." *Journal of Political Economy*, 78(4), 745-763.

POWELL, M. J. D., 1972, "Some properties of the variable metric algorithm." *AERE Harwell Report*, TP483.

—, 1971, "On the convergence of the variable metric algorithm." *J. Inst. Maths Applns.*, 7: 21-36.

POZZI, A., 2009, "Shopping cost and brand exploration in online grocery." Mimeo.

PRELEC, D. and G. LOEWENSTEIN, 1997, "Beyond time discounting." *Marketing Letters*, 8, 97-108.

PRESS, W.H., S.A. TEUKOLSKY, W.T. VETTERLING, and B.P. FLANNERY, 1994, *Numerical Recipes in C.* Cambridge, New York and Sidney: Cambridge University Press.

PUTERMAN, M. L. and M. C. SHIN, 1978, "Modified policy iteration algorithms for discounted Markov decision problems." *Management Science*, 24, 1127-1137.

RUST, J., 1987, "Optimal replacement of GMC bus engines: an empirical model of Harold Zurcher". *Econometrica*, 55(5). 999-1033.

RYSMAN, M., 2007, "An empirical analysis of payment card usage." *The Journal of Industrial Economics*, LV, 1-36.

SEETHARAMAN, P. B. and H. CHE, 2009, "Price competition in markets with consumer variety seeking." *Marketing Science*, 28(3), 516-525.

SHUM, M., 2004, "Does advertising overcome brand loyalty? Evidence from the breakfast-cereals market." *Journal of Economics and Management Strategy*, 13, 91-109.

SIMONSON, I. and R. S. WINER, 1992, "The influence of purchase quantity and display format on consumer preference for variety." *The Journal of Consumer Research*, 19(1), 133-138.

SOBEL, J., 1984, "The timing of sales." *Review of Economic Studies*, 53(3), 353-368.

SPINNEWYN, F., 1981, "Rational habit formation." *European Economic Review*, 15, 91-109.

STOLL, H., 2000, "Friction." Mimeo.

VARIAN, H., 1980, "A model of sales." *American Economic Review*, 70(4), 651-659.

VENKATESAN, M., 1973, "Cognitive consistency and novelty seeking." In S. Ward and T. S. Robertson (Eds.), *Consumer Behavior: Theoretical Sources*, Englewood Cliffs, NJ: Prentice Hall, 355-384.

WAHAL, J. P., 1997, "Entry, exit, market makers, and the bid-ask spread." *Review of Financial Studies*, 10, 871-901.

WERTENBROCH, K., 2001, "Self-rationing: self-control in consumer choice." Mimeo.

WESTON, J., 2002, "Electronic communication networks and liquidity on the Nasdaq." *Journal of Financial Services Research*, 22, 125-139.