The General Nesting Logit (GNL) Model using Aggregate Data

André de Palma, Julien Monardo

The General Nesting Logit (GNL) Model using Aggregate Data*

André De Palma† Julien Monardo‡

July 27, 2017

Preliminary and incomplete
Please do not quote

Abstract

We study the general nesting logit (GNL) model for differentiated products proposed by Fosgerau and de Palma (2016) as a member of the family of generalized entropies built by Fosgerau, Melo, de Palma and Shum (2017), to estimate demand when using aggregate data. We show that the GNL model allows products to be independent, substitutable, or complementary. While Fosgerau and de Palma (2016) show that it can be transformed into a linear regression, we show that this linear regression is very similar to that of Berry (1994) for the nested logit in that it is just a regression of market shares on product characteristics and terms related to its nesting structure. We then use the Dominick's database for estimating the demand for cereals in Chicago in 1991-1992.

Keywords. Demand estimation, Differentiated products, Discrete choice, Generalized entropy, Representative consumer.
JEL codes. C26, D11, D12, L.

---

*First Draft: November 16, 2016. We would like to thank Sophie Dantan, Alessandro Iaria, Hugo Molina and Thibaud Vergé for helpful conversations, as well as seminar participants at CREST-LEI.
†CREST, ENS Paris-Saclay, Université Paris-Saclay; andre.depalma@ens-cachan.fr.
‡CREST, ENS Paris-Saclay, Université Paris-Saclay; jmonardo@ens-cachan.fr.

1 Introduction

Discrete choice models for differentiated products treat these latter as bundles of continuous and/or categorical variables (Lancaster, 1966; McFadden, 1974). For example, Brenkers and Verboven (2006) describe a car using its price and physical characteristics (weight, lenght, etc.), as well as its production location and market segment (class). Nevo (2001) characterizes a cereal using its price and nutrient content (sugar, fiber, etc.), as well as its market segment (kids, family, etc.).

The (two-level) nested logit model is certainly the most widely used model that accounts for these discrete characteristics, which we refer to as criteria. In this model, the choice set is partitioned into several groups, referred to as nests, according to one criterion, where products in a group are closer substitutes than products of different groups. product choice is then considered as a two-step process, in which consumers first choose a group, then choose a product from that group. As an extention, the three-level nested logit (Verboven, 1996b) partitions the choice set into subgroups according to given criterion, which are then partitioned according to another criterion, thereby modelling a product choice as a three-step process.

The nested logit models are highly attractive due to their computational simplicity. They exhibit analytic formulae for both their market share functions and their inverse (Berry, 1994; Berry, Levinsohn and Pakes, 1995), and, with aggregate (sales) data at hand, they can be transformed into a simple linear regression of market shares on product characteristics and nesting terms associated to the nesting structure. However, the nested logit models have been criticized on the ground that it yields substitution patterns that are too restrictive. In addition, once the criteria have been selected, their implementation is complicated by the constraints that the modeller must place on the hierarchy of consumers' choices.

In this paper, we study the general nesting logit (GNL) model for differentiated products proposed by Fosgerau and de Palma (2016), which is a member of the family of generalized entropy models constructed by Fosgerau, Melo, de Palma and Shum (2017) as a subclass of the family of perturbed utility models (see e.g., Fosgerau and McFadden (2012); Fudenberg, Iijima and Strzalecki (2015)). They have been used to model optimization with effort (Mattsson and Weibull, 2002), rational inattention (Matejka and McKay, 2015; Fosgerau, Melo, de Palma and Shum, 2017) The family of perturbed utility models is derived from the maximization of a direct utility function of representative consumer[1], defined by the sum of expected utility and a nonlinear perturbation function. Hofbauer and Sandholm (2002) show that the choice probabilities generated by any additive random utility model (ARUM) can be derived from a perturbed utility model with a deterministic perturbation function. The family of generalized entropy models is obtained by assuming that the perturbation function is a generalized entropy function, which generalizes the Shannon entropy function. Fosgerau, Melo, de Palma and Shum (2017) show a general equivalence between ARUM and rational inattention models when the perturbation function represents the information cost functions and is defined using generalized entropy functions. Fosgerau and de Palma (2016) show that this family extends the family of ARUM. The GNL model is derived by applying the nesting operation to market shares, thereby grouping products according to several criteria.

The GNL model is closely related to the PD-GEV model of Bresnahan, Stern and Trajtenberg (1997) in the sense they both are based on the idea that markets exhibit segmentation along a small number of sources that while overcoming the hierarchical structureof the nested logit models. The difference with this model is that it belongs to the generalized extreme value (GEV) family, and therefore have the features associated to this family, while the GNL model studied here does not. The GNL model generalizes the nested logit models, and therefore embodies the logit (McFadden, 1978) and the two- and three-level nested logit models (McFadden, 1978; Verboven, 1996a) as special cases.

