Università degli Studi di Torino

Vilfredo Pareto Doctorate in Economics

Essays in Empirical Industrial Organization

Lorien Sabatino

2015-2018

- Why me?

- That is a very Earthling question to ask, Mr. Pilgrim. Why you? Why us for that matter? Why anything? Because this moment simply is. Have you ever seen bugs trapped in amber?

- Yes.

- Well, here we are, Mr. Pilgrim, trapped in the amber of this moment. There is no why.

Kurt Vonnegut Jr., Slaughterhouse-Five

Università degli Studi di Torino

Dipartimento di Scienze Economico-Sociali e Matematico-Statistiche

Vilfredo Pareto Doctorate in Economics

XXX Cycle

S.S.D.: SECS-P/01

Essays in Empirical Industrial Organization

Supervisor:

Mario Pagliero

Candidate:

Lorien Sabatino

Coordinator: Fabio Cesare Bagliano

2015-2018

Acknowledgements

I would like to thank the many people who have contributed in some way to my work over the last three years.
Firstly, I must give special thanks to my supervisor Mario Pagliero, who has provided me with a rigorous approach to my research, and whose continual guidance and patience has been greatly appreciated. Our numerous informal meetings have significantly shaped my research as well as my attitude towards applied research. I would like also to thank Christos Genakos for the hospitality during my Cambridge days, and the support during the last year of my PhD. His expertise and experience have proved invaluable. I am also indebted to Ignacio Monzon for the long talks throughout my PhD, and for the opportunity to work for him as a teaching assistant. Special thanks to Andrea Gallice for his patience and availability from the very first day of my PhD.
I must also offer my gratitude to my colleagues for the occasionally very welcome distractions from work. I am proud to have met so many special people that have influenced me so deeply. I am also very grateful to the Chief Economist Team at the European Commission for giving the opportunity to work with them. Such an experience has proved to be enormously valuable.
A special mention also goes to my family who has always supported me during the past three years. Thanks to my father Valerio for allowing me to pursue my studies and ambitions, and to Virgilio, Cinzia, and Dario for being always by mi side. I must also thank all my friends for their interest in what I do, and for their (long-distance) support. Without them I would not have gone that far...
Finally, I would like to thank the Collegio Carlo Alberto for providing support and funding for my PhD. I also thank the lady working at the self-service restaurant "Flash" in Moncalieri for feeding me during the toughest research periods. Her soups kept me alive and healthy.

Contents

List of Tables

List of Figures

Declaration of Co-Authorship

The second chapter of this thesis is a joint work with Christos Genakos (Cambridge Judge Business School, UAEB, CEP, and CEPR), Mario Pagliero (University of Turin, Collegio Carlo Alberto, and CEPR), and Tommaso Valletti (European Commission, Imperial College London, and University of Rome "Tor Vergata").

Introduction

This thesis collects three contributions in the field of Empirical Industrial Organization.

In the first chapter I investigate the effects on consumer welfare and industry profits induced by competition in proprietary aftermarket. Aftermarkets are those markets for goods that are complementary to a more complex durable product. When manufacturers effectively exclude independent sellers in providing complementary goods to the durable product, then they monopolise the *aftermarket*. I examine the effect of competition in aftermarkets on firm profits and consumer welfare. Proprietary aftermarket implies relatively high add-on prices and lower equipment prices as the manufacturer internalises the complementarity effects between equipment and add-ons. On the other hand, competition in the aftermarket implies lower add-on prices but higher equipment prices *ceteris paribus*. I analyse this trade-off by estimating profits and consumer welfare of the observed monopolised aftermarket against a counterfactual where the aftermarket is competitive in the Italian toothbrush industry. By applying a structural approach I find that, given the demand for differentiated systems in the Italian toothbrush market, aftermarket competition leads to both lower profits for the firms and lower consumer welfare, suggesting that monopolised aftermarkets are more efficient than competitive ones.

In the second chapter, I analyse the effect of the introduction of the Fixed Book Price regulation on prices and book variety in the Italian book market. The regulation implies that publishers set the price at which books are sold to final consumers. In this way governments aim to promote non-price retail competition, to foster the production of books, and to stimulate reading. However, FBP regulation relaxes booksellers price competition, thereby inducing higher prices and lower innovation at the retail level. This chapter provides the first systematic evidence of the impact of such regulation on variety and prices in the Italian book market. Results suggest that the regulation negatively affects consumers, as it raises prices without significantly affecting the number of new books in the marketplace.

The third chapter introduces discrete choice models and their applications to demand estimation. I analyse the main features of discrete choice models, with particular emphasis on the primitives and the analytical solutions of the equations to be estimated. In particular, I link these models to the standard microeconomic theory of consumer behaviour, I show how to analytically solve the random utility maximization problem, and I derive choice probabilities in a general setting. Finally, I concentrate to the analysis of Logit and Nested-Logit demand models.

CHAPTER 1

Measuring the Effects of Competition in Proprietary Aftermarket: A Counterfactual Analysis from the Italian Toothbrush Industry

1. Introduction

In many industries, firms manufacture complex goods for which consumers demand services, support, upgrades, or other complementary products, after the initial sale: standard examples include printers and toners, razors and blades, elevators and maintenance services. The common feature of these industries is that consumers invest in one durable equipment (in the *foremarket*), and then demand variable units of a perishable add-on after the initial purchase. When manufacturers effectively exclude independent sellers in providing complementary goods to the durable product, then they monopolise the *aftermarket*. This paper examines the effect of competition in aftermarkets on firm profits and consumer welfare. Proprietary aftermarket implies relatively high add-on prices and lower equipment prices as the manufacturer internalises the complementarity effects between equipment and add-ons. On the other hand, competition in the aftermarket implies lower add-on prices but higher equipment prices *ceteris paribus*. I analyse this trade-off by estimating profits and consumer welfare of the observed monopolised aftermarket against a counterfactual where the aftermarket is competitive in the Italian toothbrush industry.

In the toothbrush market, consumers can choose between two broad categories of toothbrushes, namely manual and electric toothbrushes. Electric toothbrushes (ET) are durable products that need refill heads upon consumption. Moreover, both ET and refill heads are tailored in such a way that consumers cannot use heads branded differently from the equipment, thus ET manufacturers can exert monopoly position in their proprietary aftermarket. Both manual toothbrushes (MT) and refill heads are disposable goods that need to be replaced from time to time. Therefore, consumers can choose either to purchase manual toothbrushes and replace them, or to make a high upfront payment for ET and then replace the heads.

My empirical strategy is to estimate the demand for toothbrushes, and then use the estimated cross-substitutions to derive price-cost margins (PCM) - without observing actual costs - under both proprietary and competitive aftermarket. By assuming that costs are equal in the two different market regimes, I can derive the price effect induced by aftermarket liberalisation. The first step of this strategy is to estimate the demand function, taking into account the complementarity between equipment and add-on as well as the differences in durability between MT and ET. By appropriately defining consumers' choice set I can estimate demand over differentiated *systems* using the standard *discrete choice* framework. In this setting, the demand is a function of prices, system's characteristics, parameters to be estimated, and heterogeneous consumers' preferences distributed according to the nested-logit model[1]. The estimated coefficients are then used to derive profits and consumer surplus under the observed proprietary aftermarket regime versus the competitive regime.

I model the supply side by developing a simplified version of Kende [1998], who analyses the conditions under which the profits of a system developer are greater under a *closed* system or an *open* system[2]. The model consists of two stages: in the first one the manufacturer sells the equipment, and consumers decide whether to buy equipment based on its price and the resulting second-stage aftermarket prices. In the second stage there are add-on sales, which depend on the amount of equipment bought in the first period (i.e. the *installed base*) and the actual prices of add-ons. My model departs from Kende [1998] as I do not allow for the entry of new products in the market, hence my counterfactual analysis is in the short-term. Given the supply model and the estimated demand, I compute PCM under different aftermarket regimes by changing products ownership as in Nevo [2001], and I assess profits and consumer welfare in two opposite settings. Firstly the observed one, in which systems are closed and the aftermarket is monopolised by equipment manufacturer. Secondly the competitive aftermarket, where add-ons are produced by independent firms competing in a Betrand-Nash game with differentiated products.

The results suggest that, when passing from closed to open systems, the foregone aftermarket profits almost offset the increased foremarket profits even when installed base customers are exactly equal to the number of equipment purchased in the same period. Hence, if one enlarges the installed base to equipment customers of previous periods, then manufacturer's profits decrease substantially. Moreover, also consumer

[1]Standard applications of nested-logit model include Berry [1994], Verboven [1996], and Goldberg [1995]. For an interesting comparison between nested-logit model and random-coefficients models see also Grigolon and Verboven [2014].

[2]A system is considered to be closed if the system manufacturer controls the production of compatible peripherals. The system is considered open if there is free competition in the production of peripherals.

welfare is negatively affected. This happens because, given the estimated parameters, although competition in the aftermarket slightly increases consumer surplus, the negative effect from double marginalisation arising when the aftermarket is a singleton is larger.

My predictions rely on the ability to consistently estimate demand in a market with differentiated products, some of which involving complementary add-ons. I use a two-year panel of quantities and prices for toothbrushes of the Italian market. For each administrative region I observe total quantity and average prices for toothbrushes sold through the mass retail channel distribution. The estimation faces two challenges related to the specification of the demand within the discrete choice framework in order to capture the complementarity between ET and refill heads, and the correlation of both foremarket and aftermarket prices with the econometric error term. I deal with the first issue by appropriately defining consumers choice set. In my setting, consumers can choose either a MT or one of the available *systems*, i.e. a combination of ET and refill head. Although this decision process is indeed dynamic, by keeping fixed the available alternatives, I can reduce the process to a static decision problem, and estimate a discrete-choice model (McFadden [1981], Berry [1994]).

The second challenge concerns the endogeneity of both aftermarket and foremarket prices in the demand equation. My identification strategy relies on the assumption that demand shocks are independent across regions once I control for brand, region, and time fixed effects. In this setting, product prices in other regional markets are valid instruments, and they are also correlated with local prices via the common marginal cost. Nevertheless, there are several possible reasons why exogeneity of the instruments might not hold. In fact, any demand shock affecting more than one regional market will spoil the validity of the instruments. For this purpose I divide the Italian market in four macro-areas, and I generate three instruments characterised by different degree of validity and relevance. The choice of the instrument will reflect their performance in the first-stage equation as well as the estimated coefficients in the second stage.

This paper contributes to two main strands of the economic literature. First, it contributes to the aftermarket literature which started with Shapiro [1995], Borenstein et al. [1994], Borenstein and Netz [1996], and which focuses on the assessment of consumer harm from aftermarket monopolization. Recent developments in this literature include Cabral [2014], Fong [2008], and Miao [2010]. The present work sheds new light on the effect of aftermarket monopolization on consumer welfare by empirically investigating this issue. Moreover, in the spirit of Carlton and Waldman [2009], I find that allowing for competition in the aftermarket might reduce total welfare.

Secondly, this paper contributes to the long tradition of discrete choice models (McFadden [1981], Berry [1994], Berry et al. [1995], Nevo [2001]), and in particular to discrete choice models with complementaries (Gentzkow [2007]) and interoperability (Genakos et al. [2011]). This paper contributes to this broad literature by constructing a simple static model for deriving cross-price elasticities between equipment and add-ons.

The remainder of the paper is organised as follows. Section 2 gives a brief description of the Italian toothbrush industry. In section 3, I outline the empirical model, for both demand and supply side. Section 4 presents data and descriptive statistics. In section 5 I explain the identification strategy. Section 6 shows the results from the econometric model and the implied counterfactuals. Section 7 concludes and outlines extensions.

2. The Toothbrush Industry and Competitive Landscape

Toothbrush industry is part of the wider Oral Care sector, which includes also toothpastes, dental flosses, denture care products, and mouthwashes/fresheners. As shown in Table 1, toothpaste is the more important segment in terms of sale value, followed by toothbrushes and mouthwashes. Although MTs display higher sale value, ETs show a greater growth rate. This happens because MTs are standard cleaning tools, and the segment is mature, while ETs are preferred by a growing number of consumers thanks to their perceived greater efficacy and the lower risk of damage to gums. The majority of brands in Oral Care in Italy belong to major multinational companies. In particular, toothbrush industry is fairly concentrated with the four major firms sharing almost 70% of the market (Table 2). Procter & Gamble is the

Table 1. Sale Values in the Oral Care

Segment	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
Dental Floss	43	42	41.5	41.7	42.6	43.4
Denture Care	133.7	129.7	127.2	125.9	123.6	122
Mouth Fresheners	2	2.1	2.1	2.2	2.3	2.3
Mouthwashes/Dental Rinses	176.2	179.9	187.4	189.1	198.8	206.3
Tooth Whiteners	22.7	21	19.8	19.3	18.5	17.9
Toothbrushes	279.5	280	281.6	285.6	285	285.7
- Manual	153.5	151.2	149.4	150.6	149.8	149.6
- Electric	126	128.8	132.1	135	135.2	136.1
Toothpaste	537.2	530.8	531.4	536.6	540.4	543
Total	1194.2	1185.4	1911.2	1200.4	1211.1	1220.5

Source: Euromonitor International. Data in EUR million.

leader in this industry in terms of revenue, and it exerts an almost monopoly position in the ET segment.

ETs are toothbrushes that make rapid and automatic bristle motions in order to clean the teeth. Bristles can oscillate, rotate, and pulse at different speed (frequency). The number of bristle movements (modes) together with the frequency of rotation/oscillation, vertically differentiate these products. They are relatively expensive, and consumers need to replace the head after some period of time. Moreover, ET and refill heads are tailored in such a way consumers cannot use heads from a different brand. Therefore, customers can choose either to purchase a MT or to make an upfront payment for an ET and then replace heads.

Table 2 shows market shares and revenue sales of the four major firms in the industry for 2015 and 2016. Although Unilever has the highest market share in volume sales, P&G is the leader in terms of total revenues. This is the case because it enjoys an almost monopoly position in the ET segment, with market share that lies between 62%-64%, thereby generating a similar amount or revenues as from MTs.[3] Each firm possesses a portfolio of differentiated products that allow to cover all market niches. Product differentiation is extremely important in this market, since potentially every consumer has her/his special oral-care needs. In this view, brand-proliferation is a natural outcome in this industry. For instance, out of 271 toothbrushes in the marketplace in 2016, 50 belong to P&G, 36 to CP+GABA, 20 are

[3]Notice that the table does not include revenues from refill heads. By adding revenues from refill heads, total revenues for P&G are between 36-38 million of EUR.

Table 2. Market Concentration in Toothbrush Industry

	2015		2016	
	Volume Share	**Revenue**	**Volume Share**	**Revenue**
Unilever				
- MT	24.5	18.5	25.5	18.6
- ET	-	-	-	-
- Total	24	18.5	24.9	18.6
P&G				
- MT	15.3	12	14.7	11.4
- ET	62.2	9.7	63.8	11.1
- Total	16.3	21.7	15.8	22.5
CP+GABA				
- MT	15.4	12	16.3	13.7
- ET	19.8	1	18.7	1
- Total	15.4	13	16.4	14.7
GSK				
- MT	12.7	8.4	12.2	7.8
- ET	-	-	-	-
- Total	12.4	8.4	11.9	7.8
C4				
- MT	67.9		68.7	
- ET	82		82.5	
- Total	68.1		69	

Source: Nielsen Italy. Revenues in EUR Million

owned by Unilever, and 9 by GSK. Hence almost half of the products in the marketplace are produced by one of the major four. Variety in the aftermarket reflects both foremarket volume sales, and P&G strong position in the ET segment. As of 2016, P&G owns 12 out of 29 available refill-heads.

In conclusion, the Italian toothbrush industry displays high concentration, especially in the ET sector. P&G should be considered the leader in this industry, as it is in the top three firms of the MT segment, while enjoying an almost monopoly position in the ET segment. On top of that, it exercises monopoly power in the proprietary aftermarket of refill heads.

3. Empirical Framework

My empirical strategy relates to Nevo [2001]. I model the demand side via discrete choice model, and I use the estimated cross-substitutions to predict the prices that would arise with competitive aftermarkets. In closed systems, equipment manufacturers own all brands in the relevant aftermarket, and they can set their prices simultaneously. When systems are open, add-ons are assumed to be produced by independent price-setting firms.

I model demand side using a three-level Nested-Logit model. At the first stage, consumers are assumed to choose whether they want to go for ET, MT, or none of the available alternatives. If they choose MT, they select one of the available alternatives in the MT *nest*. On the other hand, if they choose ET, they are assumed to firstly select an equipment in ET nest, and then one of the available add-ons in the relevant aftermarket.

My supply side follows Kende [1998], who studies the profitability of open versus closed systems. I model competition in this industry as a two-stage game. In the first stage foremarket competition takes place. Multi-product firms independently set foremarket prices, and customers purchase one of the available products also taking into account both foremarket and aftermarket prices. In the second stage, firms set aftermarket prices, and add-ons are sold. In proprietary aftermarket, equipment manufacturer set add-on prices simultaneously, behaving as a multi-product aftermarket monopolist. In the competitive regime, add-ons are assumed to be produced by independent firms which compete in a Bertrand-Nash game with differentiated products. I find the solution of the game under the two aftermarket regimes by backward induction.

3.1. Supply Side.

3.1.1. *Closed System.* **Second Stage**

Suppose there are $K > 1$ elements in some relevant aftermarket. The equipment manufacturer seeks to solve the following maximisation problem:

$$\max_{\pi_1,\ldots,\pi_K} I(p,\pi)\sum_{k=1}^{K}(\pi_k - c_k)\sigma_k(\pi) \tag{1}$$

where c_k and π_k are the constant marginal cost and price for add-on k respectively, $\sigma_k(\pi)$ is the market share of k in the relevant aftermarket such that $\frac{\partial\sigma_k}{\partial\pi_k} < 0$ and $\frac{\partial\sigma}{\partial\pi_l} \geq 0$ for $l \neq k$. Finally, $I(p,\pi)$ denotes the installed base customers, i.e. the size of the relevant aftermarket.[4] By taking the F.O.Cs for $l = 1, 2, .., K$ we get

$$I(p,\pi)\sigma_l(\pi) + \sum_{k=1}^{K}(\pi_k - c_k)[I(p,\pi)\frac{\partial \sigma_k}{\partial \pi_l} + \frac{\partial I}{\partial \pi_l}\sigma_k(\pi)] = 0$$

This system of K equations with K unknown takes the following matrix form:

$$I(p,\pi)\sigma(\pi) - \Lambda^M(\pi - c) = 0 \tag{2}$$

Where Λ^M is a $K \times K$ matrix whose elements are $\Lambda_{l,k} = -[I\frac{\partial \sigma_k}{\partial \pi_l} + \sigma_k \frac{\partial I}{\pi_l}]$, $I(p,\pi)$ is a scalar, and $\sigma(\pi)$ is $K \times 1$ vector. The solution for (2) is given by

$$(\pi - c) = I(p,\pi)\ [\Lambda^M]^{-1}\sigma(\pi) \tag{3}$$

which applies to all relevant aftermarkets.

First Stage

In a closed system equipment manufacturers get profits from both aftermarket and foremarket. Suppose there are J products in the foremarket: firm f produces a subset of foremarket alternatives $\mathcal{F}_f = \{\mathcal{F}_f^{MT}, \mathcal{F}_f^{ET}\}$, where $j \in \mathcal{F}_f^{ET}$ involve aftermarket components. Given relevant aftermarket and add-on prices, firm f's profits are

$$\sum_{j \in \mathcal{F}_f}(p_j - c_j)Ns_j(p,\pi) + I_f(p,\pi)\sum_{k=1}^{K_f}(\pi_k - c_k)\sigma_k(\pi) - F_f \tag{4}$$

where N denotes foremarket size, F_f is the fixed cost of production, p_j and c_j are foremarket price and the constant marginal cost of production respectively, $\mathcal{F}_f$ is the set of equipment produced by firm f, and $s_j(p,\pi)$ is product j's foremarket share, with the following properties: $\frac{\partial s_j}{\partial p_j} < 0$, $\frac{\partial s_j}{\partial p_h} \geq 0$, $\frac{\partial s_j}{\partial \pi_k} \leq 0$ for all k in the relevant aftermarket.

In toothbrush industry, all equipments produced by firm f are compatible with all the K_f add-ons produced by f, hence $I_f(p,\pi) = \sum_{j \in \mathcal{F}_f^{ET}} Ns_j(p,\pi)$. Each price p_j of any product j produced by firm f satisfies the following FOC:

$$s_h(p,\pi) + \sum_{j \in \mathcal{F}_f}(p_j - c_j)\frac{\partial s_j}{\partial p_h}(p,\pi) + V_f(\pi)\sum_{j \in \mathcal{F}_f^{ET}}\frac{\partial s_j}{\partial p_h}(p,\pi) = 0$$

[4] When the aftermarket is a singleton, i.e. $K = 1$, aftermarket demand is simply equal to $I(p,\pi)$.

where $V_f(\pi) \equiv \sum_{k=1}^{K_f}(\pi_k - c_k)\sigma_k(\pi)$. This system of J equations takes the following matrix form:

$$s(p, \pi) - \Omega(p - c) + V(\pi)\Omega\mathbb{I} = 0 \tag{5}$$

where $V(\pi) = [V_1, V_2, ..V_f, .., V_F]$, and $\mathbb{I}$ is a vector whose elements value 1 if $j \in \mathcal{F}_f^{ET}$ and zero otherwise. The entries of Ω matrix are:

$$\Omega_{h,j} = \begin{cases} -\dfrac{\partial s_j}{\partial p_h} & if \ \ \{j, h\} \in \mathcal{F}_f \\ 0 & otherwise. \end{cases}$$

The solution of (5) gives rise to the following price-cost margins:

$$(p - c) = \Omega^{-1}s(p, \pi) - V(\pi)\mathbb{I}. \tag{6}$$

3.1.2. *Open System.* **Second Stage**

Suppose there are $K \geq 1$ elements in some aftermarket, and that all products are provided by different price-setting firms. Moreover, suppose that equipment manufacturer does not compete in the relevant aftermarket. Firms face an individual demand $\sigma_k(\pi)$ and play a Betrand-Nash game with differentiated products.
Each Firm k sets price π_k so as to solve the following maximisation problem

$$\max_{\pi_k} \ (\pi_k - c_k)I(p, \pi)\sigma_k(\pi) \tag{7}$$

By F.O.C.

$$I(p, \pi)\sigma_k(\pi) + (\pi_k - c_k)[I(p, \pi)\frac{\partial \sigma_k}{\partial \pi_k} + \frac{\partial I}{\partial \pi_k}\sigma_k(\pi)] = 0$$

In matrix form the solution of this system of equations can be written as:

$$(\pi - c) = [\Lambda^C]^{-1}I(p, \pi)\sigma(\pi) \tag{8}$$

Where Λ^C is the diagonal matrix of Λ^M.