Thanks to its general nesting structure, the GNL model generates rich substitution patterns, while being parsimonious. In this respect, the GNL model provides a very product compromise parsimony/flexibility. The GNL model, as any ARUM, makes the own- and cross-price elasticities functions of a small number of parameters, and contrary to the AIDS, for which the number of parameters to be estimated grows with the number of products, for the GNL the number of parameters grows

[1] Note that Anderson, De Palma and Thisse (1988) and Verboven (1996b) construct the representative consumer's direct utility function for the logit and nested logit discrete choice models, respectively.

with the number of criteria or nests (with nest-specific nesting parameters), and not with the number of products.

It is well-known that any classical ARUM, by allowing one choice among mutually exclusive products, assume a priori that products are substitutable.[2] This also concerns the highly flexible random coefficient logit that can approximate any of them (McFadden and Train, 2000). On the contrary, by allowing for any type of demand relationship across products (complementarity, independence, or substitutability), we argue that the GNL model is a serious competitor of the AIDS model (Deaton and Muellbauer, 1980a). The GNL model is therefore particularly attractive since many important economic questions hinge on the extent to which products are independent, substitutable, or complementarity: for example, this directly affects the incentive to introduce a new product on the market (Gentzkow, 2007) and the incentive to bundle (Armstrong, 2013). It is well-known that the logit model exhibits independence from irrelevant alternatives (IIA) and that the nested logit model exhibits IIA within nests. The GNL model generalizes these models in that it relaxes these restrictions: IIA holds for products that are to that grouped together according to all the criteria (i.e. the GNL model exhibits IIA within the intersection of nests).

In this paper, we use the standard definition of complementary (resp. substitutable) products: a negative (resp. positive) compensated cross-price elasticity of demand (see Samuelson (1974) for a discussion on the different ways of defining complementarity). In our model, this definition is equivalent to the definition by the cross-price elasticity. This is because, in our model, there is no income effect, so that complementarity (resp. substitutability) between two products is the only source of the negative (resp. positive) cross-price elasticity.

The GNL model, as it is the case of any model that deals with market segmentation and involves nests, raises the question of the choice of the nests or equivalently the choice of the criteria. In principle, any categorical variable can define a crite-

[2] Gentzkow (2007) estimates a random coefficient logit with complementarity, where complementarity is captured by the discrete analogue of the cross-partial derivative of utility. He does so by redefining utility over bundles, instead of products, that is, over all the 2^J possible combinations of products. This raises a problem of high dimensionality due to the high number of parameters to be estimated and the high dimension of the integral that defines the probabilities of choice.

rion, and in turn, can be used to partition the choice set. The nested logit models constrain the modeller to impose a hierarchy on the choice pattern, so that the choice set is first partitioned into subsets according to a first criterion, then subsets are partitioned according to second criterion, and so on. On the contrary, the GNL model considers criteria without ordering them, and therefore abstract the modeller from placing constraints on the hierarchy of consumers' choices.

The GNL model is easily estimated using aggregate data and is computationally simple. Fosgerau and de Palma (2016) show that the GNL model can be transformed into a linear regression. We show that this linear regression is very similar to that of Berry (1994) for the nested logit: the GNL model is just a regression of market shares on product characteristics and terms related to its nesting structure. The GNL model can therefore be estimated using standard linear regression techniques, while accounting for price and nesting terms endogeneity. An advantage of the GNL model over the flexible random coefficient logit and the probit models is therefore that it avoids using simulation techniques, and in turn, their associated problems of global convergence (Knittel and Metaxoglou, 2014) and of numerical integration (Skrainka and Judd, 2011), which are also computationally costly and time-consuming.

The flexibility, however, comes at a cost. While there is an analytic formula for the inverse market shares, we cannot derive an analytic formula for the market shares, and in turn, we cannot obtain for the own- and cross-price elasticities independently of each other. This is because the model is formulated in the space of consumption shares and not in the space of indirect utilities as it is the case of the ARUM. The inverse market share functions required to estimate the GNL model by linear regression are thus directly obtained, while recovering the market share functions requires inverting a system of nonlinear equations. Fosgerau, Melo, de Palma and Shum (2017) shows that the invertibility of this system is insured by the theorem on global inverses of homogeneous mapping of Ruzhansky and Sugimoto (2015), but this inversion cannot be performed analytically. This inversion is the opposite operation to the Berry (1994) inversion, which consists in inverting the system of equations that equate predicted and observed market shares. Another caveat is that the GNL model, as it is the case of a large number of demand models estimated in the empirical industrial organization literature, impose a priori symmetry-like restrictions on the substitution patterns, while they do not necessarily hold at the market level and therefore shoud not be imposed a priori.[3]

We therefore resort to Monte Carlo experiments to understand the substitution patterns generated by the GNL model. We find that complementarity may or may not arise for products that are different according to their discrete characteristics, while similar products are always substitutable. One concern is that a large outside product can be decisive in generating complementarity. Using simulations, we show that the size of the outside product is not decisive in generating complementarity, at least for intermediate values of the outside product.