First Stage

When the aftermarket is competitive, equipment manufacturer does not get any profit from the aftermarket. Hence, firm f's profit- maximization problem reduces to

$$\max_{j \in \mathcal{F}_f} \sum_{j \in \mathcal{F}_f} (p_j - c_j)Ns_j(p, \pi) \tag{9}$$

Then, the solution for this system of J equations is given by

$$(p - c) = \Omega^{-1} s(p, \pi) \tag{10}$$

The difference between equations (6) and (10) reflects the complementarity effect between foremarket and aftermarket when the last one is monopolised. When add-ons are produced by independent firms, equipment manufacturer cannot control aftermarket price, hence it does not internalise the effect of aftermarket price on equipment sales. Therefore, competition in the aftermarket rises equipment prices by V_f, which in turns depends on the number of aftermarket alternatives, and add-ons PCM.

The other effect of aftermarket liberalisation is shown by equations (3) and (8). In proprietary aftermarket equipment manufacturer set add-on prices as a multi-product monopolist, thereby charging high prices, as it also internalises the substitution effects among add-ons. In competitive regime, single-product firms compete in the relevant aftermarket, driving prices down.[5]

3.2. Demand Side. The supply model in the previous section allows to predict price variations induced by competition in the aftermarket. However, it relies on the ability to consistently estimate cross-substitution patterns among products. For this purpose, I estimate the demand for this industry by using a discrete choice model (McFadden [1981]), which allows me to circumvent the *too-many-parameters* problem. In this context, a product consists of a bundle of characteristics, and consumers are assumed to choose the utility maximising alternative for a given vector of prices.

Suppose we observe $m = 1, 2, ..., M$ markets and $t = 0, 1, ..., T$ periods. In each market and at each point in time there are $N_{m,t}$ heterogeneous potential customers.

[5]When the aftermarket is a singleton *double marginalisation* arises in the competitive regime. In fact, on the one hand the unique add-on is set at monopoly price by the independent aftermarket seller, and on the other equipment manufacturer set prices according to (10).

Consumers choose among a set[6] of alternatives that includes *systems*[7], manual toothbrushes, and a composite outside option. The conditional indirect utility of consumer n from system j_k in market m at time t is

$$\tilde{v}_{n,j_k,m,t} = v_{j_k,m,t} + \epsilon_{n,j_k,m,t}$$

$$v_{j_k,m,t} = -\alpha(p_{j,m,t} + \frac{\delta}{1-\delta}(\mathbb{I}_{j\in ET}\pi_{k,m,t} + (1-\mathbb{I}_{j\in ET})p_{j,m,t}) + \beta x_{j_k} + \xi_{j_k,m,t} \qquad (11)$$

$$n = 1, ..., N_{m,t}, \qquad j_k = 0, 1, ..., [j_1, j_2, ..., j_{K_j,m,t}], .., J_{K_J,m,t}$$

where $p_{j,m,t}$ and $\pi_{k,m,t}$ are foremarket and aftermarket prices in market m at time t respectively, x_{j_k} is a row vector of product characteristics, $\xi_{j_k,m,t}$ is a mean-zero error term collecting unobservable (by the econometrician) product-specific shocks, $\mathbb{I}_{j\in ET}$ is and indicator function taking value one if j is an ET, and $\epsilon_{i,j_k,m,t}$ is the idiosyncratic random taste component of consumer's utility. Finally α, β, and δ are parameters to be estimated.

Specification (11) is consistent with a utility function that results from the present value of future instantaneous utilities.[8] When deciding whether to purchase a MT or an ET, the consumer effectively needs to choose between an option (ET) that involves an initial investment equal to p in the present and the add-on purchases for all subsequent periods, and an option (MT) that consists of a purchase/replacement process starting from the present period. Outside option represents the possibility to not choose any of the available products in the market. The utility from this outside option is normalised to $\tilde{v}_{n,0,m,t} = \epsilon_{n,0,m,t}$.

Consumers are assumed to choose the alternative that maximises their utility. Given a distribution for the stochastic term ϵ, demand for alternative j_k is equal to the share of consumers for which j_k yields the highest level of utility.[9] Since consumers heterogeneity is solely captured by the idiosyncratic term, product market shares depend on the distribution of ϵ, which I assume is distributed according to a three-level Nested-Logit model consistent with the ordered decision process represented in Figure 1. In particular, consumers are assumed to choose first whether they want to

[6]For a formal specification of the choice set, see section 7 of the Appendix.

[7]A system is defined by a combination of toothbrush and head. The set of compatible heads to some toothbrush j is the *relevant aftermarket* for j.

[8]See section 7 of the Appendix for the derivation of indirect utility in a deterministic framework.

[9]See McFadden [1981] and Train [2009] for a full treatment of *Random Utility Maximization* hypothesis.

go for a manual toothbrush, an electric one, or none of the available alternative. If they choose MT nest, then they need to pick one alternative over the manuals; if they choose ET, then they are assumed to select first one foremarket product, and then a compatible add-on in the relevant aftermarket. Nested-Logit model gives rise to

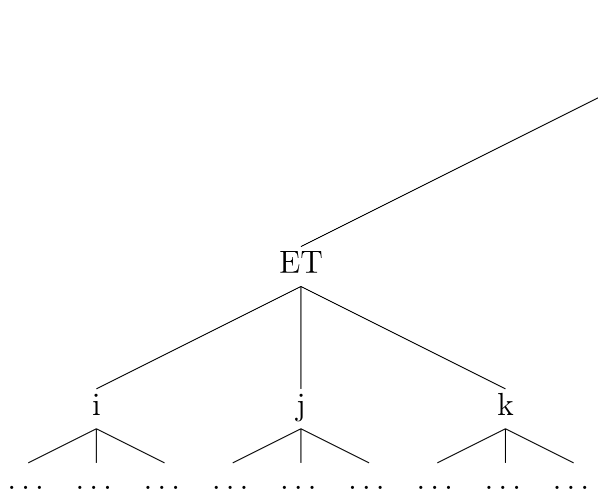


Figure 1. Consumer's Decision Process

closed-form market demand functions for *systems*[10] and it implies richer substitution patterns than standard Logit. Moreover the model can be linearised, allowing for the estimation of the parameters of interest via ordinary least-squares or two stages least-squares. The main equation to estimate is therefore given by

$$\ln(s_{j_k,m,t}) - \ln(s_{0,m,t}) = v_{j_k,m,t} + (1-\lambda_1)\ln(s_{j|\cdot T;m,t}) + (1-\lambda_2)\ln(s_{k|j;m,t}) \quad (12)$$

where $s_{k|j;m,t}$ is the aftermarket share of add-on k in the relevant aftermarket A_j in market m at time t, and $s_{j|\cdot T;m,t}$ is the foremarket share of product j in its nest, in market m and time t. So for instance, if j is a manual toothbrush, $s_{j|MT;m,t}$ will be the market share of alternative j within the nest MT. The *inclusive value* coefficients $\lambda_1, \lambda_2 \in (0,1)$ measure the degree of correlation between alternatives in the same nest and in the same relevant aftermarket respectively. As $\lambda_1 \to 0$ and $\lambda_2 \to 0$ we reach perfect correlation between alternatives in the same nest and sub-nest; conversely, as $\lambda_1 \to 1$ and $\lambda_2 \to 1$ alternatives become independent, and the model reduces to standard Logit.

Two final remarks are necessary. Firstly, my demand model predicts market shares

[10] See the Appendix.

over systems in the ET nest, i.e. a combination of toothbrush and refill-head. The demand of the specific toothbrush can be derived by summing up system market shares characterised by the same toothbrush over heads' shares. Secondly, in order to estimate (12) I need to derive system market shares from observed toothbrushes and refill heads market shares, since I do not observe the actual aftermarket choice for a given ET. This is particularly important when more than one equipment shares the same relevant aftermarket, as I only observe the marginal shares, while the joint market shares of equipment and add-on are unknown. My simplifying assumption is that the add-on choice in product j's relevant aftermarket is independent from foremarket choice. Under this condition, each system market share is given by the product between equipment foremarket share and add-on share in the relevant aftermarket.

4. Data

My dataset includes yearly product-specific total quantity of sales and revenues in the Italian toothbrush market for two years (2015-2016).[11] For each Italian region I observe volume sales and prices of toothbrushes sold through the mass retail channel - hypermarkets, supermarket and discount stores - for a total of 6,299 observations. I restrict my analysis to toothbrushes for adults. Each observation is a product sold in one region and one year ('15 or '16). The data also contains product's manufacturer, brand, and sub-brand. Moreover, I observe whether the product refers to a manual toothbrush (MT), an electric toothbrush (ET), or a pack of refill heads (RH).

In the dataset, ETs are divided in two broad categories, depending on whether they are powered by disposable batteries (BATT) or rechargeable ones (RECH). The former are identified by the string "BATT": these are low quality ETs with poor functionalities and low prices, whereas "RECH" refers to the most complex and expensive ones powered by rechargeable batteries. They both are durable goods that need refill heads from the proprietary aftermarket upon consumption. Price variation in ET segment reflects huge quality heterogeneity across products: an average BATT toothbrush costs 7 EUR, while average price for RECH is almost

[11] The data are provided by Nielsen Italy.

Table 3. Summary Statistics

	Variable	Mean	Std. Dev.	Min.	Max.	N
MT	Volume	14234.19	57375.21	11	1611513	
	Value	26878.46	88094.51	0	2148331	
	Price	2.28	1.227	0.01	16.3	4739
BATT	Volume	1553.25	4506.23	11	46816	
	Value	9819.26	27291.85	35	281840	
	Price	8.333	5.85	2.68	60	337
RECH	Volume	1288.49	4408.84	11	57558	
	Value	29437.59	85302.57	161	1086678	
	Price	38.26	23.78	9.5	153.39	582
RH	Volume	5365.05	13717.2	11	142066	
	Value	49801.34	124449.95	6	1223248	
	Price	8.11	3.35	0.49	16.8	641

Source: Nielsen Italy.

35 EUR. Descriptive statistics on prices and volumes are summarised in table 3. Manuals are the best selling toothbrushes, both in terms of number of products in the marketplace and volumes. Although RECH segment is the smallest one in terms of sales, it displays the highest growth from 2015 to 2016: total sales grow by 14.5% in RECH, by 1% in MAN, and decrease by 2% for BATT, suggesting that consumers are changing from standard to more complex toothbrushes.

Table 4. ET Products Characteristics

	Variable	Mean	Std. Dev.	Min.	Max.	N
BATT	Mode	1.05	0.47	1	3	337
	Rotation Frequency	9742.42	11808.40	0	30000	132
	Pulsation Frequency	1351.35	6231.45	0	30000	333
	Pressure Sensor	0	0	0	0	337
	Extra Hardware	0	0	0	0	337
RECH	Mode	1.43	1.14	1	6	582
	Oration Frequency	8199.28	3409.23	0	27000	558
	Pulsation Frequency	20587.63	10262.56	0	40000	582
	Pressure Sensor	0.84	0.37	0	1	582
	Extra Hardware	0.18	0.38	0	1	582

Source: Nielsen Italy.

I merge my data with products' characteristics taken from IceCat[12] and producer websites. ET have high heterogeneity in terms of functionalities, head's movement

[12]IceCat is an open catalogue with several information on product characteristics in the marketplace.

speed, and extra hardware in the pack. I include the following ET characteristics in my regression: number of modes[13], frequency of rotation/pulsation/oscillations, whether or not ET has pressure sensor[14], whether or not the pack includes extra hardware. Table 4 displays summary statistics for these product characteristics.

Table 5. REF Products Characteristics

Variable	Mean	Std. Dev.	Min.	Max.
rotation	0.772	0.42	0	1
white	0.206	0.405	0	1
doublehead	0.201	0.401	0	1
sonic	0.079	0.27	0	1
sensitive	0.084	0.278	0	1

Source: IceCat and Producer websites.

I generate five categories representing attributes for replacement heads that capture physical characteristics of the head (double vs single head) as well as their key features. Summary statistics are provided in Table 5. I recover attributes for manual toothbrushes from keywords in their sub-brand. Manufacturers emphasise products' key strength in the package through product name (sub-brand).[15] In this way I generate 13 categories that differentiate manual toothbrushes, which are summarised in Table 6.[16] ET customers need to replace the head of their toothbrush after some period of time. ETs and heads are manufactured in such a way consumers can only use refill heads of the same brand: for instance, they cannot use Phillips refill heads on an Oral-B toothbrush. This means that there is full interoperability *within* brand, but no inter-operability *across* brands, hence systems are closed. Since I observe both product's brand and whether it is a toothbrush or a refill head pack, I am able to link each equipment (ET) to its *relevant aftermarket*, defined as the set or replacement heads compatible with the ET. In this way I can match equipment and add-ons, and generate systems characterised by both foremarket, aftermarket prices,

[13]Modes refer to the different programmes available on the ET to brush teeth. Advanced toothbrushes have several modes, such as day/night cleaning, sensitive cleaning, etc.

[14]Pressure sensor signals when the consumer is pushing the toothbrush on the teeth with too much force.

[15]For instance a toothbrush called "Max White" emphasizes the whitening effect of the toothbrush: a toothbrush called "Sensitive Clean" is meant to suite sensitive teeth and gums.

[16]Ortodontico refers to special toothbrushes for people using brace; flex refers to toothbrushes with flexible heads.

Table 6. MAN Products Characteristics

Variable	Mean	Std. Dev.	Min.	Max.
whitening	0.086	0.281	0	1
sensitive	0.061	0.239	0	1
complete	0.035	0.183	0	1
ortodontico	0.028	0.166	0	1
deep clean	0.045	0.208	0	1
caries protection	0.011	0.105	0	1
plak protection	0.017	0.128	0	1
flex	0.019	0.136	0	1
classic	0.033	0.178	0	1
gum	0.012	0.108	0	1
tongue cleaning	0.037	0.189	0	1

Source: Producer websites.

and attributes. Hence, each observation is a combination of ET and one refill head in its relevant aftermarket. The size of my dataset expands to 8,664 observations.

5. Identification

5.1. Endogeneity. Identification of the econometric specification in (12) requires orthogonality of observable attributes with the error term, i.e. $E(x_{j_k}\xi_{j_k}) = 0$ across systems, region and time.[17] However, the other covariates in (12) are correlated with ξ, which includes all unobservable (to the econometrician) factors affecting system demands. Prices are endogenous since both consumers and producers internalise demand shocks when making their purchasing/pricing decisions. Neglecting this issue biases price parameters towards zero, thereby underestimating price elasticities. The logarithms of intra-nest market shares are specific covariates coming from the Nested-Logit model. Since they represent within-nests consumers choice, they are also correlated with the error term.

My identification strategy relies on the assumption that prices in other regions do not affect local demand once I control for brand, region and time fixed effects. The first control for all unobservable factors affecting products of the same brand in all regions and time - e.g. national brand advertising[18]. Regional fixed effects control

[17]This is a fair assumption in short-run perspective where firms cannot change product characteristics in response to demand shocks.

[18]P&G is famous for its aggressive ad campaign via national television, radio and web channels. Its activity is concentrated only in advertising electric toothbrushes, and this explains its strong position in the segment.

for unobservable shocks affecting products in the same regional market, while time fixed-effects control for time-specific national shocks. Accordingly, I can use product prices in other regions as instruments for local prices, since prices are correlated by the common marginal cost of production.[19]

Since I am particularly worried about the validity of product prices in other regions[20], I derive three different instruments exploiting the rich cross-section of my data and I analyse their performance in the estimation. The main idea is that, for a given regional market, prices from neighbouring regions are more relevant but less valid, since demand shocks in neighbouring regions are more likely to be correlated. Vice versa, prices from far regions are more valid but less relevant, since marginal costs are less correlated.

Let us focus on the identification of α. Consider foremarket price for product j, in market m at time t, $p_{j,m,t}$. A relevant instrument for this covariate is the average price of product j across the other markets. Let $\mathbb{I}_{j,m,t}$ be an indicator function that takes value 1 if at least one unit of product j is sold in market m at time t. Then, the instrument z for p is defined by

$$z_{j,m,t} = \frac{1}{\sum_{m=1}^{20} \mathbb{I}_{j,m,t} - 1} \sum_{l \neq m} p_{j,l,t} \tag{13}$$

For the other two instruments I proceed as follows: firstly, I divide Italian market in four macro-areas; secondly, I construct two instruments, z_1 and z_2, from other markets' product prices belonging to different macro-areas. Italian territory is divided into 20 administrative regions, and for each of them I observe market-level data on toothbrushes for 2015 and 2016. So, first I aggregate regions in four macro-areas:

- **Area 1. North-West:** Piemonte, Valle d'Aosta, Liguria, Lombardia.
- **Area 2. North-East:** Emilia-Romagna, Friuli-Venezia Giulia, Trentino, Veneto.
- **Area 3. Center:** Lazio, Toscana, Umbria, Abruzzo, Molise.
- **Area 4. South-Islands**: Basilicata, Campania, Calabria, Puglia, Sicilia, Sardegna.

[19]This identification strategy goes back to Hausman [1996]. For more on price endogeneity see Villas-Boas and Winer [1999].

[20]See Bresnahan [1996] for a critical view on this class of instruments

I construct my two instruments as follows. Consider product price $p_{j,m,t}$ in market m. One instrument z_1 will be the average price for product j across other regions $l \neq m$ within the same macro-area. For instance, the instrument for the price of product j in Piemonte is given by applying (13) over Valle d'Aosta, Liguria, and Lombardia. Similarly, I construct z_2 as the average product price across regions not belonging to the same macro-area as j, which is equal to applying (13) over the regions outside j's macro-area. Intuitively, z_2 should be more valid than both z and z_1, since prices in far away regions should be less correlated with local demand. On the other hand, we should expect z_2 to be weaker than z and z_1.
In order to identify the correlation coefficients (λ_1, λ_2), I use Berry et al. [1995] optimal instruments. The main idea is that this class of instruments proxies the competitive pressure faced by each alternative j within its nest. I use (i) the number of within-nest alternatives , (ii) the number of products produced by rival firms within the nest, and (iii) the number of alternatives in j's relevant aftermarket. These are valid instruments under the assumption that the number of alternatives is exogenous to unobserved demand factors. This is a fair assumption in a short-term horizon in which manufacturers cannot adjust their product portfolio.

6. Results

6.1. Demand Estimation. Table 7 displays the results of regression (12). In columns (1) and (2) I report results from ordinary least squares regressions. The introduction of brand-fixed effects affects drastically aftermarket price coefficient, though price coefficient remains statistically not significant. Columns (3)-(5) use various sets of instrumental variables in two-stage least squares. The first set of results, included in column (3), uses average product prices over all regions as shown in Section 5. The effects of the instruments on the parameters are significant both economically and statistically. Price coefficients move in the expected direction from ordinary least squares one, since endogeneity biases the parameters towards zero. The instrumental variables appear to work well in dealing with endogeneity, as shown also by Table 12 in the Appendix. Finally, the implied discount factor δ is estimated to be around 0.82.

The coefficient estimates of the within-nest share logarithm are quite problematic. The implied λ_1 measures substitutability across alternatives. If substitution is greater within than among nests, then $\lambda_1 \in (0, 1]$, whereas, if substitution among nests exceeds substitution within nests, then $\lambda_1 > 1$. Ordinary least squares specifications give statistically significant results, though the estimated λ_1 is very low, suggesting that within-nest alternatives are poor substitutes. The coefficient changes drastically in the IV regressions, however it is not statistically significant in any specification.

Table 7. Baseline Results

	—— OLS ——		—————— IV ——————		
Variables	(1)	(2)	(3)	(4)	(5)
Price	-0.007	-0.005	-0.059**	-0.054**	-0.060**
	(0.006)	(0.004)	(0.026)	(0.024)	(0.026)
Aftermarket Price	-0.075	-0.138**	-0.267***	-0.257***	-0.274***
	(0.049)	(0.064)	(0.080)	(0.087)	(0.072)
Log of	0.925***	0.946***	0.284	0.294	0.284
Within-Nest Share	(0.021)	(0.015)	(0.318)	(0.327)	(0.320)
Log of	1.251***	1.124***	2.081***	2.129***	2.066***
Add-on Share	(0.114)	(0.086)	(0.285)	(0.286)	(0.274)
δ			0.819	0.826	0.820
λ_1	0,075	0,054			
λ_2^*			0.540	0.551	0.546
Price			prices	prices in	prices not
Instrument				macro-area	macro-area
Brand FE	No	Yes	Yes	Yes	Yes
Observations	8,712	8,712	8,712	8,712	8,712
R-squared	0.896	0.931	0.531	0.535	0.533

*** p<0.01, ** p<0.05, * p<0.1

Clustered robust standard errors in parentheses. The dependent variable is $\ln(s_{j_k,m,t}) - \ln(s_{0,m,t})$. All regressions include time and region dummy variables, and product characteristics from table. Prices denote the average prices across regions of the product. Optimal instruments are included in all IV specifications.

The coefficient of the logarithm of add-on share allows us to estimate λ_2 which measures the substitution among alternatives in the same relevant aftermarket. Although the estimated coefficient is always positive and statistically significant among all specifications, it is persistently higher than one, which implies a $\lambda_2 < 0$ not consistent with random utility maximisation[21]. However, the interpretation of the parameter is more subtle in my setting, since I am basically reducing a dynamic decision process to a static one. Add-on market shares reflect consumers' choices in the aftermarket, which are indeed subsequent to foremarket purchasing decisions. By assuming that consumers incorporate present aftermarket conditions in their purchasing decisions, this implies that the same add-on will be selected throughout ET lifetime[22] In this sense, the estimated coefficient $(1 - \lambda_2)$ can be written as

$$(1 - \lambda_2) = \frac{\delta}{1 - \delta}(1 - \lambda_2^*)$$

where λ_2^* represents the instantaneous correlation parameter determined by the estimated coefficient from the logarithm of add-on share together with the discount factor δ. The parameter is reported in columns (3)-(5) of Table 7, and it varies between $0.54 - 0.55$.

Columns (3)-(5) display 2SLS results with different combinations of product prices for both foremarket and aftermarket prices. Table 12 in the Appendix displays the results from first stage regressions. The instruments follow the expected pattern in all specifications. Prices from other regions are always positively correlated with local price, while the number of products in the same or sub-nest are negatively correlated with the logarithm of within-nest and add-on shares. Interestingly enough, specification (3) performs relatively better in terms of relevance of the price instruments. This suggests that regional differences are not particularly large, and by taking the average price over the $M - 1$ regions I'd better refine my instrument. Hence, (3) is my preferred specification.

One main problem from the results of Table 7 is that the coefficient of the logarithm of within-nest share is not statistically significant in the IV specifications, implying that we cannot reject the null hypothesis of λ_1 being equal to one, which in turn would imply perfect correlation between alternatives in the same nest. This is an unreasonable result given the rich variation across toothbrushes, both in ET and MT nest. Moreover, the unique δ for both MT and ET implies the same durability between refill heads and manual toothbrushes. Table 8 shows results of regression

[21] Intuitively, for $\lambda_2 < 0$ product demand is an increasing function of own price. See Train [2009] for more on this issue.

[22] This argument is consistent with the deterministic model described in the Appendix.

(12) when I allow for different discount factors for MT and ET. The increased flexibility of the model allows us to account for different durability or frequency of usage between manual toothbrushes and refill heads.[23]

Table 8. Results from a more Flexible Model

	OLS		IV		
	(1)	(2)	(3)	(4)	(5)
Price	-0.000	-0.001	-0.055***	-0.051***	-0.055***
	(0.004)	(0.003)	(0.012)	(0.012)	(0.012)
Aftermarket Price	0.148*	0.116*	-0.194*	-0.185*	-0.189*
MT	(0.081)	(0.070)	(0.103)	(0.103)	(0.099)
Aftermarket Price	-0.632***	-0.517***	-0.478***	-0.447***	-0.495***
ET	(0.159)	(0.107)	(0.164)	(0.167)	(0.146)
Log of	0.972***	0.984***	0.342**	0.352**	0.352**
Within-Nest Share	(0.010)	(0.013)	(0.143)	(0.164)	(0.140)
Log of	0.934***	0.904***	1.526***	1.623***	1.478***
Add-on Share	(0.065)	(0.120)	(0.473)	(0.472)	(0.428)
δ_{MT}			0.779	0.783	0.775
δ_{ET}			0.898	0.897	0.9
λ_1	0.028	0.016	0.658	0.648	0.648
λ_2^*			0.826	0.85	0.836
Price			prices	prices in	prices not
Instruments				macro-area	macro-area
Brand FE	No	Yes	Yes	Yes	Yes
Observations	8,712	8,712	8,712	8,712	8,712
R-squared	0.945	0.964	0.650	0.651	0.662

*** p<0.01, ** p<0.05, * p<0.1

Clustered robust standard errors in parentheses. The dependent variable is $\ln(s_{j_k,m,t}) - \ln(s_{0,m,t})$. All regressions include time and region dummy variables, and product characteristics from table. Prices denote the average prices across regions of the product. Optimal instruments are included in all IV specifications. λ_2^* computed for the ET nest only.

Results from Table 8 show a similar pattern in terms of sign and significance of foremarket price coefficient as in Table 7. However, now I estimate two aftermarket price coefficients which vary across nests. The coefficient related to MT aftermarket price should be considered an additive component to the overall price sensitivity for manual toothbrush, whereas aftermarket price coefficient for ET effectively represents the average effect on utility from add-on price variation. The implied discount factors and inclusive coefficients from IV regressions are readily interpretable. ET

[23]This comes at a cost, since I am increasing the number of endogenous variables of the model. However, the results are very much in line with the first stage of the previous estimation.

consumers are more patient than MT customers, and consistently purchase a durable product. The coefficients λ_1 and λ_2^* lie between zero and one and follow the correct order of magnitude: sub-nest alternatives are relatively more similar - i.e. they have a higher coefficient - than alternatives in the same upper-nest. Since also in this specification the average price over the $M-1$ regional market works relatively better in terms of relevance, I use estimated coefficients from column (3) to derive the counterfactuals.

6.2. Predicted Prices and Welfare Evaluation. Given the demand parameters estimated in previous sections, we can derive price variation induced by aftermarket liberalisation by making use of equations (3)-(6) and (8)-(10). Equations (3)-(6) represent aftermarket and foremarket equilibrium price conditions when the equipment manufacturer monopolises the aftermarket, while equations (8)-(10) show equilibrium price conditions when single-product independent firms compete in the aftermarket. Table 9 presents prices in both aftermarket regimes for ET and refill heads. In particular, closed system's prices are the ones observed in the data, while prices in open system are derived from the combination of my supply side and demand estimation. Foremarket prices are relatively lower in closed systems, since equipment manufacturers internalise the complementarity between foremarket and aftermarket. Thus, in the open system prices unambiguously increase. Aftermarket PCM variation depends on whether the set of available add-ons is a singleton or not. Whenever there is more than one add-on in the relevant aftermarket (i.e. for P&G, CHURCH&DWIGHT, and RANIR) add-on margins decrease when the aftermarket is competitive. Interestingly enough, predicted prices are relatively higher in the liberalised aftermarket whenever an independent firm acts as a monopolist. Hence, when not mitigated by competition, aftermarket prices increase due to double marginalisation.

The magnitude of the price changes depend on the estimated elasticities from the Nested-Logit model. When the aftermarket is not a singleton and the system is open, foremarket price increase is higher both in absolute and relative terms than add-on price variation. This suggests that aftermarket competition does not compensate foremarket price increase.[24]

Although aftermarket competition increases equipment prices, such rise does not compensate the foregone aftermarket profits. Table 10 displays variable profits in the ET segment for each equipment manufacturer. The profits are computed by using predicted margins and observed demand for both toothbrushes and refill heads. In closed system, manufacturer profits are generated from the sum of both foremarket and aftermarket profits, while in an open system only foremarket profits matter.[25] If we consider the aftermarket profits generated by a single refill head, than the total profit in a closed system are almost the same as the one generated in the open system. However, if we suppose that consumers can only purchase a pack containing two refill heads, then profits in closed system are larger than those in the open system. Moreover, closed system profits in Table 10 be interpreted as lower bound, since in the computation I am not accounting for (i) customers who bought ET in the past and were already in the aftermarket, and (ii) future foregone profits.

[24]However the comparison should take into account the perishable nature of refill heads: a Δ price variation of refill heads affects lifetime system cost by Δ times $t = 1, 2, 3, ...$ periods.

[25]Profits are computed using equations 4 and 9 and the predicted margins. Aftermarket size equals the number of compatible equipments purchased in the relevant period.

Table 9. Observed and Predicted Prices for ET

Manufacturer	Prices: Closed System	Open System	Δ	Δ %
Foremarket				
CHURCH&DWIGHT	8.79	9.60	0.80	9.2
CP+GABA	7.38	8.06	0.70	9.3
KRIS	9.00	9.64	0.64	7.1
P&G	33.85	34.54	0.71	2.0
PHILIPS	25.09	25.86	0.77	3.1
PRIVATE	7.22	7.90	0.64	9.5
RANIR	14.93	15.57	0.64	4.3
Aftermarket				
CHURCH&DWIGHT	2.60	2.46	-0.13	-5.1
CP+GABA	2.68	2.81	0.13	4.8
KRIS	1.99	2.12	0.13	6.5
P&G	5.22	5.11	-0.11	-2.1
PHILIPS	6.21	6.21	0.00	0.0
PRIVATE	2.15	2.27	0.12	5.7
RANIR	2.49	2.41	-0.08	-3.1

Presented are median unit prices by manufacturer for ET and refill-heads. Prices in Closed Systems are the one observed in the data. Prices in Open System are derived from predicted elasticities using specification (3) of Table 8.

Table 10. Profits in ET Segment

	Closed System: Foremarket	Aftermarket Single RH	Total: Single RH	Total: Pack of RH	Open System Foremarket
CHURCH&DWIGHT	274199.8	21770.5	295970.3	317740.8	296199.1
CP+GABA	340202.7	104928.1	445130.8	550058.9	445130.8
KRIS	6203.188	1980.72	8183.91	10164.63	8183.91
P&G	892253.8	260154.7	1152408.5	1412563.2	1153856
PHILIPS	1683.387	569.4425	2252.83	2822.27	2269.48
PRIVATE	44665.13	13256.01	57921.14	71177.15	58066.91
RANIR	13874.53	4209.72	18084.25	22293.97	18084.25

Presented are average total profits in the ET segment by manufacturer. Profits are expressed in EUR. Margins are derived from the combination of supply side and demand estimation. Aftermarket profits refer to the unit margin multiplied by the aftermarket size.

I compute consumer welfare using the main result from Small and Rosen [1981] and I apply it to my Nested-Logit model. Omitting the constant for simplicity and summing over all consumers in each regional market, this is

$$CS = N \ln[\sum_{B}(\sum_{j \in B_k}(\sum_{k \in A_j} e^{\frac{v_{j_k}}{\lambda_2}})^{\frac{\lambda_1}{\lambda_2}})^{\lambda_1}], \tag{14}$$

where $B = \{ET, MT, 0\}$ collects all the nests, B_k is one of the nests in B, and A_j is the relevant aftermarket for some equipment j. In a closed system, equipment prices are relatively lower than in an open one, while add-on prices are relatively higher whenever the relevant aftermarket is not a singleton. Results from Table 11 show that consumer welfare decreases when the aftermarket is competitive. However, it seems that this is mainly due to the double marginalisation arising when the aftermarket is a singleton. In fact, when I compute the welfare change only for P&G customers, aftermarket competition leads to a slight increase in consumer welfare.

Table 11. Consumer Welfare Change in the ET Segment

	TOTAL			P&G		
Year	Closed System	Open System	Δ%	Closed System	Open System	Δ%
2015	29026338	27360557	-5.74	10315673	10457259	1.37
2016	31114009	29409251	-5.48	12372447	12513363	1.14

Total consumer welfare variation in the ET segment. Welfare is expressed in EUR, and it is derived from estimated parameters of the demand model.

7. Conclusions

In this paper, I use a Nested-Logit model in order to estimate the product demand for toothbrushes. The existence of electric toothbrushes that need refill heads makes the estimation more challenging, as a subset of products in the marketplace are complements rather than substitutes. I provide a new interpretation of negative value coefficient when it captures dynamic repeated choices. I use the estimated elasticities to derive counterfactuals over two opposite aftermarket regimes, given a two-stages competition model. The results suggest that allowing for competition in proprietary aftermarket reduces both consumer welfare and firms' profits. However, although profits unambiguously decrease, consumer welfare effects are heterogeneous, depending on whether the aftermarket is a singleton or not.

The results in this paper highlight the welfare enhancing effects of keeping equipments and add-ons in a closed system fully controlled by one manufacturer. In fact, on the one hand consumer welfare is only marginally positively affected, while on the other hand firms profits loss is significant. The profits in open system might be so low that firms might find it profitable to reduce variety in the foremarket, thereby reducing even more consumer welfare.

The paper does not address these long-term effects, which are nevertheless extremely important for a complete welfare analysis. The short-term horizon of the analysis presented in this paper is reflected in the demand model, in the identification strategy, and in the counterfactuals. On the demand side, consumers are assumed to observe present prices and alternative and choose the system that maximises their utility. Hence no dynamics in the choice of future add-ons is contemplated. Identification relies on the assumption that the number of alternative is exogenous to unobserved demand factors. This assumption works only within a short-term horizon, where manufacturers cannot adjust their product portfolio according to new demand shocks in the market-place. Similarly for the counterfactuals, no entry or exit is allowed, although one might well envision the possibility of new entrants in a liberalised aftermarket. All these crucial dynamic aspects are left for future research.

Appendix

Choice set and Deterministic Model. Define by F the set of alternatives in the foremarket, and let A the set of all add-ons. Both sets are discrete and finite. F includes an "outside" composite good indexed by 0.

For every $j \in F$ there exists a vector $x_j^F \geq 0$ of observable (to the researcher) attributes for alternative j such that $x_0 = 0$. Similarly, for every $k \in A$ there is a vector of observable characteristics x_k^A. Finally, let $p \gg 0$ be the (foremarket) price for alternatives in F, and $\pi \gg 0$ the price of add-ons in A.

Some of the alternatives in F need aftermarket components upon consumption, while other alternatives do not them. Hence we can partition F as $\{F_D, F_{ND}, F_0\}$ where:

- F_D collects all *durable* alternatives in F that support add-ons;
- F_{ND} includes all *non-durable* alternatives in F that cannot support any add-on;
- $F_0 = \{0\}$ includes the outside option.

For each $j \in F_D$ there exists a set of compatible add-ons $A_j \in A$. This means that there exists a collection of subsets $\tilde{A} = \{A_0, A_1, A_2, ..., A_j, ...\}$ including all *relevant aftermarket* sets, each one with $K_j \geq 0$ elements. These subsets have the following *compatibility* property:

$$\begin{cases} A_j \neq \emptyset \Leftrightarrow j \in F_D \\ A_j = \emptyset \;\; otherwise \end{cases} \tag{15}$$

A *system* is defined as the combination between foremarket alternative j and one add-on in its relevant aftermarket $k \in A_j$. So, for each j there are K_j systems.

Consumer choice set includes all systems available in the market. That is, choice set C is defined as

$$C \equiv F \times \tilde{A} \tag{16}$$

An element of the choice set can be either a combination of durable and add-on (i.e. a *system*), a non-durable alternative, or the outside option:

$$C = \{0, 1, 2, ..., [j_1, j_2, ..., j_{K_j}], ..., [J_1, J_2, ..., J_{K_J}]\}$$

It is useful to group the alternatives in C is different subsets. Define by $C_{ET} \subset C$ the set containing all alternatives with add-on, by C_0 the singleton outside option, and let $C_{MT} = C \setminus (C_{ET} \cup C_0)$ collect all manual toothbrushes. In this way we get a partition for $C = \{C_0, C_{ET}, C_{MT}\}$. Moreover we can group elements in C_{ET} characterised by the same foremarket alternative, generating a further partition in $C_{ET} = \{1, 2, ..., j, ..., J\}$.
The vector of observable characteristics for alternatives in C includes both foremarket and aftermarket attributes coming from the combination of equipment and add-on. Define by $x'_{j_k} = [x_j^F, x_k^A]$ such vector with the following properties:

$$x'_{j_k} = [x_j^F, x_k^A] \;\; j_k \in C_{ET}$$

$$x'_j = [x_j^F, 0] \;\; j \in C_{MT}$$

$$x_0 = [0, ..., 0]$$

$$X = [0, x_1, x_2, ...]$$

where X is the matrix collecting attributes vectors for alternatives in the choice set. I assume the individual has rational preferences over X, so that there exists a function $u : X \to \mathbb{R}$ representing those preferences.

Representative Utility Function. Assume $u(.)$ is additive in the argument, that is $u(x_j^F, x_k^A) = \tilde{u}(x_j^F) + \tilde{u}(x_k^A)$, and $\tilde{u}(0) = 0$. Consider $t = 0, 1, 2, ...$ periods: at $t = 0$ the consumer chooses one alternative in F and sustains an upfront payment equal to p_j. If the consumer chooses $j \in F_D$ then, at any $t > 0$, she needs to select one alternative in j's relevant aftermarket A_j. On the other hand, if $j \in F_{ND}$ is selected, then the consumer needs to choose again one perishable alternative in F_{ND}. In this simple framework, although first-stage decision is important as it implies different choice-sets in subsequent stages, these decisions are independent from choice in the first stage. Moreover, if a perishable foremarket good is chosen in the first stage, then it will be selected again in all subsequent periods.[26] The consumer is assumed to choose the system that maximizes the discounted sum of

[26]We can prove this statement by contradiction for $t = 0, 1$ and $\delta = 1$ without loss of generality. Suppose $\tilde{u}(x_j^F) - p_j + \tilde{u}(x_k^F) - p_k$ is the couple that yield the highest level of utility, and that $j, k \in F_{ND}$. Then, either $\tilde{u}(x_j^F) - p_j \geq \tilde{u}(x_k^F) - p_k$ or $\tilde{u}(x_j^F) - p_j \leq \tilde{u}(x_k^F) - p_k$. Suppose the first condition holds true; then $2(\tilde{u}(x_j^F) - p_j) \geq \tilde{u}(x_j^F) - p_j + \tilde{u}(x_k^F) - p_k$.

(indirect) utility. That is, the consumer chooses $j_k \in C$ that maximizes her indirect utility function given by:

$$V = \begin{cases} \sum_{t=0}^{\infty} \delta^t \tilde{u}(x_j^F) - p_j + \sum_{t=1}^{\infty} \delta^t (\tilde{u}(x_k^A) - \pi_k), & j_k \in C_{ET} \\ \sum_{t=0}^{\infty} \delta^t (\tilde{u}(x_j^F, 0) - p_j), & j_k \in C_{MT} \\ 0, & j_k = 0 \end{cases} \tag{17}$$

where $\delta \in (0,1)$ is consumer's discount factor. According to specification (17), indirect utility from ET is made by two components: first, a discounted stream of utility coming from foremarket durable alternative minus the upfront payment given by p. In addition she gets a discounted stream of utility - net to expenditure at every period - from $t = 1$. On the other hand, by choosing an alternative in MT, consumer simply gets the discounted sum of utility net of the expenditure at each point in time.

By noticing that $\sum_{t=1}^{\infty} \delta^t = \dfrac{\delta}{1-\delta}$ we can rewrite utility maximization problem as follows:

$$\max_{j_k \in C} \; V = \begin{cases} u(x_j^F, x_k^A) - (p_j + \dfrac{\delta}{1-\delta}\pi_k) & j_k \in C_{ET} \\ u(x_j^F, 0) - (p_j + \dfrac{\delta}{1-\delta} p_j), & j_k \in C_{MT} \\ 0, & j_k = 0 \end{cases} \tag{18}$$

Where $u(x_j^F, x_k^A) = \dfrac{1}{1-\delta}\tilde{u}(x_j^F) + \dfrac{\delta}{1-\delta}\tilde{u}(x_k^A)$ and $u(x_j^F, 0) = \dfrac{1}{1-\delta}\tilde{u}(x_j^F)$ are the lifetime utilities from system j_k conditional on being electric or manual respectively. The solution of (18) determines consumer's choice given the available alternatives, their prices and characteristics. Notice that the solution of (18) can be found recursively, starting from the utility maximizing solution in each subset and then comparing them in order to get the solution to the whole problem.

Deriving Nested-Logit Demand. Consider the following setting. There are $m = 1, 2, ..., M$ markets and $t = 0, 1, ..., T$, periods. In each market and at each point in time there are N_{mt} heterogeneous potential customers who need to choose one option over the set of feasible alternatives C_{mt}.

I capture consumers heterogeneity through an additive random component in the indirect utility. For this purpose, define with $\tilde{V}_{nmt}$ the vector of (true) utility levels for individual n over alternatives in C_{mt}, and assume that observable product attributes and prices enter the decision process of individuals as in (17). Then unobservable taste variation ϵ_{nmt} captures the difference between $\tilde{V}_{nmt}$ and *representative* utility V_{mt} :

$$\epsilon_{nmt} \equiv \tilde{V}_{nmt} - V_{mt} \implies \tilde{V}_{nmt} = V_{mt} + \epsilon_{nmt}$$

Given (17) and $\epsilon_{n,m,t}$, the share of consumers choosing $j_k \in C_{mt}$ is exactly equal to the share of consumers for whom alternative j_k yields the highest level of utility, given products characteristics and prices. That is, market share $s_{j_k,mt}$ is equal to

$$s_{j_k,m,t} = \frac{1}{N_{m,t}} \sum_n \mathbb{I}(\tilde{V}_{j_k,nmt} \geq \tilde{V}_{h_i,nmt},\ \forall h_i \in C_{mt}) =$$

$$= \frac{1}{N_{mt}} \sum_{n=1}^{N} \mathbb{I}(V_{j_k,mt} + \epsilon_{j_k,nmt} \geq V_{h_i,mt} + \epsilon_{h_i,nmt},\ \forall h_i \in C_{mt})$$

where $\mathbb{I}(.)$ is an indicator function that takes value one if argument holds.

Given a distribution for ϵ, the predicted market share $\hat{s}_{j_k,mt}$ is equal to the probability that alternative j_k yields the highest level of utility; that is[27]

$$\hat{s}_{j_k,mt} = \Pr\{V_{j_k,mt} + \epsilon_{j_k,mt} \geq V_{h_i,mt} + \epsilon_{h_i,mt},\ \forall h_i \in C_{mt}\} =$$

$$= \Pr\{\epsilon_{h_i,mt} \leq V_{j_k,mt} - V_{h_i,mt} + \epsilon_{j,mt},\ \forall h_i \in C_{mt}\} \tag{19}$$

The random component ϵ is distributed according to a three-level Nested-logit model consistent with the ordered decision process represented in Figure 1. This implies that the predicted market share for product $j_k \in C_{ET} \subset C$ is given by:

$$\hat{s}_{j_k;m,t} = \hat{s}_{ET;m,t} \times \hat{s}_{j|ET;m,t} \times \hat{s}_{k|j;m,t} =$$

$$= \frac{e^{\lambda_1 I_{l,mt}}}{1 + e^{\lambda_1 I_{ET,mt}} + e^{\lambda_1 I_{MT,mt}}} \times \frac{e^{\frac{\lambda_2}{\lambda_1} I_{j,mt}}}{\sum_{i \in C_{l,mt}} e^{\frac{\lambda_2}{\lambda_1} I_{i,mt}}} \times \frac{e^{\frac{1}{\lambda_2} V_{j_k,mt}}}{\sum_{k \in A_{j,mt}} e^{\frac{1}{\lambda_2} V_{j_k,mt}}} \tag{20}$$

where

$$I_{ET,mt} = \ln(\sum_{i \in C_{ET,mt}} e^{\frac{\lambda_2}{\lambda_1} I_{i,mt}}),\quad I_{MT,mt} = \ln(\sum_{i \in C_{MT,mt}} e^{\frac{\lambda_2}{\lambda_1} I_{i,mt}}),$$

[27]Let $F_\epsilon(t)$ be the c.d.f. for ϵ. If that distribution admits a p.d.f. $f\epsilon(t)$, then choice probability is given by:

$$\hat{s}_{j_k} = \int_{\epsilon_{j_k}} \frac{\partial F_\epsilon}{\partial \epsilon_{j_k}} (V_j - V_k + \epsilon_{j_k})$$

$$I_{j,mt} = \ln(\sum_{h \in A_{j,mt}} e^{\frac{1}{\lambda_2} V_{j_h,mt}})$$

are the *inclusive values* for nest C_{ET}, C_{MT}, and A_j respectively: they are the expected surplus consumer gets from the selected nest. So for instance, I_{MT} is the expected surplus from manual toothbrushes nest, and it is a function of all prices and product characteristics of alternatives in C_{MT}. Similarly, I_j is the expected surplus from product j relevant aftermarket, which depends on prices and add-ons characteristics in A_j. An increase in add-on price affects negatively I_j, making product j less desirable.