The remainder of this paper is organized as follows. In Section 2, we introduce the family of generalized entropy models and study its properties. In Section 3, we describe the GNL model and provide evidence that it allows for rich substitution patterns. In Section 4, we derive the econometric model and show that any GE model can be written a (non)linear regression model where the error term is additive. In particular, we show that with aggregate data at hand, the GNL can be easily estimated by a linear regression of market shares on product characteristics and additional terms related to its nesting structure. In Section 5, we use the Dominick's database to estimate the demand for ready-to-eat cereals in Chicago in 1991-1992.

2 The Generalized Entropy Model

2.1 The generalized entropy model

Consider a representative consumer with income y who faces a choice set $\mathcal{J} = \{0, 1, \dots J\}$ of differentiated products (J inside products indexed $j > 0$ and an

[3]Davis and Schiraldi (2014) show that all models that belong to the GEV family embody the proportional symmetry restriction, and, among them, all models with utility linear in price impose symmetry. They also show that the random coefficient logit satisfying additive separability in observed and unobserved characteristics, additive separability in observed characteristics and price, and linearity in price impose such a restriction. The flexible coefficient multinomial logit model they propose is able to generate asymmetric substitution patterns, but only in the case in which the model has no underlying welfare function.

outside option indexed $j = 0$[4]), and a *numéraire* good with price 1.

Following Fosgerau, Melo, de Palma and Shum (2017), the family of generalized entropy models is obtained from the maximization of the consumer's utility function, defined by the sum of an expected utility and a generalized entropy function, where the latter is a nonlinear and deterministic function that depends on the vector of choice $q = (q_0, \ldots, q_J)$.

The representative consumer chooses $q \in \Delta$, where Δ is the unit simplex in $\mathbb{R}^{J+1}$, i.e. $\Delta = \left\{ q \in \mathbb{R}^{J+1} | \sum_{j=0}^{J} q_j = 1 \text{ and } q_j \geq 0 \text{ for all } j \in \mathcal{J} \right\}$, so as to maximize her utility function subject to her budget constraint, i.e. she solves

$$\max_{q \in \Delta} u(q, y) = \alpha z + \sum_{j=0}^{J} v_j q_j + \Omega(q) \quad \text{subject to} \quad \sum p_j q_j + z \leq y,$$

where z is consumption of the *numéraire* good, $\alpha > 0$ is the marginal utility of income, v_j and p_j are respectively the quality and price of product j, and $\Omega(q)$ is a generalized entropy function.

The budget constraint is binding so that the consumer's program is rewritten as follows

$$\max_{q \in \Delta} u(q, y) = \alpha y + \sum_{j=0}^{J} \delta_j q_j + \Omega(q),$$

where $\delta_j = v_j - \alpha p_j$ is the net utility that the consumer derives from the consumption of product j.[5]

The generalized entropy function $\Omega : [0, \infty)^{J+1} \to \mathbb{R} \cup \{-\infty\}$ is given by

$$\Omega(q) = \begin{cases} -\sum_{j=0}^{J} q_j \ln S^{(j)}(q), & q \in \Delta \\ -\infty, & q \notin \Delta \end{cases}, \tag{1}$$

where the function $S(\cdot) = \left(S^{(0)}(\cdot), \ldots, S^{(J)}(\cdot)\right) : (0, \infty)^J \to (0, \infty)^J$ is a *flexible generator*, which satisfies the four conditions outlined in Axiom 1:

[4]The reader should keep in mind that outside option just means a good whose (net) utility can be normalized to zero.

[5]In the empirical industrial organization literature (Berry, 1994; Berry, Levinsohn and Pakes, 1995), δ refers to as the mean utility.

Axiom 1 (Properties of the flexible generators)**.** *The flexible generators S satisfies the following conditions:*

1. *$S(q)$ is twice continuously differentiable, and homogeneous of degree 1.*
2. *$\Omega(q)$ is concave over Δ, or equivalently $\mathbf{J}_{\ln S}$, is positive semi-definite.*
3. *At any $q \in \Delta$, the Jacobian of $\ln S$, $\mathbf{J}_{\ln S}$, is symmetric, and $-\frac{\partial \Omega(q)}{\partial q_k} = \ln S^{(k)}(q) + 1$.*
4. *$S(q)$ is globally invertible, and its inverse is denoted by $H = S^{-1}$.*

The following lemma shows that, together with Condition 1, Condition 3 is equivalent to Condition 3 in Fosgerau, Melo, de Palma and Shum (2017) given by

$$\sum_{j=0}^{J} q_j \frac{\partial \ln S^{(j)}(q)}{\partial q_k} = 1,\ k = 0, \ldots, J. \tag{2}$$