Equation (20) specifies the demand function for system j_k. Notice however that we are interested in the effect of add-on price variation on foremarket sales. So foremarket demand for alternative j is derived by summing up equation (20) over the available alternative in the relevant aftermarket $A_{j,mt}$:

$$s_{j,mt}(p, \pi|x) = \sum_{k \in A_{j,mt}} s_{j_k,mt}(p, \pi|x) \tag{21}$$

and we are interested in analysing the effects of add-on price variations on $s_{j,mt}(p, \pi|x)$. It worth noting that a price variation in add-on $k \in A_j$ affects s_j via the inclusive value I_j that determines the expected surplus from the equipment j's aftermarket.

Table 12. First Stage of Table 7

Instrumental Variables	Overall (3)		In Macro-Area (4)		Out Macro-Area (5)	
Price	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.	Coef.	Robust Std. Err.
Formarket Price	0.977	0.001	0.938	0.001	0.902	0.002
Aftermarket Price	0.023	0.005	0.010	0.016	0.101	0.008
N products Manufacturer	0.013	0.002	0.007	0.001	0.013	0.002
N products Rivals	0.019	0.005	0.021	0.005	0.015	0.005
N add-ons	0.004	0.064	0.272	0.052	-0.095	0.054
Partial R^2	0.797		0.817		0.749	
F-test	4.4e+06		8.5e+05		9.2e+05	
Aftermarket Price						
Formarket Price	0.000	0.000	0.000	0.000	0.000	0.000
Aftermarket Price	0.972	0.011	0.909	0.034	0.950	0.019
N products Manufacturer	-0.003	0.001	-0.003	0.001	-0.003	0.001
N products Rivals	0.008	0.003	0.010	0.004	0.008	0.003
N add-ons	0.159	0.026	0.207	0.043	0.161	0.027
Partial R^2	0.736		0.721		0.714	
F-test	15606		2741		23546	
Log of Within-Nest Share						
Formarket Price	-0.088	0.004	-0.079	0.003	-0.082	0.004
Aftermarket Price	-0.304	0.121	-0.286	0.105	-0.303	0.116
N products Manufacturer	-0.011	0.001	-0.010	0.001	-0.011	0.001
N products Rivals	-0.008	0.007	-0.008	0.007	-0.007	0.007
N add-ons	0.221	0.079	0.171	0.074	0.231	0.080
Partial R^2	0.158		0.143		0.152	
F-test	6297		4988		5726	
Log of Addon-Share						
Formarket Price	0.000	0.001	0.000	0.001	-0.001	0.001
Aftermarket Price	0.001	0.030	-0.005	0.028	0.010	0.034
N products Manufacturer	0.002	0.001	0.002	0.001	0.002	0.001
N products Rivals	-0.004	0.002	-0.004	0.002	-0.004	0.002
N add-ons	-0.367	0.018	-0.366	0.017	-0.370	0.019
Partial R^2	0.145		0.145		0.145	
F-test	15173		39022		9018	

Standard errors clustered on brand. Regressions based on 8712 observations All regressions include time and region dummy variables, and product characteristics from table. Each column refers to first-stage results using different combination of product prices.

CHAPTER 2

The Impact of Fixed Book Price Regulation on Prices and Variety: Evidence from the Italian Book Market

1. Introduction

Many European countries apply the Fixed Book Price (FBP) regulation. This regulation implies that publishers set the price at which books are sold to final consumers. In this way, governments aim to promote non-price retail competition to foster the production of books and to stimulate reading. Since books are the main vehicle of knowledge diffusion, this would create positive spillovers for the society as a whole. However, it is not clear a priori that by restricting retail price competition, the regulation promotes reading or innovation at the retail level or leads to the production of a larger variety of books.

Despite the considerable political debate across Europe, this is the first paper to provide systematic evidence of the impact of such regulation on book variety and prices. For this purpose we use a unique dataset of all the books published in Italy and their monthly sales from 2008 until 2014. Since September 2011 the *Levi law* regulates the Italian book market. The regulation imposes a 15% maximum discount booksellers can deduct from publishers prices. Hence, the regulation imposes a *Resale Price Maintenance* (RPM) by law. This paper empirically investigates the effects of the introduction of FBP in the Italian book market.

We use the introduction of regulation as an exogenous event that affected both the upstream publishing decisions and retail prices. Our analysis proceeds in two steps. First, we analyse the causal effect of the book price regulation on the variety of books in the Italian market. Our measure of book variety consists of the number of *new books* published at each point in time. New books are defined as all the book titles that have been published for the first time without taking into consideration re-editions. The idea is that by focusing on first editions only, we are able to effectively measure the change in the number of publications that does affect cultural variety in the marketplace. To consistently estimate the causal effect of the regulation on variety, we utilize a difference-in-differences (DiD)[1] framework. By analysing the discount policies before the introduction of regulation, we identify classes of books for which the regulation is not binding and use them as control group. Our results

[1]Early applications of this approach are found in Ashenfelter and Card [1985], Card [1992], Card and Krueger [1994], and Card and Krueger [2000]; more recent applications in industrial economics include, for example, Ashenfelter et al. [2013].

indicate the introduction of the regulation had no significant effect on the number of books published.

Second, we estimate the demand for books in order to capture the effect of the regulation on consumer welfare. We follow the literature on discrete choice models as applied to market data (McFadden [1981], Berry [1994], Berry et al. [1995], Verboven [1996]). We use the introduction of the regulation as an instrumental variable affecting retail prices but not directly correlated with demand. As one would expect, we find that regulation pushes prices up, thereby negatively affecting consumer welfare. Since we do not find any significant positive effect on book variety, we conclude that the introduction of regulation is detrimental for consumer welfare.

The remainder of the paper is organised as follows. Section 2 introduces a brief analysis of the book industry and a detailed description of the Italian FBP regulation. Section 3 presents our empirical strategy to identify the regulation's effect on variety and to estimate the demand for books. Section 4 describes the data. Results from both variety effects and demand estimation are presented in Section 5. Section 6 concludes and outlines extensions.

2. Institutional Background

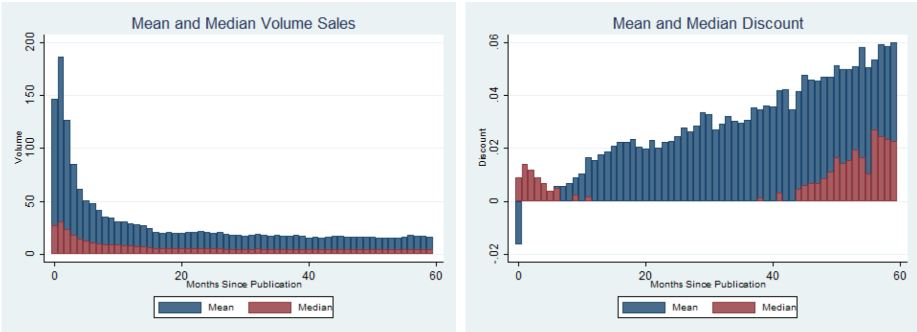
2.1. Some Economics of Books. From an economic perspective books are durable products with uncertain demand. Demand uncertainty derives from the fact that books are experience goods, as consumers can establish their quality only after consumption. Books are also very different from each other, in the way that each book has many close, but no perfect substitutes. This happens not only because cookbooks are poor substitutes for fiction books, but also because book titles within the same *genre* are not perfect substitutes either. In this sense books are unique.[2]

On the supply side, publishers focus on the production of books characterised by increasing returns to scale. Publishers have to incur in high fixed costs, most of them sunk, and negligible marginal costs. Fixed costs refer to the costs of all the activities necessary to create the book concept: the acquisition of property right,

[2]It is very hard to believe 'The Sun Also Rises' by Hemingway is a perfect substitute for 'The Great Gatsby' by Fitzgerald albeit they belong to thesame category, have been written by two American novelists of the same historical period, and have been published within two years.

editing, reviewing, and more.[3] Once the concept has been created, the cost of printing extra copies is extremely low. Then publishers determine the *cover price* of the book, which constitutes a reference for both wholesale and retail prices. At the downstream level, retailers sell books to final consumers. The margin book stores get from selling a book title is given by the difference between retail price and the discounted cover price.

Books are very heterogeneous in terms of popularity and sales, though they share



(a) Sales Trend **(b)** Discount Trend

Figure 1. Source: GfK POS Tracking.

similar trends over their lifetime. Figure 1(a) displays the mean and median volume sales trend after the publication of a book title. The difference between mean and median highlights the heterogeneity of book sales. However, most of the sales happen within the first two years after publication, with a spike when books just hit the market, suggesting that books experience the highest popularity just after release.[4] Accordingly, most of profits from book titles are generated within the first months after publication, and only best sellers and "classics" keep selling in the long-run. Figure 1(b) shows the same kind of trend for discounts, which are defined as the percentage difference between cover price and retail price. Discounts increase as books become older and less popular, hence retail prices reduce together with consumer demand. Thus, although books are durable products, the lifetime of the average book-title is quite short, as it lasts for no more than two years. We take into account these facts in our analysis of the effects of the regulation on the variety of

[3]Accordingly, upstream level is intrinsically oligopolistic: the existence of exclusive rights makes the demand faced by firms inelastic, allowing them to exercise market power.

[4]The trend does not change significantly when we plot the same graph over different genres.

books in the marketplace. In particular, since sales happen in the very first months after publication, and old books are discounted anyway, we focus on books that have been published for no more than two years.[5]

2.2. The Italian Fixed Book Price Regulation. Since September 2011 the *Levi law* regulates the Italian book market. The aim of the regulation is to relax retail price competition by imposing a maximum discount retailers can deduct from the cover price. Such kind of regulation, sometimes referred to *Fixed Book Price Agreement* (FBP) is not new in Europe: 14 out of 27 EEA member states apply FBP in different forms.[6] The reasons for such regulation relate to the special role of books as a vehicle of knowledge diffusion. The idea is that, by relaxing price competition, book supply should increase both in terms of variety (i.e. more books published) and in terms of efficacy of the retail distribution. This would create positive spillovers for knowledge diffusion in society.

The Italian book market experienced two waves of regulations. The first one, from 2001 to 2004, was a first attempt to impose a FBP in order to analyse its effects on the market. From 2005 to August 2011 the book market has been unregulated, hence retailers were able to apply any discount. The Levi law defines what a book is in legal terms and fixes a maximum discount for retailers. In particular, book stores cannot charge a discount higher than 15% off the cover price.[7] The regulation applies to all new books[8] - both printed and e-books.[9]

The effects of FBP regulation can be explained using the theory of *Resale Price Maintenance* (Mathewson and Winter [1998], Overstreet [1984]). Under FBP publishers determine the prices at which retailers sell book titles to consumers, thus removing the possibility for booksellers to compete on prices. The higher retail margins should foster both the production of books and the creation of an extensive

[5]This is also consistent with one of the exclusions allowed by the Italian regulation. See next sub-section for further details.

[6]For instance, in France the *Lang law* limits retail discounts to 5% off the cover price. In Germany retailers cannot charge any additional discount.

[7]The law allows two types of exceptions. First, publishers can make promotional campaigns on specific book titles once per year, and for a period no longer than one month, excluding December. Second, the regulation does not apply to book titles that have been published for at least 20 months and that have not been purchased by book-stores for at least 6 months.

[8]Hence second hand books are exempted.

[9]The only exception is for primary school books, of which prices are set by the Ministry of Culture every year.

network of booksellers.[10] However, by relaxing price competition among booksellers, the regulation induces higher prices and less innovation at the retail level. Hence, the tension is between the positive effect on book variety in terms of more publications, and the negative effect induced by higher prices and lower innovation at the retail level.

3. Empirical Strategy

3.1. Variety Effect. Our analysis on the effect of FBP regulation on variety in the Italian market relies on the ability to find a suitable control group. Our main idea is to look for classes of books for which the regulation is not binding, and use them as controls. We start by considering books that have been published for no more than two years. We focus on these books since most of the sales happen within the first two years after publication. Then, we use the GfK third-digit classification of genres[11] and we analyse the distribution of discounts by genre in the unregulated period. Since the regulation effectively imposes a discount ceiling at 15%, we define the control group as all the books in categories for which discounts are constantly below 15% in the un-regulated period covered by our data, hence from January 2008 to August 2011. Formally, denote by d the discount and let F_g be the c.d.f. of d for the $g = 1, 2, ..., G$ genre for the selected data. Let $\tilde{d}_g$ be the discount at the 95th percentile of the distribution of F_g, that is

$$\tilde{d}_g = F_g^{-1}(\frac{95}{100}).$$

Then we define the control group C by the following rule:

$$C = \{g = 1, 2, ..., G \mid \tilde{d}_g \leq \frac{15}{100}\}, \tag{22}$$

while the treatment group is defined by all the genres such that (22) does not hold. The control group identifies book categories for which the regulation is not binding, since discounts higher than 15% would not apply anyway. Since all the books in $g \in C$ are not affected by the regulation, then the publishing decision for these books are not affected by the regulation either. Table 4 displays some of the third-digit

[10] In fact the first article of Levi law states that "[The Fixed Book-Price Regulation] aims to contribute to the development of the publishing sector, to support literature creativity, to promote book reading, together with the diffusion of culture, and the protection of pluralism of information."

[11] See Section 4 for more on GfK classification.

genres distributions. Overall, we find 19 out 98 categories satisfying (22).
Identification of the impact of the policy change on the variety of books published is then obtained within a DiD framework. Denote by V_{gt} the number of books published in genre g, in month t. The basic empirical specification is of the form:

$$\ln(V_{gt}) = \beta_0 + \beta_1 Post_t + \beta_2 Treat_g + \beta_3 Post_t \times Treat_g + \tau_t + z_g + e_{gt} \tag{23}$$

where $Post_t$ is an indicator variable equal to one after regulation, $Treat_g$ is an indicator variable for book genres most likely affected by the regulation (treatment group), $Post_t$ x $Treat_g$ denotes their interaction; τ_t denotes the time fixed effects and z_g the book genres fixed effects and e_{gt} is a random shock with $E(e_{gt}|Post_t,\ Treat_g) = 0$. β_3 is the crucial parameter capturing the impact of the policy change.
The key identifying assumption is that number of books published in each genre would be the same in the treatment and control groups in the absence of treatment. We test the robustness of this assumption by also adding a differential trend for each book genre and by estimating a dynamic model interacting $Treat_{\mathrm{g}}$ with indicator variables for 12 months before and after the policy change,

$$\ln(V_{gt}) = \beta_0 + \beta_1 Post_t + \beta_2 Treat_g +$$

$$+\beta_{3,T-12}\left[Treat_g \text{ x } D^{T-12}\right] + \beta_{3,T-11}\left[Treat_g \text{ x } D^{T-11}\right] + \dots$$

$$\dots + \beta_{3,T+11}\left[Treat_g \text{ x } D^{T+11}\right] + \beta_{3,T+12}\left[Treat_g \text{ x } D^{T+12}\right] + \tau_t + z_g + e_{gt} \tag{24}$$

where $D^{T-i} = 1$ in the i^{th} month before deregulation.[12] The last period (T+12) includes all the observations 12 or more months after the policy change.[13]

3.2. Demand Estimation. For estimating the demand for books we follow the literature on discrete choice models as applied on aggregate market-level data (McFadden [1981], Berry [1994], Berry et al. [1995], and Verboven [1996]).
Suppose the utility function of consumer i is given by:

$$\mathrm{u_{ij}} = \mathrm{x_j}\beta + \xi_{\mathrm{j}} - \alpha \mathrm{p_j} + \varepsilon_{\mathrm{ij}} \equiv \delta + \varepsilon_{\mathrm{ij}} \tag{25}$$

[12] The results are robust with respect to the choice of the window around the policy change.
[13] The omitted indicator variable covers the period 12 or more months before regulation. See David Autor [2003] or Laporte and Windmeijer [2005] for a discussion of this approach.

where

$$\delta_{\mathrm{j}} \equiv \mathrm{x_j}\beta \;-\; \alpha \mathrm{p_j} + \xi_{\mathrm{j}},$$

$$i = 1, ..., I, \quad j = 1, \ldots, J$$

x_j is a K-dimensional vector of the observed product characteristics, ξ_j stands for the mean of consumers' valuation of the unobserved by the econometrician product characteristics and ε_{ij} is a mean-zero stochastic term.
A consumer will purchase the book that will maximize his utility. Thus, consumer i chooses product j if and only if:

$$u_i\left(x_j, p_j, \xi_j; \theta\right) > \; u_i\left(x_r, p_r, \xi_r; \theta\right) \qquad r = 0, 1, \ldots, J$$

where product 0 stands for the outside good (of not buying any book) and θ is a vector that includes all parameters to the model. As customary, the utility of the outside good is normalized to zero.
Hence, for a given θ the set that leads to a choice of good j is

$$A_j\left(x,\; p,\; \xi;\; \theta\right) = \{\zeta : u_{ij} \geq u_{ir}\}, \quad r = 0, 1, \ldots, J$$

where A_j is the set of values for ζ that induces the choice of good j. If $\mathrm{f}(\zeta)$ is the distribution of preferences in the population, then the market share of good j is:

$$s_j\left(x,\; p; \theta\right) = \int_{\zeta \in A_j(\theta)} \mathrm{f}(\zeta)\; \mathrm{d}\zeta \tag{26}$$

Assuming that the potential market size, M, is the whole adult population, then the market shares can be calculated as $s_i = q_i/M$ and $s_0 = (M - \sum q_k)/M$

Assuming that ε_{ij} is identically and independently distributed across products and consumers with the "type I extreme value" distribution then, the market share of product j is given by the logit formula:

$$s_j\left(\delta\right) = \; e^{\delta_j} / \sum_{k=0}^{N} e^{\delta_k} \tag{27}$$

In addition, we also allowed for a more flexible substitution pattern among books, by allowing ε_{ij} to be correlated across books of the same genre rather than independently distributed. This assumption gives rise to the Nested Logit model, where the utility is now given by:

$$u_{ij} = \delta_j + v_{ij}\left(\sigma\right) + \left(1 - \sigma\right)\varepsilon_{ij} \tag{28}$$

And the market share for good j is:

$$s_j\left(\delta\right) = \frac{e^{\delta_j/(1-\sigma)}}{I_g} \frac{I_g^{\ 1-\sigma}}{\sum_g \left(\ I_g^{\ 1-\sigma}\ \right)} \tag{29}$$

where: $I_g = \sum_{j \in g} e^{\delta_j/(1-\sigma)}$

since the utility of the outside good is zero, then δ_0 and I_0 are equal to zero and one respectively.

Hence, following Berry (1994) our Logit demand function is

$$\ln\left(s_j\right) - \ln\left(s_0\right) = \beta x_j - ap_j + \xi_j \tag{30}$$

whereas the Nested Logit demand function is

$$\ln\left(s_j\right) - \ln\left(s_0\right) = \beta x_j - ap_j + \sigma ln(s_{j|g}) + \xi_j \tag{31}$$

Identification of the coefficients on price and within group market shares is based on an assumption and a vector of instrumental variables. We use two types of instruments. First, following our earlier discussion, we use the introduction of regulation as an exogenous change that affects the price of books. Second, following Berry et al. [1995] we assume that the unobserved product level errors are uncorrelated with the observed product characteristics. In other words, that the location of products in the characteristics space is exogenous. Given this exogeneity assumption, we calculate the number of own and rival books in each genre for each publisher.

4. Data

Our dataset comes from GfK Point of Sales (POS) tracking, and it consists of a retail panel with actual sales data of the Italian book market. The data cover a period starting from January 2008 to December 2015, with a total of more than 24 million of observations. Each observation is a book - identified by an ISBN code - for which we observe quantity of sales and average retail prices by distribution channel on a monthly basis. Three distribution channels are considered, namely independent book stores, vertically integrated book stores, and e-commerce.[14] We also observe the cover price of the book, together with its title, author, publisher, format, and date of publication. These are all book idiosyncratic characteristics that do not vary over time. The percentage difference between cover price and retail price determines book discount.[15] Finally, we use the difference between the date of sale and the date of publication to compute a measure of the "age" of the book. Table 1 displays summary statistics.

Table 1. Summary Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
Volume Sales	23,750,090	23.34	189.97	1	143086
Price	23,750,090	14.83	10.34	1	100
Cover Price	23,750,090	16.03	11.01	1	406
Discount	23,750,090	0.07	0.14	-1	0.99
Age	23,750,090	59.70	66.81	0	1272

Source: GfK POS Tracking. Prices are expressed in EUR. Age is on a monthly basis, and it is defined as the difference between the date of sale and the date of publication.

Each ISBN is uniquely allocated by GfK over three-digit categories identifying the "genre" of the book. The upper level consists of 5 categories, while the middle and lower levels collect 26 and 98 sub-genres respectively. Table 2 shows how data are distributed across genre's upper level, distribution channels, and format. We use the third-genre classification (lower level) in both our demand estimation and variety analysis. In the demand estimation, we neglect the distribution channel differentiation. We carefully select our data in order to keep only the book titles that have been published before the regulation, and that keep selling when FBP is in place. In this way we get rid of all the books that "expired" before the regulation, as well as of those that have been published when the regulation was already in place.

In order to correctly identify the impact of the regulation on variety, we need to clean our dataset from re-editions. This is because we are interested in the true number of new publications affecting the book marketplace. The combination of

[14]Independent book stores are local brick and mortar shops distributed over the whole national territory. Vertically-integrated book stores refer to brick and mortar shops that are directly controlled by the publishers. Finally, e-commerce includes purchases from Internet websites.

[15]Therefore discounts vary over time and channel distributions together with retail prices.

Table 2. Distribution Over Categories

	Frequency	Percent	Cumulative
Distribution Channel			
VI Book stores	9,104,345	38.33	38.33
E-Commerce	7,670,319	32.30	70.63
Independent Book stores	6,975,426	29.37	100
Format			
Not Defined	11,369	0.05	0.05
Standard Format	19,752,442	83.17	83.22
Pocket	3,986,279	16.78	100
Genre: Upper Level			
Not Defined	194,361	0.82	0.82
Kids	3,279,076	13.81	14.62
Fiction	5,429,826	22.86	37.49
History, Science, Art	10,531,094	44.34	81.83
Miscellaneous	4,315,733	18.17	100

Source: GfK POS Tracking.

author and title together with the date of publication allows us to identify whether the ISBN refers to a book that has been published for the first time or to a successive edition. In particular, when we observe variation of ISBN for the same title/author of the book, we identify as first edited the book with the earliest date of publication. However, the identification is limited to the period covered by our data, thus we cannot ensure that book titles that appear as first edited in our dataset are in fact successive edition of books published before 2008. In order to overcome this issue, we merge our original dataset with data provided by Informazioni Editoriali (IE).[16] The IE dataset contains a complete list of books published before 2008. To each ISBN they associate an identification number called "Opera ID". Each Opera ID identifies ISBN codes that refer to the same book. The variation of the ISBN within the Opera ID refers to re-editions of the same book. We include this information in the original GfK data in order to further clean our dataset from re-editions. Table 3 displays the number of titles in our dataset, with the identification of the first editions. According to our merged data, almost 80% of the book titles in our dataset are books published for the first time.

[16]IE provides the most complete database on books published in Italy. The database is mainly used by libraries and book-stores.

Table 3. First Editions vs. Successive Editions

N Titles	Frequency	Percent	Cumulative
Successive Editions	160,516	21.7	21.7
First Editions	579,341	78.3	100
Total	739,857	100	

Source: GfK POS Tracking merged with IE data. Total number of titles (ISBN) by first vs. successive editions.

5. Empirical Results

5.1. The Effect of Regulation on Variety. Table 5 reports the results on the impact of regulation on the number of books for different symmetric windows of time around the time of regulation (columns 1-9) and using the whole dataset (columns 10-12). Fixing a symmetric one-year window around the change in regulation and without any fixed effects seems to produce a positive and significant at 10% positive effect in column 1. However, once we control for either genre (column 2) or in addition time fixed effects (column 3) turns the coefficient on the interaction of $Post_t$ x $Treat_j$ insignificant, hence indicating that the introduction of regulation had no discernible impact on the number of books published for book genres affected by the regulation compare to the ones not affected.

In columns 4-6 and 7-9 we repeat the same exercise augmenting the window around the regulation to either 2 or 3 years respectively, hence allowing more time to the publishers to react. Results in columns 6 and 9 when we control for both time and genre fixed effects seem to indicate a negative and significant impact. In other words, publishers in book genres most likely affected by the introduction of regulation seem to react by *reducing* the number of titles published, in total contrast to the main argument made in favour of this regulation that retail price maintenance would allow them to publish a larger variety of books. Moreover, looking at the results across the different time windows and comparing coefficients in columns 3, 6 and 9 indicates that this effect increases rather than decreases over time, so given more time publishers seem to react even more by reducing the number of books published. A similar picture is drawn when we look at the whole dataset available in the last three columns, where again the effect of regulation is negative and significant.

In Table 6 we test the robustness of our findings by estimating the most restrictive specification with time and genre fixed effects and also allowing for differential trends across book genres. Results now seem more online with our original intuition in that publishers seem to react positively in the short run (column 1 for the one year symmetric window) but this effect fades away when one enlarges the time window around the introduction of regulation. The coefficient on the interaction is only significant at the 10% in column 2 (two year window) and insignificant for the last two columns when we look at the three year window or when looking at the whole dataset.

In Table 7 we report the results by estimating the dynamic effects of the introduction of regulation by estimating model 2. In column 1 we report the results from the baseline specification, whereas column 2 adds individual trends by book genre. The respective estimated coefficients are plotted in Figure 2(a) and 2(b) to ease the comparison. What we observe is a temporary increase in the estimated coefficients immediately after the introduction of regulation (two and three months after the introduction) but no significant impact of regulation overall if one calculates the average of the coefficients before and after the introduction of regulation (solid grey lines in the two graphs).

Therefore, taken all together our investigation of the impact of regulation on the number of books published seem to strongly indicate that the introduction of regulation had no significant effect. In other words, publishers do not seem to react by either producing more or fewer books by genre.

5.2. Estimating The Demand for Books. Table 8 summarizes the results from estimating the Logit demand model. In columns 1-4 we report results without instrumenting the price coefficient and by experimenting with different fixed effects. The coefficient on price is negative and significant as expected and it is mostly affected when controlling for time (columns 3 and 4) rather than genre (column 2) fixed effects. However, either controlling for month and year fixed effect separately (column 3) or jointly (column 4) does not seem to make much of a difference.

In the next four columns we repeat the same steps, but now instrumenting price with the number of own and rival number of books by genre and publisher and using a variable called *regulation* that takes a value equal to one after the introduction of regulation. In the last column, we also experiment by altering the *regulation* instrument and making it one only for the book genre most likely affected by regulation (those with discounts at 95 percentile as we discussed earlier for the variety of books). Results seem to indicate that instrumenting for price is important as it significantly affects the coefficient and moving it away from zero, as suggested by the previous literature.

Table 9 summarizes the results for our Nested-Logit specification. First three columns report different specifications where we progressively control for more fixed effects, whereas the last four columns we instrument both the price and the within-nest market share. Looking at the most restrictive specification in column 6, we can see that the coefficient on price is negative and significant and about half the magnitude of the one estimated with the simple Logit model. The coefficient on the within nest is positive and statistically different from either zero or one as indicated by theory, highlighting relatively small substitutability across nests (book genres) as one would expect.

We compute the variation in consumer welfare induced by the introduction of the FBP regulation by applying the main result of Small and Rosen [1981] to our Logit and Nested-Logit demand models. Omitting the constant for simplicity, consumer welfare in the Nested-Logit model is given by[17]

$$CS = M \ln(\sum_{g}(\sum_{j \in g} \exp(\frac{\delta_j}{1-\sigma}))^{(1-\sigma)}) \tag{32}$$

Given our parameter estimates[18], we compute the observed consumer welfare under FBP versus the counterfactual in which FBP regulation does not take place. The average price change induced by the regulation is derived from the first-stage estimates of Table 10. Table 11 shows the results for both Logit and Nested-Logit model. As expected, the upward shift in prices induced by the regulation reduces consumer welfare. The welfare change predicted by the Nested-Logit model is significantly higher than the Logit one, which is consistent with the coefficient of the

[17] Consumer welfare from Logit model can be derived by setting $\sigma = 0$ in (32).

[18] We use coefficients from Table 8 column 9 and from Table 9 column 7. Thus, consistently with the analysis on variety, we use the specifications where the instrumental variable is the interaction between the treatment group and the post dummy.

instrumental variable $Treat \times Post$ of Table 10 column 2. Thus, Logit model predicts a welfare increase of 24% if the regulation would not take place, while Nested-Logit model predicts an increase of almost 54% if the market was unregulated.

6. Concluding Remarks

In this paper, we present the first systematic evidence of the impact of FBP regulation in the Italian book market. Despite its positive intent, we find no significant evidence of any effect of the regulation on the publication of new books. On the contrary, we find that the regulation raises retail prices, thereby affecting negatively consumer welfare.

Although our results are quite robust and economically sizeable, we neglect two important factors. First, we do not analyse the substitution effects among distribution channels. E-commerce is definitely the channel that has been most heavily affected by the introduction of the regulation, hence e-commerce customers are those who have been mostly affected by the introduction of the FBP regulation. Secondly, we do not analyse how publishers internalise the new regulation in their pricing decision. This is a non-trivial problem as one would like to compare two similar books published under different market conditions, and analyse the price differential. However, since even books in the same genre are not close substitute, such comparison is not an easy task.

Finally, we do not look at re-editions and that this is not a cost-benefit analysis of all aspects of this regulation (for example we do not look at the number of book stores), yet still this is the first systematic analysis on the central premise that by killing retail competition publishers could maintain a larger variety.

Table 4. Discount Distribution By Genre (Third-Digit)

	Genre3	mean	p50	p75	p90	p95	p99	N
1	Not Defined	0.036	0.000	0.034	0.121	0.190	0.500	18873
2	Other/Fiction	0.036	0.015	0.059	0.145	0.200	0.338	18769
3	Other/Sport	0.007	0.003	0.031	0.118	0.159	0.302	13738
4	Animals	0.007	0.005	0.022	0.091	0.149	0.317	9750
.								
.								
.								
29	Economics & Finance	0.056	0.020	0.100	0.161	0.207	0.376	56721
30	Encyclopedia	0.030	0.016	0.073	0.133	0.200	0.474	8516
31	Fantasy & Science Fiction	0.047	0.019	0.058	0.138	0.202	0.380	14271
32	Fiction - Primary School	0.022	0.015	0.060	0.124	0.166	0.258	44635
33	Fiction - Secondary School	0.043	0.019	0.070	0.145	0.215	0.302	25070
34	Fiction - Young	0.029	0.020	0.044	0.101	0.146	0.301	7722
35	Philosophy	0.060	0.025	0.098	0.176	0.235	0.416	68645
36	Fitness	0.028	0.010	0.038	0.118	0.149	0.294	7700
.								
.								
.								
96	Thrillers	0.036	0.022	0.068	0.149	0.202	0.309	67285
97	Humor	0.022	0.009	0.029	0.093	0.155	0.289	5465
98	Wine & Drink	0.009	0.006	0.024	0.109	0.177	0.299	4861

Discount Distribution by genre from GfK POS Data. Statistics are derived using data from January 2008 to August 2011, and for books that have been published for no longer than 24 months. Categories with discounts lower than 15% at the 95th percentile belongs to the control group.

Table 5. Estimating The Variety Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	1 Year Window			**2 Year Window**			**3 Year Window**			**All Data**		
Dependent Variable	ln(number books)			ln(number books)			ln(number books)			ln(number books)		
Treat X Post	0.104*	0.00670	-0.0427	-0.0693	-0.112**	-0.134**	-0.0990*	-0.139***	-0.173***	-0.116*	-0.167**	-0.214***
	(0.058)	(0.0492)	(0.0517)	(0.0559)	(0.0486)	(0.0508)	(0.0560)	(0.0493)	(0.0533)	(0.0677)	(0.0677)	(0.0712)
Post	-0.157***	-0.118***	0.503***	-0.0364	-0.0397	0.559***	0.00879	0.00553	0.592***	0.0185	0.0186	0.468***
	(0.047)	(0.0410)	(0.0996)	(0.0465)	(0.0413)	(0.108)	(0.0485)	(0.0429)	(0.102)	(0.0606)	(0.0621)	(0.123)
Treat	0.324			0.439*			0.450*			0.444*		
Genre FE	(0.265)			(0.262)			(0.254)			(0.251)		
Observations	1,994	1,994	1,994	3,934	3,934	3,934	5,840	5,840	5,840	7,678	7,678	7,678
Adjusted R2												
Clusters	97	97	98	98	98	98	98	98	98	98	98	98
Genre FE		yes	yes		yes	yes		yes	yes		yes	yes
Time FE			yes			yes			yes			yes

The dependent variable is the logarithm of the number of new titles by genre.Post takes value one after September 2011. Treat takes value one if (22) is not satisfied. Standard errors clustered at the book genre third-level are reported in parenthesis below coefficients: *significant at 10%; **significant at 5%; ***significant at 1%.

Table 6. Estimating the Variety Effect: Robustness

	(1)	(2)	(3)	(4)
	1 year window	**2 year window**	**3 year window**	**all data**
Dependent Variable	ln(number books)	ln(number books)	ln(number books)	ln(number books)
Treat X Post	0.328***	0.139*	0.0493	0.0286
	(0.088)	(0.0761)	(0.0728)	(0.0560)
Post	-4.758***	-11.47***	-15.86***	22.33***
	(0.975)	(2.267)	(3.343)	(3.932)
Observations	1,994	3,934	5,840	7,678
Clusters	98	98	98	98
Genre FE	yes	yes	yes	yes
Time FE	yes	yes	yes	yes
Trend X Genre	yes	yes	yes	yes

The dependent variable is the logarithm of the number of new titles by genre.Post takes value one after September 2011. Treat takes value one if (22) is not satisfied. Standard errors clustered at the book genre third-level are reported in parenthesis below coefficients: *significant at 10%; **significant at 5%; ***significant at 1%.

Table 7. Dynamic Impact of FBP on Variety

	(1)	(2)
	All Data	**All Data**
Dependent variable	ln(number books)	ln(number books)
$Treat \times Post_{t-12}$	-0.375**	-0.312*
	(0.187)	(0.184)
$Treat \times Post_{t-11}$	0.058	0.120
	(0.149)	(0.143)
$Treat \times Post_{t-10}$	-0.016	0.0385
	(0.132)	(0.132)
$Treat \times Post_{t-9}$	0.299	0.369*
	(0.183)	(0.194)
$Treat \times Post_{t-8}$	-0.152	-0.0724
	(0.173)	(0.179)
$Treat \times Post_{t-7}$	-0.313**	-0.228
	(0.143)	(0.140)
$Treat \times Post_{t-6}$	-0.348***	-0.268**
	(0.119)	(0.122)
$Treat \times Post_{t-5}$	-0.076	0.0236
	(0.138)	(0.149)
$Treat \times Post_{t-4}$	-0.225*	-0.138
	(0.130)	(0.134)
$Treat \times Post_{t-3}$	-0.240	-0.145
	(0.175)	(0.168)
$Treat \times Post_{t-2}$	-0.389**	-0.299*
	(0.155)	(0.156)
$Treat \times Post_{t-1}$	0.000	0.0954
	(0.223)	(0.232)
$Treat \times Post_{t}$	-0.259	-0.161
	(0.181)	(0.167)
$Treat \times Post_{t+1}$	0.023	0.127
	(0.147)	(0.150)
$Treat \times Post_{t+2}$	0.214*	0.330**
	(0.128)	(0.137)
$Treat \times Post_{t+3}$	0.282	0.397**
	(0.170)	(0.188)
$Treat \times Post_{t+4}$	-0.466**	-0.344
	(0.213)	(0.219)
$Treat \times Post_{t+5}$	-0.359**	-0.232
	(0.147)	(0.150)
$Treat \times Post_{t+6}$	-0.061	0.0707
	(0.204)	(0.224)
$Treat \times Post_{t+7}$	-0.165	-0.0394
	(0.151)	(0.159)
$Treat \times Post_{t+8}$	-0.379***	-0.243*
	(0.121)	(0.133)
$Treat \times Post_{t+9}$	-0.171	-0.0393
	(0.141)	(0.155)
$Treat \times Post_{t+10}$	-0.382*	-0.252
	(0.195)	(0.176)
$Treat \times Post_{t+11}$	-0.211	-0.0597
	(0.146)	(0.150)
$Treat \times Post_{t+12}$	-0.283***	-0.0513
	(0.087)	(0.101)
Observations	7,678	7,678
Adjusted R2		
Clusters	98	98
Genre FE	yes	yes
Time FE	yes	yes
Trend X Genre		yes

Standard errors clustered at the book genre third-level are reported in parenthesis below coefficients: *significant at 10%; **significant at 5%; ***significant at 1%.

Table 8. Estimating Logit Demand for Books

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Estimation method	OLS	FE	FE	FE	IV	IV	IV	IV	IV
Price	-0.021***	-0.0119***	-0.0103***	-0.0103***	-0.221***	-1.125***	-0.710**	-0.697**	-0.616**
	(0.002)	(0.00152)	(0.00148)	(0.00147)	(0.0287)	(0.276)	(0.330)	(0.326)	(0.281)
Book age	-0.003***	-0.00340***	-0.000855	-0.000808	-0.00744***	-0.0214***	-0.0128**	-0.0127**	-0.0112**
	(0.001)	(0.00101)	(0.000775)	(0.000772)	(0.00137)	(0.00567)	(0.00609)	(0.00600)	(0.00526)
First Edition	-0.659***	-0.591***	-0.593***	-0.591***	-0.383***	0.310	-0.0220	-0.0310	-0.0965
	(0.053)	(0.0415)	(0.0397)	(0.0396)	(0.0998)	(0.323)	(0.277)	(0.273)	(0.233)
First Edition x Book age	0.002**	0.00274***	0.00124*	0.00121*	0.00432***	0.00746**	0.00462*	0.00458*	0.00412*
	(0.001)	(0.000906)	(0.000732)	(0.000729)	(0.00111)	(0.00357)	(0.00278)	(0.00274)	(0.00243)
Observations	2,259,461	2,259,461	2,259,461	2,259,461	2,259,461	2,259,461	2,259,461	2,259,461	2,259,461
Adjusted R2	0.005								
Clusters	98	98	98	98	98	98	98	98	98
Genre FE		yes	yes	yes		yes	yes	yes	yes
Year and Month FE			yes				yes		
Time FE				yes				yes	yes

Depend variable is $\ln(s_j) - \ln(s_0)$. Standard errors clustered at the book genre third-level are reported in parenthesis below coefficients: *significant at 10%; **significant at 5%; ***significant at 1%.

Table 9. Estimating Nested-Logit Demand for Books

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Estimation method	OLS	FE	FE	IV	IV	IV	IV
Price	-0.027***	-0.0252***	-0.0252***	-0.458***	-0.316***	-0.319***	-0.217***
	(0.004)	(0.00427)	(0.00428)	(0.0920)	(0.0658)	(0.0668)	(0.0607)
ln(Within-Nest Share)	0.546***	0.546***	0.548***	0.391**	0.248**	0.257**	0.153**
	(0.053)	(0.0522)	(0.0525)	(0.162)	(0.108)	(0.111)	(0.0740)
Book age	-0.001	0.00134*	0.00142*	-0.0114***	-0.00668***	-0.00666***	-0.00461***
	(0.001)	(0.000800)	(0.000794)	(0.00280)	(0.00195)	(0.00197)	(0.00162)
First Edition	-0.525***	-0.511***	-0.508***	0.0353	-0.180	-0.170	-0.337***
	(0.135)	(0.139)	(0.139)	(0.196)	(0.127)	(0.129)	(0.103)
First Edition x Book age	0.001	-0.000395	-0.000447	0.00571***	0.00338***	0.00335***	0.00253**
	(0.001)	(0.000628)	(0.000625)	(0.00177)	(0.00122)	(0.00123)	(0.000982)
Observations	2,259,461	2,259,461	2,259,461	2,259,461	2,259,461	2,259,461	2,259,461
Adjusted R2	0.005						
Clusters	98	98	98	98	98	98	98
Year and Month FE		yes			yes		
Time FE			yes			yes	yes

Depend variable is $\ln(s_j) - \ln(s_0)$. Standard errors clustered at the book genre third-level are reported in parenthesis below coefficients: *significant at 10%; **significant at 5%; ***significant at 1%.

Table 10. First Stages: Logit Model & Nested-Logit Model

	Logit Model	**Nested-Logit Model**	
Dependent Variable	Price	Price	$\ln(s_{j\|g})$
	(1)	(2)	(3)
Book Age	-0.016	-0.019	-0.001
	(0.004)	(0.004)	(0.000)
First Edition	0.816	1.091	-0.500
	(0.240)	(0.483)	(0.046)
First Edtion x Book Age	0.004	0.007	0.001
	(0.003)	(0.003)	(0.001)
Treat X Post	0.414	2.237	0.136
	(0.068)	(1.196)	(0.114)
Within-Nest:			
N Publishers books	-0.003	-0.007	0.001
	(0.002)	(0.002)	(0.000)
N Rivals Book	0.000	-0.001	-0.001
	(0.000)	(0.000)	(0.000)
F-Test	19.57	10.75	228.33

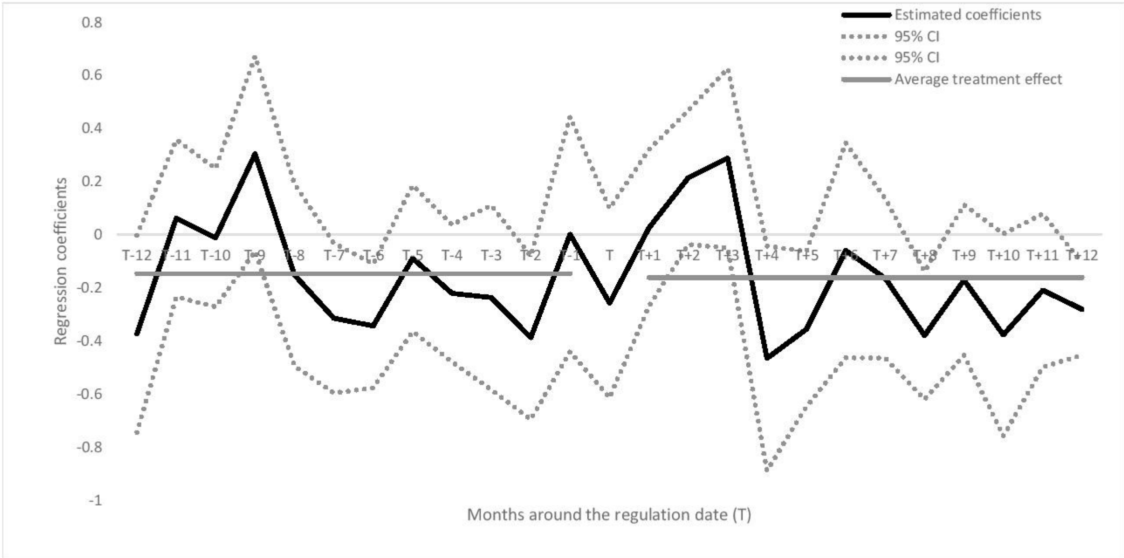
The table shows estimated coefficients from first-stage regressions, and F-tests. Column 1 refers the first stage of the model in column 9 in Table 8. Columns 2 and 3 are first-stage regressions of model in column 7, Table 9. Standard errors clustered at the book genre third-level

Table 11. Welfare Evaluation

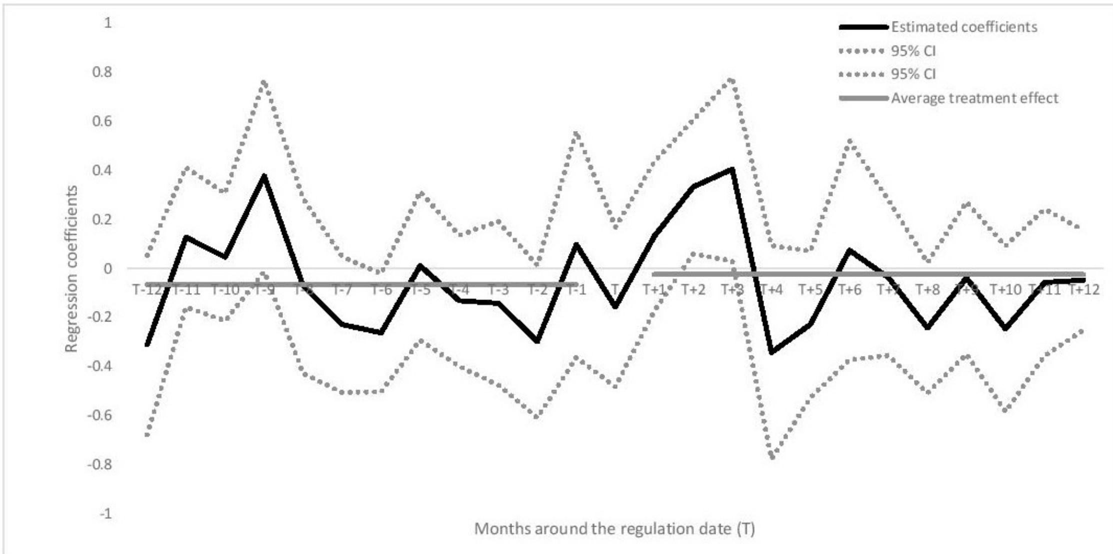
	FBP Regulation		
	YES	NO	Δ%
Logit	90885116	112323333	23.59
Nested-Logit	114924705	176655771	53.71

Data in EUR. The Table displays total consumer welfare computed using date from September 2011 to August 2013. Counterfactuals are derived using estimated coefficient from Table 8 column 9 and from Table 9 column 7.