Lemma 1. *Assume that S is twice continuously differentiable and homogeneous of degree 1. Then, the Jacobian of $\ln S$, $\mathbf{J}_{\ln S}$, is symmetric if and only if (2) is satisfied.*

Proof. See Appendix A.1. □

Condition 3 implies that $-\partial\Omega(q)/\partial q_k = \ln S^{(k)}(q) + 1$, which, as show in Theorem 1, allows to derive a tractable and familiar form for demand. We discuss in the following subsection the economic content of these 4 conditions. To summarize, the GE models are therefore derived from the maximization of the direct utility function given by

$$u(q, y) = \alpha y + \sum_{j=0}^{J} \delta_j q_j - \sum_{j=0}^{J} q_j \ln S^{(j)}(q), \tag{3}$$

where $S(q)$ satisfies Axiom 1.[6]

[6] The utility (3) embodies two different components: the linear term $\sum_{j=0}^{J} \delta_j q_j$ and the nonlinear term $\sum_{j=0}^{J} q_j \ln S^{(j)}(q)$. If the utility were only composed of the first component, then the representative consumer would maximize her utility by only choosing the product j^* with the highest net utility δ_{j^*}. This component therefore describes the net utility derived from the consumption of the

Maximization of the utility function (3) lead to demands that are given as follows:

Theorem 1. *Let $S(q)$ be a flexible generator satisfying the conditions outlined in Axiom 1. Maximization of utility (3) leads to a demand system, with interior solution*

$$q_i(\delta) = \frac{H^{(i)}\left(e^{\delta}\right)}{\sum_{j=0}^{J} H^{(j)}\left(e^{\delta}\right)},\ i = 0, \ldots, J. \tag{4}$$

where $H^{(i)} = S^{-1(i)}$.

Net utility δ and demand q are related through the flexible generator S by

$$\delta = \ln S(q) + c, \tag{5}$$

for some $c \in \mathbb{R}$. Consumers' surplus is given by

$$CS = \alpha y + \ln\left(\sum_{j=0}^{J} H^{(j)}\left(e^{\delta}\right)\right). \tag{6}$$

Proof. See Appendix Fosgerau, Melo, de Palma and Shum (2017). □

Maximization of utility (3), when $\Omega(q) = -\mu \sum_{j=0}^{J} q_j \ln(q_j)$ is the Shannon entropy function, or equivalently when $S(q) = q$, gives rise to the logit demand (Anderson, De Palma and Thisse, 1988):[7]

$$q_i(\delta) = \frac{e^{\delta_i/\mu}}{\sum_{j=0}^{J} e^{\delta_j/\mu}}. \tag{7}$$

The first part of Theorem 1 shows that the demand system (4) generalizes the logit demand. In the same way that different ARUM are obtained from different

vector q in the absence of interaction among products. In this respect, the net utility δ_j measures the intrinsic contribution to one unit of product j. In contrast, if the utility were only composed of the second component, then the representative consumer would maximize his or her utility by choosing all the products with equal share $1/(J+1)$. This component therefore expresses the variety-seeking behavior of the representative consumer.

[7]Everything else being equal, the larger is μ, the greater is the preference for diversity: when $\mu \to 0$, diversity is not valued per se and the consumer solely buys the product j with the largest net utility δ_j; when $\mu \to \infty$, consumption is divided equally among all products in the choice set.

assumptions about the distribution of the random part of the utility, different GE models are obtained from different specifications of the flexible generator function S. Fosgerau and de Palma (2016) provide different ways of generating flexible generators and show that some generalized entropy models are not consistent with any ARUM. They also show that for any ARUM there exists a GE model that leads to the same demand. The second part of Theorem 1 shows that any GE model generates the relation (5) between net utility δ and demand q, which will be used as a basis for demand estimation. The third part of Theorem 1 show that, in the same way as in the logit model, the consumers' surplus in any GE model is simply the log of the denominator of the demand system (4).

2.2 Properties of the Generalized Entropy Model

We begin this subsection with the following definition that involves integrability and rationality:[8]

Definition 1 (Lewbel, 2001)**.** A demand system is defined to be *integrable* if it satisfies Adding up, Homogeneity, and Slutsky symmetry. A demand is defined to be *rational* if it is integrable and it also satisfies negative semidefiniteness. *Adding up* is that the total value of demands is total expenditure. *Homogeneity* is that demands are homogeneous of degree 0 in prices. *Slutsky symmetry* is that the cross-price derivatives of demands are symmetric. *Negativity* is that the matrix of own- and cross-price derivatives is negative semidefinite.

Intuitively, integrability implies that the demands satisfy the first-order conditions for utility maximization, while rationality correspond to the conditions for demands to be derived from utility maximization. The following proposition shows that the demand system (4) is indeed rational and helps us in understanding the economic content of Conditions 1 to 4:

Proposition 1. *The demand system (4) satisfies Adding-up, Homogeneity, Slutsky symmetry, and Negativity.*

[8] Due to the linearity of utility in income, demands will be independent of income, so that the Slutsky matrix reduces to the matrix of cross-price derivatives.

Proof. See Appendix A.2. □

In other words, adding-up means $\sum_{j=0}^{J} p_j q_j = y - z$, homogeneity implies that for all scalar $\lambda > 0$ and for all j $q_j(\lambda p) = q_j(p)$, Slutsky symmetry means that for all $j \neq i$ $\partial q_i / \partial p_j = \partial q_j / \partial p_i$, and negativity means that the matrix $(\partial q_i / \partial p_j)$ is negative semidefinite.

Conditions 1 to 4 presented in Axiom 1 are grounded in economic theory. Conditions 1 and 3 make demands integrable. Condition 1, together with Condition 3, ensures that Roy's identity holds. Condition 3, together with Condition 1, is a sufficient condition for the Slutsky matrix to be symmetric. We can indeed show that every demand system satisfying Conditions in Axiom 1 exhibits a symmetric Slutsky matrix, while such a symmetry condition can fail to hold for demand systems for which Condition 3 does not hold. Condition 4, together with Condition 1, ensures the existence of the solution of utility maximization, while Condition 2 ensures that it is unique.

Complementarity and substitutability in the GE model. The matrix of own- and cross-price elasticities being defined, it is interesting to study the patterns of substitution the generalized entropy models allow for.

We use the standard definition of complementary (substitutable) products, i.e. a negative (positive) compensated cross-price elasticity (or equivalently derivative) of demand (see Samuelson (1974) for a discussion on the different ways of defining complementarity).[9] Due to the linearity of utility in income, there is no income effect, so that complementarity (substitutability) between products is the only source of the negative (positive) cross-price elasticity. The elements of the Slutky matrix are therefore the cross-derivatives of demand, which implies the following definition:

Definition 2. Products i and j are substitutable if $\frac{\partial q_i}{\partial p_j} > 0$, independent if $\frac{\partial q_i}{\partial p_j} = 0$, and complementary if $\frac{\partial q_i}{\partial p_j} < 0$.

In the GE model, the flexible generator S gives all the relevant information on the type of demand relationship across products (i.e., complementarity, indepen-

[9]Note that in this paper we do not pretend to explain what the sources of complementarity are.

dence, or substitution) it allows. To show this, we compute the cross-price derivatives by first totally differentiating the first-order conditions of the utility maximization program, and then by matrix manipulation (see e.g., Deaton and Muellbauer (1980b, Chapter 2)). We summarize this as follows:

Proposition 2. *In the GE model, the derivative of demand q_j with respect to price p_i is given by*

$$\frac{\partial q_j}{\partial p_i} = \alpha \frac{(-1)^{i+j} \det (M_{ij})}{\det (M)},$$

where $M = -\begin{bmatrix} \mathbf{J}_{\ln S} & \mathbf{1} \\ \mathbf{1}^\intercal & 0 \end{bmatrix}$ *is a negative definite matrix with* $\mathbf{1} = (1, \ldots, 1)^\intercal$.

Therefore, $\frac{\partial q_j}{\partial p_j} < 0$ *and* $\frac{\partial q_j}{\partial p_i} \lesseqgtr 0$ *, i.e. goods can be substitutes or complements, depending on the magnitude of the derivative of* $\ln S$.

Proof. See Appendix A.3. □

As a consequence, since in any ARUM, including the highly flexible random coefficient logit, all products are substitutes, we get the following result:

Corollary 1. *Some GE models leads to demand systems that cannot be rationalized by any ARUM.*

However, in the following subsection, we show that the converse is true, i.e., any ARUM can be generated by a GE model. Lastly, given the demand system (4), we derive the formula for the matrix of own- and cross-price elasticities given as follows:

Proposition 3. *The matrix of own- and cross-price elasticities are given by*

$$\Sigma = \left[\frac{\partial q_j}{\partial p_i}\frac{p_i}{q_j}\right] = -\alpha \mathit{diag}\left(p \cdot S\left(q\right)\right) \cdot \left[\mathbf{J}_S\left(q\right)^\intercal\right]^{-1} \cdot \left(\mathit{diag}\left(1/q\right) - J_J\right),$$

where J_J is the $(J \times J)$ unit matrix and $\mathbf{J}_S(q)$ is the Jacobian of S evaluated at q.

Proof. See Appendix A.4. □

For example, with the logit model we have $\text{diag}\left(S\left(q\right)\right) = \text{diag}\left(q\right)$ and $\mathbf{J}_S\left(q\right) = I_J$, where I_J is the $J \times J$ identity matrix, so that the own-price elasticities are $\Sigma_{jj} = -\alpha\left(1 - q_j\right)p_j$ and the cross-price elasticities are $\Sigma_{ij} = \alpha q_i p_j$.