Figure 2. Dynamic Impact of Regulation on Book Variety



(a) The figure plots the regression coefficients from column 1 Table 7, capturing the dynamic impact of regulation on the logarithm of number of books. Each period corresponds to a month. The 95 percent confidence interval is based on standard errors clustered at the genre level.



(b) The figure plots the regression coefficients from column 2 Table 7, capturing the dynamic impact of regulation on the logarithm of number of books. Each period corresponds to a month. The 95 percent confidence interval is based on standard errors clustered at the genre level.

CHAPTER 3

The Microeconomics of Discrete Choice Models

1. Introduction

The aim of discrete choice models is to describe decision makers' choices among alternatives. I refer to him or her as *agent*, or *consumer*, but we should keep in mind that discrete choice framework can be used to describe any decision process that satisfies the fundamentals of the model.[1]

Discrete choice models date back to the seminal work by McFadden et al. [1973]. However, due to data limitations, only in the nineties these models received significant attention, especially after the work of Berry [1994] and Berry et al. [1995]. These contributions develop applications of discrete choice models to estimate demand for differentiated products. Nowadays, these are standard methods in empirical Industrial Organization (IO).

In this paper, I analyse the main features of discrete choice models, with particular emphasis on the primitives of the models and the analytical solutions of the equations to be estimated. In particular, I link these models to standard microeconomic theory of consumer behaviour, I show how to solve analytically random utility maximization, and I derive choice probabilities in a general setting. Finally, I concentrate to the analysis of Logit and Nested-Logit models.

Demand estimation is extremely relevant in microeconomics for many reasons. First, it allows to understand the main factors affecting consumers choices: how does a price increase affect sales? What are the main product characteristics affecting consumers choice? What is the benefit from the introduction of new goods? What are the welfare implications of a new policy affecting prices? Second, demand estimation gives us a better understanding of firms' business practices, as demand affects firms' profitability and competitive landscape: what is the degree of substitutability with rivals' products? What is the expected profit from the introduction of a new product? Is it profitable to enter the market?

It is therefore not surprising that demand estimation is the point of departure of many economic analyses in empirical IO. However, demand estimation faces many different computational challenges over the years. The first challenge is known as the *too-many-parameters* problem. Markets are characterised by many substitute

[1]Therefore the decision of a firm over a set of available investment choices can also be coherently described by discrete choice models.

products, and a price increase of one product affects both own-demand and the demand of all other goods. To capture these own and cross-price effects, one would like to regress quantity of each product over all prices, which means estimating a system with as many parameters as moment conditions. In a market for differentiated product, one would also like to add products attributes in order to capture similarities among the goods in the market-place. However one additional covariate implies that the number of parameters is greater than moment conditions, making the system unidentified. Discrete choice models reduce significantly the parameter space, providing a solution to the *too-many-parameters* problem.
The second challenge refers to the computational feasibility of discrete choice models, as they are often non-linear in parameters, and involving the computation of integrals that might not have closed form solution. The contributions of Berry [1994] and Berry et al. [1995] provide solutions to these computational problems, allowing for the estimation of complex demand systems with rich substitution patterns.
The remainder of the paper is organised as follows. Section 2 introduces the primitives of discrete choice models and the random utility maximization problem. In section 3 I derive choice probabilities for Logit and Nested-Logit models, and I introduce the standard set-up for the demand estimation. Finally, in section 4 I analyse the endogeneity problem of these models.

2. Primitives

2.1. Deterministic Model. Discrete choice models are characterised by the fundamental assumption that decision makers have to choose among a set of countable and mutually exclusive alternatives. The agent then chooses, among the feasible alternatives, the one (and only) she prefers the most. Indeed, according to Train [2009], the set of alternatives, called *choice set*, has to satisfy the following three properties in order to fit discrete choice framework:

(1) Alternatives in the choice set must be *mutually exclusive*;
(2) The choice set must be *exhaustive*;
(3) The number of alternatives must be *finite*.

Thus, discrete choice models are able to describe those decision processes where agents need to choose just **one** alternative over the set of **all** feasible alternatives which have to be **finite** in number.

The decision maker is assumed to select the alternative she likes the most. We describe this step by making use of standard microeconomic tools. In particular, the agent is assumed to have a *rational* preference order over alternatives which can be represented by utility function. In this context, the alternative actually chosen is simply the one that yields the highest level of utility over the alternatives.

Consider the following setting. Let C be the finite discrete choice set. For each alternative $j \in C$ there exists a vector of attributes $x_j \in X$ with L elements. The agent is assumed to have rational preferences over X, hence, there exists a utility function $U : X \to R$ representing those preferences. Finally, the individual is assumed to choose the alternative j^* that yields the highest level of utility, i.e.

$$U(x_{j^*}) \geq U(x_j) \;\; j \in C$$

In this context, each alternative is seen as a bundle of characteristics over which the individual has a complete and transitive preference order.[2] Thus, she chooses the most valuable bundle, which is determined by comparing utility levels over available alternative in C.

The researcher does not observe the preference order - i.e. the utility levels - that the individuals get from each feasible alternative. However, we do observe *actual* choices, and assumes that, given individual's preferences, such choices are the ones yielding the maximum level of utility. We also recognise that there might be other *unobserved* factors, not included in X, that might affect agents' decision. This happens because we acknowledge the existence of *unmeasurable* factors affecting the decision process. We treat these factors as random components of individual's utility, hence we move from a deterministic to a probabilistic model of choice decisions. The following example should clarify these aspects.

Recurring Example: Laura's Choice for Dinner

To make things clear consider the following example. Laura needs to choose what

[2]A preference order is a relation that is both complete and transitive. Broadly speaking, preferences are complete if, for every pairwise comparison, the agent is always able to determine which alternative likes the most. Transitivity implies that, if A is preferred to B and B is preferred to C, then A is preferred to C too.

to have for dinner. She can choose between order take away sushi, pizza, thai, or cook something by herself. She considers all the alternatives, and she decides to go for thai. We describe this decision process using the discrete choice framework as follows.

Laura's choice set is $C = \{S, P, T, 0\}$ where capital letters stay for Sushi, Pizza, and Thai respectively, and 0 is the outside option - which in this case means to cook something else by her own. Laura has preferences over alternatives in the choice set that can be represented by a utility function $U : C \to \mathbb{R}$; hence, for every alternative, we have a real-valued utility level associated to each alternative.[3] Thus, we can define the collection of utility levels by a vector $U \in \mathbb{R}^4$, $U = \langle U_S, U_P, U_T, U_0 \rangle$.

Let us suppose Laura chooses Thai. By assuming that Laura is a utility maximizer we mean that

$$U_T \geq U_k \ \ \forall k \in C.$$

That is, Laura has chosen thai because that's the alternative yielding the highest utility given her preferences.

The researcher cannot observe the actual levels of utility Laura gets from alternatives, therefore utilities are unknown from our point of view. What we observe is instead a set of alternatives attributes which we assume enter in the decision process of the agent. In our example, observed attributes might be calories, delivery time, price, meal size, and we assume that Laura take these factors into consideration when making her decision. However we are not able to observe, say, the *pleasure* Laura gets from each meal; that is, we are not able to observe Laura's idiosyncratic factors that affect her utilities, and ultimately her choices. We recognise the existence of such factors, as well as the impossibility to measure them, and we take such factors as a random component of Laura's utility. Consequently U is a collection of random variable - or alternatively a random vector - in the space $\mathbb{R}^4$.

2.2. Probabilistic Choice System and Random Utility Maximization. In this section I follow the original setting from Manski et al. [1981]. This implies the specification of a probability choice system, and the link with random utility

[3] Recall that, in principle, utility levels are determined by the underlying alternative attributes. However we skip this passage in this example.

maximization problem. Random utilities arise because we acknowledge the existence of *unmeasurable* factors affecting the decision process, and we treat these as random component of agent's utility. In this framework, the probability to choose one alternative j is exactly equal to the probability that the utility from j is the highest among the feasible alternatives.

A *Probabilistic Choice System* (PCS) is defined by the the vector $(I, Z, \xi, \mathcal{B}, S, P)$ where:

- $I = \{1, 2, ..., \tilde{J}\}$ is the index set for alternatives;
- $Z = \{z_1^j, z_2^j, ..., z_K^j\}$ is the collection of vectors of attributes (or characteristics) over alternatives;
- $\xi : I \rightarrow Z$ is a mapping specifying the observed or measured (by the researcher) attributes;
- $\mathcal{B}$ is the family of budget sets (choice sets);
- S is the space of individuals' characteristics;
- $P : I \times \mathcal{B} \times S \rightarrow [0, 1]$ is the choice probability.

The index set I labels the *available* alternatives faced by the decision maker. For each of these alternatives, we have a $K \times 1$ vector of attributes z_j. Unfortunately, we only observe $k \leq K$ characteristic, and the observed characteristics are determined by the function ξ. For simplicity, denote by x_j the vector of observable attributes for alternative j, that is $\xi(j) \equiv x_j$.
Only a subset of I available alternatives are indeed also *feasible*: the set collecting feasible alternative is defined by $B \in \mathcal{B}$.[4] Hence the individual selects one feasible alternative in B, given her own (unobserved) preferences, and a vector of individual attributes s.

Choice probability $P(j|B, s)$ specifies the probability of choosing $j \in I$ given that the choice has to be made from $B \in \mathcal{B}$ and given individual characteristics $s \in S$. Choice probabilities are assumed to satisfy the following two conditions:[5].

[4]Feasibility arises because of environmental constraints that restrict individual's choice possibility. They might reflect, for instance, cultural and/or geographical limitations faced by the decision maker.

[5]Hence, if $j, k \in B$, and $\xi(j) = \xi(k)$, then $P(j|B, s) = P(k|B, s)$

(1) $P(B|B,s) = \sum_{j\in B} P(j|B,s) = 1$;

(2) Choice probabilities depend only on the measured attributes of alternatives and on individual characteristics:
If for $B = \{1, ..., J\}$ and $B' = \{1', ..., J'\}$ we have that $\xi(j) = \xi(j') = x_j$ for all $j = 1, 2, ..., J$ then $P(j|B,s) = P(j'|B',s)$.

The hypothesis of Random Utility Maximization (RUM) is defined formally by the vector (I, Z, ξ, S, μ) where μ is a probability measure depending on $s \in S$ on the space of utility functions on I.
Consider the set I, and assume there exists a function $U : I \to \mathbb{R}^I$. Hence we have a vector $U = \langle U_1, ..., U_I \rangle$ collecting utilities associated to alternatives. By assumption, this vector is assumed to be random, with probability distribution conditional on characteristics s given by μ. In the usual context of demand estimation, the probability μ gives the distribution of tastes over alternatives among individuals with characteristics s. Therefore it is the exogenous distribution of random tastes given s.
Let μ^B be the restriction of μ to B, hence the restriction of probability distribution over the feasible alternatives. The following assumptions are imposed on μ:

(1) The restriction of μ to the space of utility values on a finite set of alternatives B depends on the measured attributes of these alternatives:
If for $B = \{1, ..., J\}$ and $B' = \{1', ..., J'\}$ we have that $\xi(j) = \xi(j') = x_j$ for all $j = 1, 2, ..., J$ then $\mu^B = \mu^{B'}$.

(2) The probabilities of "ties" is zero:

$$\mu(\{U \in \mathbb{R}^I | U_1 = U_2\}, s) = 0$$

(3) Each RUM (I, Z, ξ, S, μ) and family of choice sets $\mathcal{B}$ generate a PCS $(I, Z, \mathcal{B}, S, P)$ via the following mapping:

For $B = \{1, ..., J\} \in \mathcal{B}$, $s \in S$, then for $k = 1, 2, ..., n$

$$P(j|B,s) = \mu^B(\{U \in \mathbb{R}^I | U_j \geq U_k,\ \forall\, k \in B\}, s) \tag{33}$$

Assumption 1 implies that two consumers with the same vector of individual characteristics s face the same utility distribution over alternatives. Assumption 2 guarantees the existence of a utility maximizing alternative, so that choice probabilities are well-defined and sum to 1. Assumption 3 imposes that choice probabilities are derived from maximization of random utility; that is, the probability of choosing alternative j given individual characteristics and choice set, is exactly equal to the probability to pick a vector of utility such that $U_k \geq U_j$ for all $j \in B$.

This is the foundation of discrete choice models. On the one hand we define choice probabilities from a set of feasible alternatives. The restrictions on choice probabilities are straightforward: they sum to one and they only depend on the vector of observable attributes for alternatives in the choice set. On the other hand, we define random utilities and we impose that choice probabilities arise exactly from random utility maximization. Moreover, observed characteristics of the alternatives determine both choice probabilities and random utilities

When μ^B can be represented by a probability density function f^B so that $\mu^B = \int f^B(U_1, ..., U_J|s)dU_1...dU_J$ then choice probability can be rewritten as:

$$P(j = 1|B, s) = \int_{U_1=-\infty}^{+\infty} \int_{U_2=-\infty}^{U_1} ... \int_{U_j=-\infty}^{U_1} f^B(U_1, ..., U_J, s)dU_1...dU_J =$$

$$= \int_{U_1=-\infty}^{+\infty} F_1^B(U, ..., U, s)dU \tag{34}$$

where F^B is the c.d.f. of f^B, and F_1^B denotes the derivative of F^B with respect to U_1.[6]

When U_j's are i.i.d random variables, choice probabilities take the following form:[7]

$$P(j = 1|B, s) = \int_{U_1=-\infty}^{+\infty} \prod_{j=2}^{J} F^B(U)f^B(U)\, dU. \tag{35}$$

In fact, since U_j's are independent, then $\Pr\{U_j \leq U_1, \ \ \forall j \neq 1\} = \Pr\{U_2 \leq U_1\} \times \Pr\{U_2 \leq U_1\} \times ... \times \Pr\{U_J \leq U_1\}$. Moreover, as they are all identically distributed,

[6]Notice that by integrating f^B over all $U_j \neq U_1$, we get F_1^B. By further averaging F_1^B over all the possible values of U_1, we get the formula for the choice probability of alternative $j = 1$.

[7]This is exactly the case of the Logit model.

then $\Pr\{U_j \leq U_1\} = \Pr\{U_h \leq U_1\}$ for all $j, h \in B$. By integrating over U_1 we get the formula (35) for the choice probabilities.

Laura's Choice for Dinner I

In order to understand the primitives of discrete choice models let us turn back to Laura's choice for dinner. Here we take the simplest case where Laura can choose between order a take away meal or cook something by herself. We do not know the exact values of utility she gets from any option, but given a limited knowledge of Laura's preferences we are able to put utilities in a range of possible values, and we assume that they are independently uniformly distributed in their range. We formalize the problem as follows.

Let $B = \{T, 0\}$ be Laura's choice set, and let $U = \langle U_T, U_0 \rangle$ be the vector of utilities associated to the available alternatives. Assume the following distributions for $\langle U_T, U_0 \rangle$:[8]

$$U_T \sim U[0,1]$$
$$U_0 \sim U[\frac{1}{4}, \frac{4}{5}]$$

We want to compute the probability that Laura will choose take away. Let F_j define *cumulative distribution function* and f_j be the *probability density function* for U_j. Assuming that Laura is a (random) utility maximizer means that we need to find the probability that U_T takes higher values than U_0:

$$P(T|B) \equiv \Pr\{U_T \geq U_0\} = \Pr\{U_0 \leq U_T\} = \Pr\{U_0 \leq U_T|U_T\}\Pr\{U_T\} =$$

$$= \int_0^1 F_0(x) f_T(x)\, dx = \int_0^1 \frac{x - 1/4}{4/5 - 1/4}\, dx = \int_0^1 (x - \frac{1}{4})\frac{20}{11}\, dx =$$

$$= \int_0^1 \frac{20}{11} x\, dx - \int_0^1 \frac{5}{11}\, dx = \frac{5}{11}$$

So, the probability that Laura chooses to cook something by herself is $1 - \Pr\{U_T \geq U_0\} = 6/11$. It is easy to verify that solving the problem by searching for the probability to choose "cook at home" gives exactly the same result.

[8] Notice that this is exactly the probability measure μ^B for this problem.

3. Discrete Choice Models

3.1. Standard Set-up. We want to analyse the choice of N consumers who have to choose one alternative over the choice set B that includes $J+1$ alternatives. For each alternative $j = 0, 1, 2, ..., J$, the decision maker, indexed by n, gets utility levels $U_n = [U_{n,0}, U_{n,1}, U_{n,2}, ..., U_{n,J}]$. Alternative $j = 0$ is said to be the composite *outside option*, representing the choice "none of the alternatives ". It must be included in order to account for the fact that choice set has to be *exhaustive.*
By using the same notation as in section 2, define with $s_n = [s_{n,1}, s_{n,2}, ..., s_{n,L}]$ the vector of attributes for the nth consumer, and let $S = [s_1, s_2, ..., s_N]$.
We are not able to observe the actual levels of utility consumers get from alternatives; thus they are unknown from our point of view. What we observe is instead a set of alternative attributes which we assume enter agents' decision problem. We denote *observed* (or *measured*) attributes by the vector $x_j = [x_1, x_2, ..., x_K]$ for all of the $J+1$ alternatives, and we can build the $(J+1) \times K$ matrix of observable products' characteristics $\mathbf{x} = [x_0, x_1, x_2, ..., x_J]' \in X$.
In addition to measured products' characteristics, there are unobserved (by the researcher) factors affecting consumers' utility. We recognize the existence of such factors, as well as the impossibility to measure them, and we take them as random components of decision makers' utility. We formalize this fact as follows.

Define the function $v : X \times S \to \mathbb{R}$ as the *representative (or mean)* utility function. For each individual n and product j, there exists a real valued function $v_{j,n} \equiv v(x_j, s_n)$. The difference between true utility level and mean utility is the random component $\epsilon_{j,n}$:

$$\epsilon_{j,n} \equiv U_{j,n} - v_{j,n} \tag{36}$$

Without loss of generality we can express $U_{j,n}$ as

$$U_{j,n} = v_{j,n} + \epsilon_{j,n} \tag{37}$$

and given (4), we can define the (random) vector of utilities for agent n as

$$U_n = [v_{0,n} + \epsilon_{0,n}, v_{1,n} + \epsilon_{1,n}, v_{2,n} + \epsilon_{2,n}, ..., v_{J,n} + \epsilon_{J,n}]$$

According to RUM, individual n's choice probability for any product $j \in B$ is given by

$$P(j|B, s_n) = \Pr\{v_{j,n} + \epsilon_{j,n} \geq v_{k,n} + \epsilon_{k,n} \ \forall k \in B\} \tag{38}$$

Two considerations about equation (38). First, since $v_{.,n}$ is deterministic, choice probabilities depend on the distribution of the unobserved portion of demand $\epsilon_{.,s}$, that is on the distribution of the random vector $\epsilon_n = \langle \epsilon_{0,s}, \epsilon_{1,s}, \epsilon_{2,s}, ..., \epsilon_{J,s} \rangle$ conditional on individual characteristics s_n of agent n.
Second, only differences in *representative* utility matter. Indeed, we can rewrite (5) as

$$P(j|B, s_n) = \Pr\{\epsilon_{k,n} - \epsilon_{j,n} \leq v_{j,n} - v_{k,n} \ \forall k \in B\}$$

and we notice that is the difference between $v_{.,n}$'s that enter in the argument of the probability. The following example should clarify this aspect.

Laura's Choice for Dinner II

Let us come back to Laura's choice between take-away vs. home-cooked dinner. Here we add the decomposition of Laura's utility function into the two components introduced in previous section.
We have the usual choice set $B = \{T, 0\}$, and a vector of utility values for Laura given by

$$U = [U_T, U_0] = [v_T + \epsilon_T, v_0 + \epsilon_0]$$

Suppose, for simplicity, that ϵ's are independently distributed. The probability to choose take-away is given by

$$P(T|B) = \Pr\{v_T + \epsilon_T \geq v_0 + \epsilon_0\} = \Pr\{\epsilon_0 \leq v_T - v_0 + \epsilon_T\} =$$

$$= \int_{\epsilon_T} F_{\epsilon_0}(v_T - v_0 + x) f_{\epsilon_T}(x)\, dx$$

Where F_{ϵ_0} is the c.d.f. for ϵ_0 and f_{ϵ_T} is the p.d.f. for ϵ_T. Thus, the difference between *mean* utilities for the two alternatives affects Laura's choice probability via the F_{ϵ_0}, i.e. through the distribution of the unobserved portion of utility for the alternative "home-cooked dinner".

Assume ϵ are i.i.d. uniformly distributed in $[0,1]$ and that $v_T - v_0 = 1/8$. Then Laura's probability to choose take-away is given by

$$P(T|B) = \int_0^1 (v_T - v_0 + x)\, dx = v_T - v_0 + \frac{1}{2} = \frac{1}{8} + \frac{1}{2} = \frac{5}{8}$$

The distribution of the unobserved component of utility represents taste variations among decision makers that possess the same attributes. Such distribution must yield probability choices that are computationally feasible. As shown in equation (34) the general formula for computing choice probabilities involves a multiple integral that might not have a solution.