3 The GNL Model

3.1 The GNL Model

In the spirit of Bresnahan, Stern and Trajtenberg (1997) and following Fosgerau and de Palma (2016), let us introduce the GNL model. Consider a market for differentiated products that exhibits segmentation according to a given number C of *dimensions*, indexed c. Each dimension c taken separately potentially provides a source of market segmentation and define a finite number of *nests* according to the modalities that it takes. The dimensions taken together define *groups* and product *types*. Products of the same type are those that are grouped together according to all the dimensions and that therefore belong to the same group. Each dimension defines a concept of product closeness (or similarity), so that products of the same type would be closer substitutes than products of different types. In general, a nest may or may not be a subnest of another nest, thus defining nonoverlapping or overlapping nests (but it will always allow for nonoverlapping nests).

Let $\sigma_c(j)$ be the set of products that are grouped together with product j on dimension $c = 1, \ldots, C$, and $q_{\sigma_c(j)} = \sum_{i \in \sigma_c(j)} q_i$. The GNL model is derived from the maximization of utility (3), where $S = \left(S^{(1)}, \ldots, S^{(J)}\right)$ defined by

$$S^{(j)}(q) = \begin{cases} q_0, & \text{if } q \in \Delta,\ j = 0, \\ q_j^{\mu_0} \prod_{c=1}^{C} q_{\sigma_c(j)}^{\mu_c}, & \text{if } q \in \Delta,\ j > 0, \\ -\infty & \text{if } q \notin \Delta. \end{cases} \tag{8}$$

with $\mu_0 + \sum_{c=1}^{C} \mu_c = 1$ and $\mu_c > 0$ for all $c = 0, \ldots, C$, is a flexible generator that satisfies Conditions 1 to 4.

The NL model (McFadden, 1978) is a special case of the GNL model that arises when the market exhibits segmentation according a single dimension. The NL model is highly attractive due to its computational simplicity. It exhibits an analytical formulae for both its market share functions and their inverse (Berry, 1994), which renders it very easy to study and to estimate. In addition, with aggregate data, it can be transformed into a linear regression of market shares on product characteristics and a nesting term. However, it has been criticized on the ground that it yields substitution patterns that are too restrictive. As we show in the following, the GNL model provides flexible substitution patterns, while remaining as computationally simple and very easy to estimate with aggregate data as the NL model. However, the GNL model does not exhibit an analytic formulae for the market shares functions, rending its study a little bit more challenging.

To understand how the GNL model works, consider a market for cereals that exhibit segmentation according to two dimensions: the first one measures the advantage provided by the brand name (Kellogg's or General Mills), and the second one measures the closeness between products that are suitable for kids and those that are not (kids-friendly or not). The three-level NL model (Verboven, 1996a) is able to deal with this market segmentation, but will treat the two dimensions hierarchically as for example in Figure 1 (with the nests for brand name on top). In this example, the hierarchy implies that the modeller does not consider the two dimensions independently of each other, but instead only consider the closeness between products that have the same brand name but that belong to different segments (as defined by the dimension kids-friendly or not).

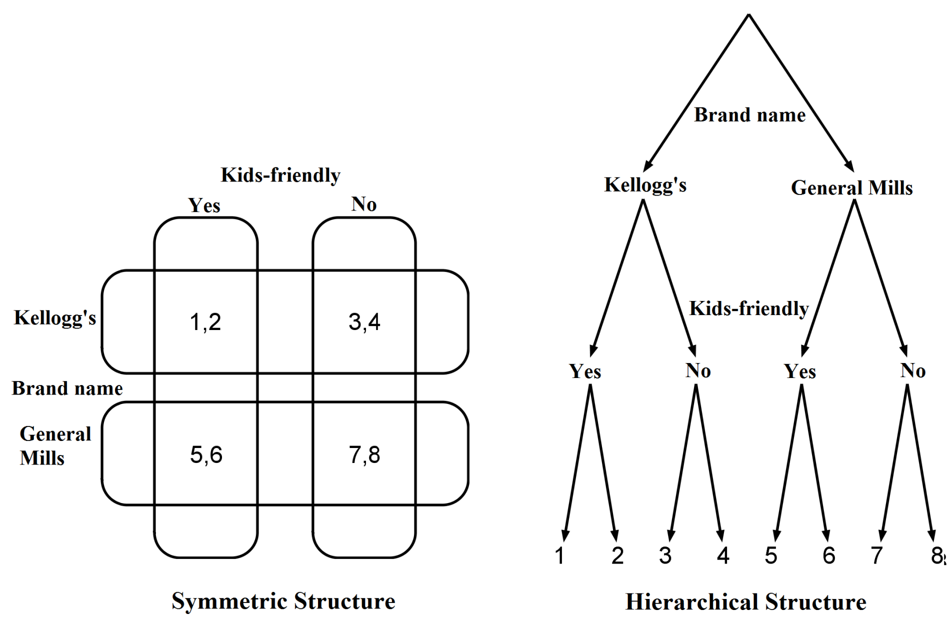

Figure 1: Cereals: symmetric vs. hierarchical structure

In this particular case, the generator is given by $S^{(j)}(q) = q_j^{\mu_0} q_{\sigma_1(j)}^{\mu_1} q_{\sigma_2(j)}^{\mu_2}$, for $j = 0, \ldots, J$, with $\mu_0 + \mu_1 + \mu_2 = 1$, where the $\sigma_{2(j)}$'s partition the $\sigma_{1(j)}$'s , i.e., the choice set is partitioned into G nests, where each nest g is partitioned into H_g subnests, where each subnests h contains J_{hg} products.

The generator can be inverted to recover the three-level NL logit demand:

$$q_j = \frac{\exp\left(\delta_j / \left(1-\gamma_2\right)\right)}{\exp\left(I_{hg} / \left(1-\gamma_2\right)\right)} \frac{\exp\left(I_{hg} / \left(1-\gamma_1\right)\right)}{\exp\left(I_g / \left(1-\gamma_1\right)\right)} \frac{\exp\left(I_g\right)}{\exp\left(I\right)},$$

where the inclusives values I_{hg}, I_g and I are defined as follows

$$\begin{aligned} I_{hg} &= \left(1-\gamma_2\right) \ln\left(\sum_{j=1}^{J_{hg}} \exp\left(\delta_j / \left(1-\gamma_2\right)\right)\right), \\ I_g &= \left(1-\gamma_1\right) \ln\left(\sum_{h=1}^{H_g} \exp\left(I_{hg} / \left(1-\gamma_1\right)\right)\right), \\ I &= \ln\left(\sum_{g=1}^{G} \exp\left(I_g\right)\right), \end{aligned}$$

with $\gamma_2 = \mu_1 + \mu_2$ and $\gamma_1 = \mu_1$.[10]

The GNL overcomes this hierarchical structure by treating the two dimensions symmetrically and independently of each other, as in Figure 1.

In three-level NL logit model, the researcher must choose between the two contradictory models with hierarchical structure (i.e., between the model with the nests for brand name on top and the one with the nests for segments), that is, determine *a priori* which of the two dimensions is perceived as more important by the consumers. In contrast, the GNL model avoids the researcher to choose between them.

[10]Note that the constraints on the parameters are consistent. For the three-level NL model to be consistent with random utility maximization, the parameters must satisfy $0 \leq \gamma_1 \leq \gamma_2 \leq 1$. For the GNL model to be a GE model, the parameters must satisfy: $\mu_1 \geq 0$, $\mu_2 \geq 0$, and $\mu_1 + \mu_2 \leq 1$, which is equivalent to $0 \leq \gamma_1 \leq \gamma_2 \leq 1$.

3.2 Properties of the GNL model

Independence from irrelevant alternatives (IIA). IIA can be seen as a restriction on the ratios of probabilities (the ratio of probabilities between products j and k, does not depend on any products other than j and k). To study this, we consider the relation between δ and q derived in Theorem 1, which for the GNL model is given $e^{\delta_j - c} = q_j^{\mu_0} \prod_{c=1}^{C} q_{\sigma_c(j)}^{\mu_c}$ for all $j > 0$, i.e., for any pair of products j and k

$$\frac{q_j(\delta)}{q_k(\delta)} = \exp\left(\frac{\delta_j - \delta_k}{\mu_0}\right) \exp\left(\sum_{c=1}^{C} \frac{\mu_c}{\mu_0} \ln\left(\frac{q_{\sigma_c(k)}(\delta)}{q_{\sigma_c(j)}(\delta)}\right)\right). \tag{9}$$

Then, for products j and k of the same type, Equation (9) reduces to $\frac{q_j}{q_k} = \exp\left(\frac{\delta_j - \delta_k}{\mu_0}\right)$, so that IIA holds for products of the same type. Otherwise, for products of different types, IIA does not hold.

Complementarity/Substitutability. We have already showed that the GE model allows for complementarity. We now show that it is also the case of the GNL model. To do so, consider a particular case with 3 inside products and one outside option. Inside products are grouped according two dimensions. For the first dimension, product 1 is in one nest, and products 2 and 3 are in a second nest. For the second dimension, products 1 and 2 are in one nest, and product 3 is in a second nest. We proceed as above by first totally differentiating the first-order conditions of the utility maximization program, and then by matrix manipulation. We show (see Appendix B.1) that the cross-derivative $\frac{\partial q_3}{\partial p_1}$ is negative for δ that satisfies $\mu_0 (q_1(\delta) + q_2(\delta))(q_2(\delta) + q_3(\delta)) - \mu_1\mu_2 q_0(\delta) q_2(\delta) < 0$. Therefore, complementarity arises in the GNL model and is a local property.