3.2. Logit Model. The Logit model is by far the easiest and most used discrete choice model. In this model, choice probabilities take a simple form which are readily interpretable. Moreover, as we will see, choice probabilities involve the *Independence of Irrelevant Alternatives* (IIA) property: the probability to choose one alternative, relative to another one, does not depend on any other alternative in the choice set. We will see the implications of such assumption, as well as its limitations. To derive choice probabilities in this model we are going to use the approach so far developed. We will start from the distribution of the vector of unobserved utility, and derive choice probabilities consistent with the RUM hypothesis. Given the specific distribution of the Logit model, the integral (1) is easily solvable, and choice probability will take a closed form.

In Logit model, $\epsilon_{j,n}$'s are i.i.d. type I extreme value random variables. That is, for every j

$$F_{\epsilon_{j,n}}(t) = F_\epsilon(t) = e^{-e^{-t}} \qquad f_{\epsilon_{j,n}}(t) = f_\epsilon(t) = e^{-t}e^{-e^{-t}} \tag{39}$$

Given the distribution for the unobserved portion of utility, we have all the ingredient to find the choice probability of agent n with individual characteristics s_n. In particular, the probability to choose alternative $j \in B$ over the $J+1$ alternatives is given by

$$P(j|B, s_n) = \Pr\{v_{j,n} + \epsilon_{j,n} \geq v_{k,n} + \epsilon_{k,n} \ \ \forall k \in B\} =$$

$$= \Pr\{\epsilon_{k,n} \leq v_{j,n} - v_{k,n} + \epsilon_{j,n} \forall k \in B | \epsilon_{j,n}\} \Pr\{\epsilon_{j,n}\} =$$

$$= \int \prod_{k \neq j} F_\epsilon(v_{j,n} - v_{k,n} + t) f_\epsilon(t)\, dt \tag{40}$$

By assuming that ϵ is distributed according to (39), we get our closed form expression for choice probability:[9]

$$P(j|B, s_n) = \frac{e^{v_{j,n}}}{\sum_{k \in B} e^{v_{k,n}}} \tag{41}$$

It is easy to verify that choice probabilities sum to one. Moreover, they exhibit IIA property: the probability to choose one alternative relative to another is not affected by all the other un-chosen alternatives.

Laura's Choice for Dinner III

Consider the usual decision problem for Laura. The main setting is the same: Laura has to choose between two alternatives, i.e. $B = \{T, 0\}$, with vector of utilities associated to the alternatives

$$U = [U_T, U_0] = [v_T + \epsilon_T, v_0 + \epsilon_0]$$

Since now we know that only difference in mean utilities matters, without loss of generality we can normalize $v_0 = 0$. Moreover, we assume that the ϵ's are i.i.d extreme value type I random variables as in the Logit model. In this setting, the probability choices are given by

$$P(T|B) = \frac{e^{v_T}}{1 + e^{v_T}} \quad P(0|B) = \frac{1}{1 + e^{v_T}}$$

For any two alternatives $j, k \in B$ the ratio of the Logit choice probabilities is

$$\frac{P(j|B, s_n)}{P(k|B, s_n)} = \frac{e^{v_{j,n}}}{e^{v_{k,n}}} = e^{v_{j,n} - v_{k,n}} \tag{42}$$

This ratio is not affected by any alternative other than j and k. That is, the relative probability to choose j over k does not depend on the other available alternatives in the choice set. This means that, if we restrict the choice set, by eliminating some

[9] You can find the algebra in Train [2009].

(irrelevant) alternatives, this will not effect the relative choice probability of alternative j over k.
While the IIA property is realistic in some choice situations, it is clearly inappropriate in others. The famous *red-bus,blue-bus* problem should clarify this aspect.

Red-Bus Blue-Bus Problem

Consider the following problem. A traveller has to choose between taking the car or a blue-bus to go to work. Let P_c and P_{bb} be the choice probabilities to take the car and the blue-bus respectively. Assume that

$$P_c = P_{bb} = \frac{1}{2} \Longrightarrow \frac{P_c}{P_{bb}} = 1$$

Now suppose that another alternative, the red-bus, is introduced in the choice set of our traveller and let P_{rb} be the choice probability for the red-bus. Suppose that the traveller is indifferent between taking the blue or the red bus. Hence we must have

$$\frac{P_{bb}}{P_{rb}} = 1$$

However notice that IIA property of Logit model implies that $P_c/P_{bb} = 1$ no matter what. The only distribution that satisfies these two conditions is

$$P_c = P_{bb} = P_{rb} = \frac{1}{3}$$

However, one would have expected to have

$$P_c = \frac{1}{2}, \quad P_{bb} = P_{rb} = \frac{1}{4}$$

That is, one would expect to observe similar choice probabilities for similar alternatives. In this sense, IIA might lead to unreasonable substitution patterns, especially when subsets of alternatives are more similar than other ones.

3.2.1. *Logit Demand Estimation.* Imagine to observe market-level data for some industry characterized by differentiated products. We have units sold for each product in the market, their prices, and measurable products' characteristics. We are interested in estimating aggregate demand in this market.

The starting point for demand estimation consists in specifying consumers' utility function, and then aggregate preferences through the unobserved portion of consumers' utility.[10]

Suppose in our industry there are $J + 1$ differentiated products, where option 0 states for the option "none of the available products". By choosing this option, the consumer decides to not purchase any of the available alternatives in the market. For each product j, we have a vector of observed product-characteristics x_j as defined before. .

We have N consumers who choose one product among the available alternatives. We define N as the *size* of this market. A simple specification for consumers utility takes the following form:

$$U_{j,n} = v_j + \epsilon_{j,n} = x_j\beta - \alpha p_j + \xi_j + \epsilon_{j,n} \tag{43}$$

where p_j is the price for product j, x_j is the vector of measurable attributes for j, and ξ_j is a mean-zero random component that includes all unobservable (by the researcher) product characteristics that enter in the decision problem of the consumer. Since, in this specification, the parameters to be estimated (β, α) are assumed to be same across consumers, then the only source of heterogeneity in preferences is given by $\epsilon_{j,n}$. In this case, the unobserved portion of utility describes the distribution of *random* tastes among the population of consumers.

In this setting, the share of people purchasing product j consists of the share of customers for which the utility from product j is higher than the one from any other available alternative in the market. Let $\epsilon_{j,n}$ be i.i.d extreme value I distributed. Then we know that the *predicted* share of consumers purchasing product j, named $\hat{s}_j$ will be

$$\hat{s}_j(\beta, \alpha) = \frac{e^{x_j\beta - \alpha p_j + \xi_j}}{\sum_{k=0}^{J} e^{x_k\beta - \alpha p_k + \xi_k}}$$

Since only difference in utilities matters, we can normalize the representative utility of the outside option to zero. By doing this we get the usual predicted market share

[10] I use the same specification as Berry et al. [1995].

for good j

$$\hat{s}_j(\beta,\alpha) = \frac{e^{x_j\beta-\alpha p_j+\xi_j}}{1+\sum_{k=1}^{J} e^{x_k\beta-\alpha p_k+\xi_k}}$$

Let us turn back to market level data. If we observe total quantity purchased of product j, q_j, then we are able to derive *observed* market shares[11]

$$s_j = \frac{q_j}{N} \;\; j \neq 0, \quad s_0 = 1 - \sum_j s_j$$

Finally, we get the $J+1$ moment conditions for demand estimation:

$$E[\mathbf{s} - \hat{\mathbf{s}}(\beta,\alpha)] = \mathbf{0} \tag{44}$$

A nice feature of the Logit demand model is that it can be easily linearised. Notice that, for all $j \neq 0$

$$\frac{\hat{s}_j}{\hat{s}_0} = e^{x_j\beta-\alpha p_j+\xi_j} \quad \Rightarrow \quad \ln(\hat{s}_j) - \ln(\hat{s}_0) \equiv \hat{\delta}_j = x_j\beta - \alpha p_j + \xi_j$$

And we get the following J moment conditions

$$E[\delta - \hat{\delta}(\beta,\alpha)] = \mathbf{0} \tag{45}$$

Given the linearity of the model, the solution for (12) is the standard OLS estimator for (β,α).

3.2.2. *Logit Elasticities.* An elasticity is the percentage change in one variable that is associated to a one percent change in another variable. Usually we are interested in the elasticity of demand for product j with respect to its own price, or other product prices. Define with $\eta_{j,k}$ the (cross-price) elasticity of product j with respect to the price of product k as follows:

$$\eta_{j,k} = \frac{\partial s_j}{\partial p_k}\frac{p_k}{s_j} \tag{46}$$

In logit demand model, elasticities take the following simple forms:

$$\eta_{j,j} = -\alpha p_j(1-s_j) \tag{47}$$

$$\eta_{j,k} = -\alpha p_k s_k \quad \forall k \neq j \tag{48}$$

[11]Note that here we use the fact that consumers can purchase just one unit of the product.

We notice that own-price elasticity for product j depends solely upon price and market share of good j, while cross-price elasticities of any good $j \neq k$ with respect to the price of product k depends only upon price and market share of product k. These formula are directly a consequence of the IIA assumption underlying the Logit model, which assumes that ϵ's are i.i.d. type I across products.

Consider cross-price elasticity $\eta_{j,k}$. Equation (16) implies that the effect of a price increase of product k affects the sales of all the other products different than k by the same amount. However, one would expect to get different responses, depending on the degree of differentiation between the goods considered. For instance, if product i is more similar to k than j in terms of product characteristics, one would expect

$$|\eta_{i,k}| > |\eta_{j,k}|$$

However the model predict that they are equal. This is an important limitation of Logit demand model: substitution patterns are very poor, and they do not discriminate products across their relative degree of substitutability. Nested-Logit model solves, to some extent, this issue, allowing for richer substitution patterns among alternatives in the choice set.

3.3. Nested-Logit. Nested-Logit model relaxes, to some extent, the IIA assumption of the Logit model. It allows for richer substitution patterns by identifying alternatives that are closer to each other and including them in subsets of the choice set, called *nests*.

When implementing Nested-Logit model we move in two directions. First, we identify alternatives in the choice set that have similar features, i.e. we identify observable characteristics that are common to some of the available alternatives, and we include them in a nest. By doing this, we essentially *partition* the choice set into nests.

Second, we structure the cognitive decision process of the agent. In particular, we assume that the decision process follows some predetermined order across nests and alternatives. The decision maker is assumed to choose first the nest (and eventually sub-nest) and then one alternative from the chosen nest.

It is important to stress the fact that this model does not involve a two-stage decision process; that is, the agent does not effectively choose first the nest and than one alternative in the nest. What we are modelling here is the cognitive decision process: the decision maker chooses instantly her best alternative, but she follows the process described. The following example should clarify what we mean by that.

Laura's Choice for Dinner IV

Let us turn back to Laura's choice for dinner. Consider the very first example, where the choice set was $B = \{S, P, T, 0\}$ where S states for sushi, P for pizza, T for thai, and 0 is the composite outside good which means "cook something by herself at home".
In Nested-Logit model, we assume two-stage decision process for Laura. First, she chooses a nest, that is a subset of alternatives, and then one alternative over the nest. In this problem we can group alternatives in the following way

- Let $B_1 = \{S, P, T\}$ be the "take-away" nest. This include all the option that need to be paid by Laura, and that involve delivery time.
- Let $B_0 = \{0\}$ be the "cook-at-home" nest. Given our choice set this nest is a singleton, however it might not be the case.

Given the partition for the choice set, we impose the following decision process for Laura:

(1) First choose one (and only one) nest:"Should I cook or order something?"
(2) Then choose one (and only one) alternative from the nest selected:"sushi, pizza, or thai?"

In this framework, we can decompose alternatives choice probabilities into two components: the probability to choose the specific alternative given previous nest decision; and the probability to choose exactly that nest. In short, we can express the probability to choose thai as follows:

$$P(T|B) = \Pr\{T|B_1\}\Pr\{B_1\}$$

Where $\Pr\{T|B_1\}$ is the probability to choose thai, given the choice of nest B_1, and $\Pr\{B_1\}$ is the probability to choose nest B_1.

The first step for Nested-Logit model is to partition the choice set into nests. Let B be the usual choice set of $J+1$ alternatives, and let $\{B_0, B_1, ..., B_K\}$ be a partition for B. The partition is specified by the researcher depending on the decision problem she needs to analyse.[12]

The second step involves the distribution of the vector of the unobserved component of demand $\epsilon = \langle \epsilon_0, \epsilon_1, \epsilon_2, ..., \epsilon_J \rangle$.[13] In Nested-Logit model, the vector ϵ is distributed according to the following c.d.f.:

$$F(\epsilon) = \exp[-\sum_{k=0}^{K} (\sum_{j \in B_k} e^{-\dfrac{\epsilon_j}{\lambda_k}})^{\lambda_k}] \tag{49}$$

Given the multivariate distribution (49) we can compute choice probabilities consistent with RUM hypothesis by using (2). Since in this case the solution involves tedious algebraic manipulation, we go directly to the final formula for the choice probability.[14]

The Nested-Logit probability to choose alternative $j \in B_k \subset B$ is

$$P(j|B, \{B_0, B_1, ..., B_K\}, s_n) \equiv P_{j,n} = \frac{e^{v_{j,n}/\lambda_k} (\sum_{i \in B_k} e^{v_{i.n}/\lambda_k})^{\lambda_k - 1}}{\sum_{l=0}^{K} (\sum_{i \in B_l} e^{v_{i,n}/\lambda_l})^{\lambda_l}} \tag{50}$$

Notice that for $\lambda_k = 1$ for all k, Nested-Logit probability choices are exactly equal to the Logit ones. Indeed, the parameter λ_k is and indicator of the degree of correlation among alternatives that belong to the nest B_k: $\lambda_k = 1$ indicates no correlation among alternatives in nest B_k, and choice probabilities become simply Logit. As λ_k approaches zero the degree of correlation among alternatives in B_k increases.

We can express choice probabilities in a more intuitive way. Let $v_{j,n} = w_{k,n} + y_{j,n}$ where

- $w_{k,n}$ includes across-nest variations in representative utility, i.e. all those observable characteristics that do not vary within the same nest.

[12] Clearly, the standard properties for a partition of a set need to be satisfied. In particular, $B_l \cap B_h = \emptyset$ and the union of all the subsets $B_0, B_1, ..., B_K$ is equal to B.

[13] We eliminate the subscript n for the time being with the understanding that the distribution is conditioned to decision maker's characteristics s_n.

[14] You can find the algebra in the Appendix.

- $y_{j,n}$ includes within-nest variations in representative utility, that is all observable characteristics that vary among alternatives belonging to the same nest.

Then, we can express choice probabilities in the following way:[15]

$$P_{j,n} = \frac{e^{y_{j,n}/\lambda_k}}{\sum_{j \in B_k} e^{y_{j,n}/\lambda_k}} \frac{e^{w_{k,n}+\lambda_k I_k}}{\sum_{l=1}^{K} e^{w_{l,n}+\lambda_l I_{l,n}}} = P_{j,n|B_k} \times P_{B_k,n} \tag{51}$$

Where $I_{k,n} = \ln \sum_{j \in B_k} e^{y_{j,n}/\lambda_k}$ is called *inclusive value*: it is the expected surplus from nest k.

Nested-Logit model yields substitution patterns with the following two properties:

(1) IIA holds within nests, i.e. the relative choice probability of the alternatives in the same nest is independent from any other alternative in the choice set.

(2) IIA does not hold across nests, i.e. the relative likelihood to choose one alternative from another belonging in a different nest can depend on attributes of other alternatives in the two nests.

3.3.1. *Nested-Logit Demand.* We use the set-up described in previous section and apply it to the Nested-Logit model. We observe market level data (units sold for each product in the market, prices, measurable product characteristics) and we assume that consumers can choose only one of the available goods in the market. Suppose to have $J + 1$ alternatives in the market, and recall that option zero is the outside option. Each product belongs to a nest B_k. Assume that there are $K + 1$ nests, and that $B_0 = \{0\}$. For each $j \neq 0$ we have a vector of measurable product attributes x_j that enters in the decision process of the N consumers via their utility function. We use the same utility specification as in (11)

$$U_{j,n} = x_j\beta - \alpha p_j + \xi_j + \epsilon_{j,n}.$$

We assume that now $\epsilon_{j,n}$ are distributed according to the Nested-Logit distribution (17) with $\lambda_k = \lambda$ for all k, and we normalize the mean utility from the outside option to zero. In this set-up, predicted market share for product $j \in B_k$ is given by

$$\hat{s}_j(\beta, \alpha) = \frac{e^{(x_j\beta - \alpha p_j + \xi_j)/\lambda} \left(\sum_{i \in B_k} e^{(x_i\beta - \alpha p_i + \xi_i)/\lambda}\right)^{\lambda - 1}}{\sum_{l=0}^{K} \left(\sum_{i \in B_l} e^{(x_i\beta - \alpha p_i + \xi_i)/\lambda}\right)^{\lambda}} \tag{52}$$

[15]Proof in the Appendix.

Given the special structure of the decision process implied by Nested-Logit model, we might want to decompose predicted market shares as we did in (19):

$$\hat{s}_j(\alpha, \beta) = \hat{s}_{j|B_k} \times \hat{s}_{B_k}$$

with

$$\hat{s}_{j|B_k} = \frac{e^{(x_j\beta - \alpha p_j + \xi_j)/\lambda}}{\sum_{j \in B_k} e^{(x_j\beta - \alpha p_j + \xi_j)/\lambda}}$$

$$\hat{s}_{B_k} = \frac{e^{w_k + \lambda I_k}}{\sum_{l=1}^{K} e^{w_l + \lambda I_l}}$$

Where $I_l = \ln \sum_{j \in B_k} e^{(x_j\beta - \alpha p_j + \xi_j)/\lambda_k}$ is the *inclusive* value, and w_k include observable variations across nests; $\hat{s}_{B_k}$ is the predicted share for nest B_k, while $\hat{s}_{j|B_k}$ is the predicted share of product j given the choice of nest B_k.

Given market-level data we are able to compute *observed* market shares for alternatives given market size N:

$$s_j = \frac{q_j}{N} \;\; j \neq 0, \quad s_0 = 1 - \sum_j s_j$$

Moreover:

$$s_{B_k} = \frac{\sum_{j \in B_k} q_j}{N} \quad s_{j|B_k} = \frac{q_j}{\sum_{i \in B_k} q_i}$$

And again, here we are interested in find the parameters that solve the system of $J + 1$ equations

$$E[\mathbf{s} - \hat{\mathbf{s}}(\beta, \alpha)] = 0$$

A nice feature of the nested-logit model is that it can be linearised, taking the following form:[16]

$$\ln(\hat{s}_j) - \ln(\hat{s}_0) \equiv \hat{\delta} = X_j\beta - \alpha p_j + (1 - \lambda)\ln(s_{j|B_k}) + \xi_j \tag{53}$$

And we get the J moment conditions to estimate the parameters of interest

$$E[\delta - \hat{\delta}(\beta, \alpha)] = 0$$

Clearly also this model is affected by endogeneity problem, as the one in (13), and for the same sorts of reasoning. Hence we need exogenous and relevant instruments in order to consistently estimate (β, α).

[16] We show the algebra involved in the Appendix.

Given our specification for the utility, Nested-Logit model implies the following own and cross-price elasticity for demand of product $j \in B_k$:

- For $j \in B_k$ *and* $i \in B_l$

$$\eta_{j,i} = \alpha p_i s_{B_l} s_{i|B_l} \tag{54}$$

- For $i, j \in B_k$

$$\eta_{j,i} = \alpha p_i s_{i|B_k} (\frac{1}{\lambda} - (1 - s_{B_k})) \tag{55}$$

- Own price elasticity for some $j \in B_k$ is given by

$$\eta_{j,j} = -\alpha \frac{1}{\lambda} p_j (\frac{1}{\lambda}(1 - s_{j|B_k}) + (1 - s_{B_k}) s_{j|B_k}) \tag{56}$$

It is important to recognize the richer substitution patterns implied by Nested-Logit model with respect to the Logit model. By comparing (22) and (23) we observe that, if prices and market shares are equal, than the cross-price elasticity between two goods in the same nest is higher than the cross-price elasticity between two alternatives in different nests. The model captures differences in substitutability among products, however this crucially depends on how the researcher specifies the nests. These might be a problem in many applications, since in principle there might be many suitable partitions of the choice sets each of them leading different results and interpretations.

Laura's Choice for Dinner V

Suppose now that Laura needs to choose between Thai or Pizza, but she can also decide whether to make them or take away. In this situation we have four alternatives in the choice set $C = \{P_H, P_T, T_H, T_T\}$, where subscripts mean home (H) and take away (T). Nested-Logit model implies a partition for C, and it is easy to see that in this case we have two very intuitive possible partitions: $C_1 = \{\{P_H, P_T\}\{T_H, T_T\}\}$ and $C_2 = \{\{P_H, T_H\}\{P_T, T_T\}\}$. The former, implies that Laura chooses first if she prefer pizza or thai, and then whether she wants to take away or to cook. The second partition refers to a different decision process: Laura decides first whether she wants to cook or take away, and then one of the two alternatives.

Both possibilities are reasonable, however they lead to different results. Suppose we select C_1, then the probability to choose "cook-pizza" is given by

$$P(P_H|C_1) = P(P) \times P(H|P) = \frac{e^{I_P}}{e^{I_P} + e^{I_T}} \times \frac{e^{\delta_{P_H}/\lambda}}{e^{\delta_{P_H}/\lambda} + e^{\delta_{P_T}/\lambda}}$$

Where I_P and I_T are the inclusive values for pizza and thai respectively; δ is the mean utility, and λ measure the degree of correlation among alternatives in the same nest.

The same choice probability consistent with partition C_2 would be:

$$P(P_H|C_2) = P(H) \times P(P|H) = \frac{e^{I_H}}{e^{I_H} + e^{I_T}} \times \frac{e^{\delta_{P_H}/\lambda}}{e^{\delta_{P_H}/\lambda} + e^{\delta_{T_H}/\lambda}}$$

Hence $P(P_H|C_1) \neq P(P_H|C_2)$.

4. Endogeneity

Consider the linear econometric model from Logit demand estimation

$$\ln(s_j) - \ln(s_0) = x_j\beta - \alpha p_j + \xi_j \tag{57}$$

The mean-zero error ξ_j represents unobserved (by the researcher) product characteristics: it reflects the difficulty of quantifying aspects of style, reputation, prestige, that affect the demand for different products. Consumers internalize these features when making their purchasing decision: for a given (p_j, x_j) we should expect higher demand for alternatives displaying higher ξ_j, as consumers utility rises.

On the supply side, firms set prices as a function of marginal cost and demand. Following Villas-Boas and Winer [1999], one can envision observed prices in the market as a result of firms profit maximization. In this setting, optimal prices are function of marginal cost of production and demand features.

Monopolist Problem and Linear Demand

Consider the following example. A monopolist faces a linear downward-sloping demand curve $D(p) = a - p$, and has a constant marginal cost of production $c < a$. Optimal prices comes from the usual profit maximization problem

$$\max_p \pi(p) = (p - c)(a - p)$$

with solution $p^* = \frac{a+c}{2}$.
Now suppose $a = \tilde{a} + \mu$, where μ represents demand shocks not observable by the researcher. Then $p^* = \frac{a+c+\mu}{2}$, hence μ affects positively firm's optimal price.

In general, we might want to describe optimal prices as a parametric function of marginal cost and demand shock as

$$p_j^* = p(c_j, \mu_j; \theta)$$

where μ_j represents unobservable demand shocks. Hence μ_j is correlated with ξ_j, and $E(p_j \xi_j) \neq 0$. Unobserved product's characteristics affect positively demand and prices, hence the correlation between the endogenous covariate in (25) and the error term is positive. As such, endogeneity biases the price coefficient towards zero.

Appropriate instrumental variables are needed in order to consistently estimate the price coefficient. The Instrument needs to be relevant, i.e. it must be correlated with the endogenous covariate, and it must be valid, i.e. orthogonal to the error term. Since the price coefficient is usually the main parameter of interest in demand estimation, most of the effort has to be devoted to find relevant and valid instruments for price.
A class of instrument that has been used in the literature includes cost-shifters, such as price of raw materials, firm's operating costs (e.g. wage), or exogenous shock on the supply side. These factors are correlated with prices via marginal costs, but are not directly correlated with demand. Unfortunately, such data might not be available, and other more sophisticated identification strategies are often needed. Another class of instrumental variables that has been used in the literature consists in rival product characteristics (Berry et al. [1995],Verboven [1996]). Identification is based on the assumption that firms compete in a Bertrand-Nash game. Products with similar characteristics are more substitute, hence they have lower mark-ups. Therefore, the higher the similarity in attributes between products, the lower prices will be. Rival product characteristics approximate the competitive pressure among alternatives in the market place. This justifies the assumption that they are orthogonal to the unobservable term of the econometric specification, we ensure the

validity of such instruments.

Another widely used identification strategy consists of making use of product prices in other markets as instruments for local prices (Hausman [1996], Nevo [2001]). This approach can be implemented when data cross-section is particularly rich, and many local market are observed at each point in time. The validity of the instrument is based on the assumption that foreign prices do not affect local demand, which means that consumers in one particular local market do not consider foreign prices when making their purchasing decision. A proper specification of individuals choice sets in each market is necessary to validate exogeneity of the instrument. Relevance comes from the usual correlation via common marginal cost of production.

Appendix

Deriving Choice Probabilities in Nested-Logit Model. In Nested-Logit model, the vector ϵ is distributed according to the following c.d.f.:

$$F(\epsilon) = \exp[-\sum_{k=1}^{K}(\sum_{j\in B_k} e^{-\frac{\epsilon_j}{\lambda_k}})^{\lambda_k}]$$

We use the procedure developed in first section in order to find probability choice from this distribution. First, we note that the vector U is distributed according to F:

$$F(\epsilon) = \Pr(\epsilon_j \leq t_j,\ \forall j) = \Pr(v_j + \epsilon_j \leq v_j + t_j,\ \forall j) = \Pr(U \leq u) = F(u)$$

Now, without loss of generality, we focus on alternative i_1:

$$F_1(u) = \frac{\partial F}{\partial u_1} = \exp(-\sum_{k=1}^{K}(\sum_{j\in B_k} e^{-(u_j/\lambda_k)})^{\lambda_k})(\sum_{j\in B_1} e^{-(u_j/\lambda_1)})^{\lambda_1 - 1} e^{-(u_1/\lambda_1)} \tag{58}$$

This has to be evaluated at u_1. This means that $v_j + \epsilon_j = v_1 + \epsilon_1\ \forall j$. Then $\epsilon_j = v_1 - v_j + \epsilon_1\ \forall j$:

$$F_1(u) = \exp(-\sum_{k=1}^{K}(\sum_{j\in B_k} e^{-((v_1 - v_j + \epsilon_1)/\lambda_k)})^{\lambda_k})(\sum_{j\in B_1} e^{-((v_1 - v_j+\epsilon_1)/\lambda_1)})^{\lambda_1 - 1} e^{-((v_1+\epsilon_1)/\lambda_1)} \tag{59}$$

We can rewrite it as follows:

$$F_1(u) = \exp(-\sum_{k=1}^{K}(e^{-(\epsilon_1/\lambda_k)}\sum_{j\in B_k} e^{-((v_1 - v_j +)/\lambda_k)})^{\lambda_k})(\sum_{j\in B_1} e^{-((v_1 - v_j)/\lambda_1)})^{\lambda_1 - 1} e^{-(\epsilon_1)}$$

$$F_1(u) = \exp(-e^{-\epsilon_1}\sum_{k=1}^{K}(\sum_{j\in B_k} e^{-((v_1 - v_j +)/\lambda_k)})^{\lambda_k})(\sum_{j\in B_1} e^{-((v_1 - v_j)/\lambda_1)})^{\lambda_1 - 1} e^{-(\epsilon_1)} \tag{60}$$

Then the probability of choosing alternative i_1 is given by:

$$P_1 \equiv P(i_1|B,s) = \int_{U_1=-\infty}^{+\infty} F_1 d\epsilon_1 \tag{61}$$

In order to solve the integral define $x = e^{-\epsilon_1}$. Notice that as $\epsilon_1 \to \infty,\ x \to 0$, and as $\epsilon_1 \to -\infty,\ x \to \infty$. Moreover $-dx = e^{-\epsilon_1}d\epsilon_1$. By substituting in the integral we get:

$$P_1 = \int_0^{\infty} \exp(-x \sum_{k=1}^{K} (\sum_{j \in B_k} e^{-((v_1 - v_j)/\lambda_k)})^{\lambda_k}) (\sum_{j \in B_1} e^{-((v_1 - v_j)/\lambda_1)})^{\lambda_1 - 1} dx =$$

$$-[\frac{\exp(-x \sum_{k=1}^{K} \sum_{j \in B_k} e^{-((v_1 - v_j +)/\lambda_k)})^{\lambda_k}) (\sum_{j \in B_1} e^{-((v_1 - v_j)/\lambda_1)})^{\lambda_1 - 1}}{\sum_{k=1}^{K} (\sum_{j \in B_k} e^{-((v_1 - v_j +)/\lambda_k)})^{\lambda_k}}]_0^{\infty} =$$

$$= \frac{(\sum_{j \in B_1} e^{-((v_1 - v_j)/\lambda_1)})^{\lambda_1 - 1}}{\sum_{k=1}^{K} (\sum_{j \in B_k} e^{-((v_1 - v_j +)/\lambda_k)})^{\lambda_k}} = \frac{(e^{-v_1})^{1-(1/\lambda_1)} (\sum_{j \in B_1} e^{v_j/\lambda_1})^{\lambda_1 - 1}}{e^{-v_1} \sum_{k=1}^{K} (\sum_{j \in B_k} e^{v_j/\lambda_k})^{\lambda_k}}$$

$$P_1 = \frac{e^{v_1/\lambda_1} (\sum_{j \in B_1} e^{v_j/\lambda_1})^{\lambda_1 - 1}}{\sum_{k=1}^{K} (\sum_{j \in B_1} e^{v_j/\lambda_k})^{\lambda_k}} \tag{62}$$

$$= \frac{e^{v_j/\lambda_k}}{\sum_{j \in B_k} e^{v_j/\lambda_k}} \frac{(\sum_{j \in B_k} e^{v_j/\lambda_k})^{\lambda_k}}{\sum_{k=1}^{K} (\sum_{j \in B_k} e^{v_j/\lambda_k})^{\lambda_k}} =$$

$$= \frac{e^{(w_k + y_j)/\lambda_k}}{\sum_{j \in B_k} e^{(w_k + y_j)/\lambda_k}} \frac{(\sum_{j \in B_k} e^{(w_k + y_j)/\lambda_k})^{\lambda_k}}{\sum_{k=1}^{K} (\sum_{j \in B_k} e^{(w_k + y_j)/\lambda_k})^{\lambda_k}} =$$

$$= \frac{e^{w_k/\lambda_k} e^{y_j/\lambda_k}}{e^{w_k/\lambda_k} \sum_{j \in B_k} e^{y_j/\lambda_k}} \frac{e^{w_k} (\sum_{j \in B_k} e^{+y_j/\lambda_k})^{\lambda_k}}{\sum_{k=1}^{K} e^{w_k} (\sum_{j \in B_k} e^{y_j/\lambda_k})^{\lambda_k}} =$$

$$= \frac{e^{y_j/\lambda_k}}{\sum_{j \in B_k} e^{y_j/\lambda_k}} \frac{e^{w_k + \lambda_k I_k}}{\sum_{k=1}^{K} e^{w_k + \lambda_k I_k}} = P_{j|B_k} \times P_{B_k} \tag{63}$$

Where $I_k = \ln \sum_{j \in B_k} e^{y_j/\lambda_k}$. Last equality uses the fact that $x^a = e^{a \ln x}$.

Linearising Nested-Logit Model. We want to linearise the Nested-Logit model. To do so, suppose that there are $J+1$ alternatives and $K+1$ nests. There exists an outside option $j = 0$ in nest B_0, hence it is a singleton. Moreover, we normalise the mean utility of the outside option equal to zero.

In this framework choice probabilities are given by

$$P_j = \frac{e^{y_j/\lambda_k}}{\sum_{j \in B_k} e^{y_j/\lambda_k}} \frac{e^{w_k + \lambda_k I_k}}{1 + \sum_{l=1}^{K} e^{w_l + \lambda_l I_l}} = P_{j|B_k} \times P_{B_k} \tag{64}$$

$$P_0 = P_{0|B_0} P_{B_0} = p_{B_0} = \frac{1}{1 + \sum_{l=1}^{K} e^{w_l + \lambda_l I_l}} \tag{65}$$

By taking the natural logarithm of (31) and (32) we get

$$\ln(P_j) = \frac{y_j}{\lambda_k} - I_k(1-\lambda_k) + w_k - \ln(1 + \sum_{l=1}^{K} e^{w_l + \lambda_l I_l}) \tag{66}$$

$$\ln(P_0) = -\ln(1 + \sum_{l=1}^{K} e^{w_l + \lambda_l I_l}) \tag{67}$$

By taking the difference between (33) and (34) we get

$$\ln(P_j) - \ln(P_0) = \frac{y_j}{\lambda_k} - I_k(1-\lambda_k) + w_k \tag{68}$$

Notice that we can express $I_k = y_j/\lambda_k - \ln(P_{j|B_k})$, therefore by substituting in (29)

$$\ln(P_j) - \ln(P_0) = y_j + w_k + (1-\lambda_k)\ln(P_{j|B_k}) \tag{69}$$

Recall that $y_j + w_k = v_j(s) = X_j\beta - \alpha p_j + \xi_j$ hence we have a linear-in-parameters regression, and standard instrumental variables techniques can be applied. Hence we get the following linear model where standard instrument techniques can be applied:

$$\ln(s_j) - \ln(s_0) = X_j\beta - \alpha p_j + (1-\lambda_k)\ln(s_{j|B_k}) + \xi_j$$

Deriving Own/Cross-Price Elasticities in Nested-Logit Model. In this section we want to derive own and cross price elasticities according to the expected demand of the Nested-Logit model. For this purpose, let us define $P_i \equiv s_i = s_{B_k} \times s_{i|B_k}$, as the expected demand for good $i \in B_k \subset C$. The first-order derivative of s_i with respect to the price p_j for some $j \in B_k$ is given by

$$\frac{\partial s_i}{\partial p_j} = \frac{\partial s_{i|B_k}}{\partial p_j} s_{B_k} + \frac{\partial s_{B_k}}{\partial p_j} s_{i|B_k} \tag{70}$$

If $j \notin B_k$ then $(\partial s_{i|B_k}/\partial p_j) = 0$ and

$$\frac{\partial s_i}{\partial p_j} = \frac{\partial s_{B_k}}{\partial p_j} s_{i|B_k} \tag{71}$$

We observe that in nested-logit model, price variation for alternative j has two effects on the expected demand for alternative i. It affects the within nest demand as well as the demand for the nest containing alternative i. When alternative j does not belong to the same nest of i, then there is no within-nest effect, and price variation affects the probability to choose the nest containing i. Define the cross-price elasticity of alternative i with respect to the price of j as follows:

$$\eta_{ij} \equiv \frac{\partial s_i}{\partial p_j}\frac{p_j}{s_i} \tag{72}$$

Let us derive the cross price elasticity of (i, j) for $j \notin B_k$ by using (38) and (39):

$$\eta_{ij} = \frac{\partial s_{B_k}}{\partial p_j} s_{i|B_k} \frac{p_j}{s_i} = \frac{\partial s_{B_k}}{\partial p_j} s_{i|B_k} \frac{p_j}{s_{B_k} s_{i|B_k}} = \frac{\partial s_{B_k}}{\partial p_j}\frac{p_j}{s_{B_k}} \tag{73}$$

If $j \in B_k$ then we use (37) and (39):

$$\eta_{ij} = [\frac{\partial s_{i|B_k}}{\partial p_j} s_{B_k} + \frac{\partial s_{B_k}}{\partial p_j} s_{i|B_k}]\frac{p_j}{s_{B_k} s_{i|B_k}} = \frac{\partial s_{i|B_k}}{\partial p_j}\frac{p_j}{s_{i|B_k}} + \frac{\partial s_{B_k}}{\partial p_j}\frac{p_j}{s_{B_k}} \tag{74}$$

Now we need to compute the partial derivatives from the formulas for s_{B_k} and $s_{i|B_k}$ derived in (30):

$$\frac{\partial s_{i|B_k}}{\partial p_j} = -\frac{1}{\lambda_k}\frac{\partial y_j}{\partial p_j}\frac{e^{y_j/\lambda_k} e^{y_i/\lambda_k}}{(\sum_{j \in B_k} e^{y_j/\lambda_k})^2} = -\frac{1}{\lambda_k}\frac{\partial y_j}{\partial p_j} s_{j|B_k} s_{i|B_k} \tag{75}$$

$$\frac{\partial s_{i|B_k}}{\partial p_i} = \frac{1}{\lambda_k}\frac{\partial y_i}{\partial p_i}\frac{e^{y_i/\lambda_k}\sum_{j \in B_k} e^{y_j/\lambda_k} - (e^{y_j/\lambda_k})^2}{(\sum_{j \in B_k} e^{y_j/\lambda_k})^2} = \frac{1}{\lambda_k}\frac{\partial y_i}{\partial p_i} s_{i|B_k}(1 - s_{i|B_k}) \tag{76}$$

In computing $\partial s_{B_k}/\partial p_j$ we must distinguish whether or not alternative j belongs to the same nest as i:

$$\frac{\partial s_{B_k}}{\partial p_j} = \lambda_k \frac{\partial I_k}{\partial p_j}\frac{e^{w_k+\lambda_k I_k}\sum_{l=1}^{K} e^{w_k+\lambda_k I_k} - (e^{w_k+\lambda_k I_k})^2}{(\sum_{l=1}^{K} e^{w_k+\lambda_k I_k})^2} =$$

$$= \lambda_k \frac{\partial I_k}{\partial p_j} s_{B_k}(1 - s_{Bk}) \quad for\ j \in B_k \tag{77}$$

For $j \in B_l$ we have:

$$\frac{\partial s_{B_k}}{\partial p_j} = -\lambda_l \frac{\partial I_l}{\partial p_j}\frac{e^{w_k+\lambda_k I_k} e^{w_l+\lambda_l I_l}}{(\sum_{l=1}^{K} e^{w_l+\lambda_l I_l})^2} = -\lambda_l \frac{\partial I_l}{\partial p_j} s_{B_k} s_{B_l} \tag{78}$$

Now we are able to derive the elasticities for the nested-logit demand model.
For $j \in B_l, i \in B_k$ we use (40) and (45):

$$\eta_{ij} = -\lambda_l \frac{\partial I_l}{\partial p_j} p_j s_{B_l} \tag{79}$$

For $i, j \in B_k$ we have

$$\eta_{i,j} = -\frac{1}{\lambda_k}\frac{\partial y_j}{\partial p_j} p_j s_{j|B_k} + \lambda_k \frac{\partial I_k}{\partial p_j} p_j (1 - s_{Bk}) \tag{80}$$

Own-price elasticity for alternative i is given by:

$$\eta_{ii} = -\frac{1}{\lambda_k}\frac{\partial y_i}{\partial p_i} p_i (1 - s_{i|Bk}) + \lambda_k \frac{\partial I_k}{\partial p_i}(1 - s_{B_k}) \tag{81}$$

Finally, we need to analyze $\partial I_k/\partial p_j$. Given the usual linear specification for the representative utility, $\partial y_j/\partial p_j$ is simply the coefficient associated to price.
Let us recall first that for $k = 1, 2, ..., K$

$$I_k = \ln \sum_{j \in B_k} e^{y_j/\lambda_k}$$

therefore for $j \in B_k$, and $k = 1, 2, ..., K$

$$\frac{\partial I_k}{\partial p_j} = \frac{(1/\lambda_k)(\partial y_j/\partial p_j) e^{y_j/\lambda_k}}{\sum_{j \in B_k} e^{y_j/\lambda_k}} = \frac{1}{\lambda_k}\frac{\partial y_j}{\partial p_j} s_{j|B_k} \tag{82}$$

We have now all the elements to compute the elasticities for the nested-logit demand by substituting (49) onto (46), (47) and (49):

- For $i \in B_k$ *and* $j \in B_l$

$$\eta_{ij} = -p_j s_{B_l} s_{j|B_l} \frac{\partial y_j}{\partial p_j} \tag{83}$$

- For $i, j \in B_k$

$$\eta_{ij} = -p_j \frac{\partial y_j}{\partial p_j} s_{j|B_k} \left(\frac{1}{\lambda_k} - (1 - s_{B_k})\right) \tag{84}$$

- Own price elasticity for some $i \in B_k$ is given by

$$\eta_{ii} = \frac{1}{\lambda_k}\frac{\partial y_i}{\partial p_i} p_i \left(\frac{1}{\lambda_k}(1 - s_{i|B_k}) + (1 - s_{B_k}) s_{i|B_k}\right) \tag{85}$$

Bibliography

O. Ashenfelter and D. Card. Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs. *The Review of Economics and Statistics*, 67(4):648–660, November 1985.

O. C. Ashenfelter, D. S. Hosken, and M. C. Weinberg. The price effects of a large merger of manufacturers: A case study of maytag-whirlpool. *American Economic Journal: Economic Policy*, 5(1):239–61, 2013.

S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

S. T. Berry. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262, 1994.

S. Borenstein and J. S. Netz. Exercising market power in proprietary aftermarkets. *October (at http://haas. berkeley. edu/borenste/aftmkt. pdf)*, 1996.

S. Borenstein, J. K. MacKie-Mason, and J. S. Netz. Antitrust policy in aftermarkets. *Antitrust LJ*, 63:455, 1994.

T. Bresnahan. Comments on 'valuation of new goods under perfect and imperfect competition' by jerry hausman. *The Economics of New Goods, Studies in Income and Wealth*, 58, 1996.

L. Cabral. Aftermarket power and foremarket competition. *International Journal of Industrial Organization*, 35:60–69, 2014.

D. Card. Using regional variation in wages to measure the effects of the federal minimum wage. *ILR Review*, 46(1):22–37, 1992.

D. Card and A. B. Krueger. Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania. *The American Economic Review*, 84(4):772–793, 1994.

D. Card and A. B. Krueger. Minimum wages and employment: a case study of the fast-food industry in new jersey and pennsylvania: reply. *American Economic Review*, 90(5):1397–1420, 2000.

D. W. Carlton and M. Waldman. Competition, monopoly, and aftermarkets. *Journal of Law, Economics, and Organization*, page ewp006, 2009.

H. David Autor. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, 21(1): 1–42, 2003.

Y.-f. Fong. When does aftermarket monopolization soften foremarket competition. 2008.

C. Genakos, K.-U. Kühn, and J. Van Reenen. Leveraging monopoly power by degrading interoperability: Theory and evidence from computer markets. Technical report, National Bureau of Economic Research, 2011.

M. Gentzkow. Valuing new goods in a model with complementarity: Online newspapers. *The American Economic Review*, 97(3):713–744, 2007.

P. K. Goldberg. Product differentiation and oligopoly in international markets: The case of the us automobile industry. *Econometrica: Journal of the Econometric Society*, pages 891–951, 1995.

L. Grigolon and F. Verboven. Nested logit or random coefficients logit? a comparison of alternative discrete choice models of product differentiation. *Review of Economics and Statistics*, 96(5):916–935, 2014.

J. A. Hausman. Valuation of new goods under perfect and imperfect competition. In *The economics of new goods*, pages 207–248. University of Chicago Press, 1996.

M. Kende. Profitability under an open versus a closed system. *Journal of Economics & Management Strategy*, 7(2):307–326, 1998.

A. Laporte and F. Windmeijer. Estimation of panel data models with binary indicators when treatment effects are not constant over time. *Economics Letters*, 88 (3):389–396, 2005.

C. F. Manski, D. McFadden, et al. *Structural analysis of discrete data with econometric applications*. Mit Press Cambridge, MA, 1981.

F. Mathewson and R. Winter. The law and economics of resale price maintenance. *Review of Industrial Organization*, 13(1-2):57–84, 1998.

D. McFadden. Econometric models of probabilistic choice. *Structural analysis of discrete data with econometric applications*, 198272, 1981.

D. McFadden et al. Conditional logit analysis of qualitative choice behavior. 1973.

C.-H. Miao. Consumer myopia, standardization and aftermarket monopolization. *European Economic Review*, 54(7):931–946, 2010.

A. Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69(2):307–342, 2001.

T. R. Overstreet. *Resale price maintenance: Economic theories and empirical evidence*. The Bureau, 1984.

C. Shapiro. Aftermarkets and consumer welfare: Making sense of kodak. *Antitrust Law Journal*, pages 483–511, 1995.

K. A. Small and H. S. Rosen. Applied welfare economics with discrete choice models. *Econometrica: Journal of the Econometric Society*, pages 105–130, 1981.

K. E. Train. *Discrete choice methods with simulation*. Cambridge university press, 2009.

F. Verboven. International price discrimination in the european car market. *The RAND Journal of Economics*, pages 240–268, 1996.

J. M. Villas-Boas and R. S. Winer. Endogeneity in brand choice models. *Management science*, 45(10):1324–1338, 1999.