The properties of the GNL model are summarized as follows:

Proposition 4. *The GNL model generated by the flexible generator (8) has the following properties:*

1. *it is observationally equivalent to the three-level NL with nesting parameters* (γ_1, γ_2) *when there is a hierarchical structure on nests and when the parameters satisfy* $\gamma_2 = \mu_1 + \mu_2$ *and* $\gamma_1 = \mu_1$.

2. *it does not exhibit IIA. IIA holds for products of the same type.*

3. *it allows for complementary, independent, or substitutable products.*

4. *it is always a GE model and sometimes an ARUM.*

Proof. See Appendix B.1. □

Elasticities. Given the demand system (4), we can compute an analytic formula for the matrix of own- and cross-price elasticities, but we cannot derive an analytical formula for the market share functions, and in turn, we cannot obtain the own- and cross-price elasticities independently of each other. Recall that $\sigma_c(j)$ be the set of products that are grouped together with product j on dimension c. The $J \times J$ matrix Θ_c is the nesting structure matrix for dimension c, i.e.,

$$(\Theta_c)_{ij} = \begin{cases} 1, & \text{if } i \in \sigma_c(j), \\ 0, & \text{otherwise.} \end{cases}$$

The following proposition gives the expression of the matrix of own- and cross-price elasticities:

Proposition 5. *The matrix of own- and cross-price elasticities is given by*

$$\Sigma = \left[\frac{\partial q_j}{\partial p_i}\frac{p_i}{q_j}\right] = \text{-}\alpha diag\,(p \cdot q) \cdot \rho\left(\mu; (\Theta_c)_{c=1,\ldots,C}\right) \cdot \left(diag\,(1/q) - J_J\right),$$

where

$$\rho\left(\mu; (\Theta_c)_{c=1,\ldots,C}\right) = \left[\mu_0 I_J + \sum_{c=1}^{C} \mu_c \frac{q_j}{q_{\sigma_c(j)}} \cdot \Theta_c\right]^{-1}$$

is the market structure matrix and J_J denotes the $J \times J$ unit matrix (where every element is equal to one).

Proof. See Appendix B.2. □

The market structure matrix summarizes the information on the different sources of market segmentation. For example, it just reduces to the identity matrix in the logit model.

We have just established the properties of the GNL model. Some numerical properties are also provided in Appendix **??**. Our findings are that (i) products belonging to the same group (i.e., those of the same type) are never complementary, while products belonging to different groups may or may not be complementary; (ii) higher values of the outside option, at least for intermediate ones, do not generate complementarity; (iii) the size of the cross-elasticities depends on the degree of closeness between products as measured by the belonging to the same nests and groups and the value of the nesting parameters.

The next section deals with the estimation of GE models and especially with the estimation of the GNL model.

4 Econometric Setting

Estimating GE models. Estimating GE demand (4) for differentiated products at the product level requires data on market shares q_{jt}, prices p_{jt}, and characteristics $\mathbf{X}_{jt}$ of each product j in each market t.

We follow Fosgerau and de Palma (2016) while making clear the distinction between parameters in the linear part of the utility and the ones in the nonlinear part. Consider T markets ($t = 1, \ldots, T$) with J inside products ($j = 1, \ldots, J$) and an outside option ($j = 0$). Let ξ_{jt} be the unobserved product characteristics of product j in market t. We assume that net utility is parametrized as the linear combination $\delta_{jt}\left(\mathbf{X}_{jt}, p_{jt}, \xi_{jt}; \theta\right) = \beta_0 + \mathbf{X}_{jt}\beta - \alpha p_{jt} + \xi_{jt}$, with $\theta = (\alpha, \beta_0, \beta)$. The intercept β_0 captures the value of consuming an inside product instead of the outside option; the parameter vector β represents the consumers' taste for the characteristics $\mathbf{X}_{jt}$; and the parameter $\alpha > 0$ is consumers' price sensitivity (i.e. the marginal utility of income).

Then, from Theorem 1, we have

$$\ln S^{(jt)}\left(q; \theta_2\right) = \delta_{jt}\left(\mathbf{X}_{jt}, p_{jt}, \xi_{jt}; \theta_1\right) + c_t,$$

for all $j = 0, \ldots, J$ and $t = 1, \ldots, T$, and where $c_t \in \mathbb{R}$, and $\theta = (\theta_1, \theta_2)$ are the demand parameters to be estimated. The parameters θ_1 belong to the linear part of the utility and are labelled *linear parameters*, while the parameters θ_2 belong to the nonlinear part of the utility and are labelled *nonlinear parameters*. Assuming that $\delta_{0t} = 0$ for all $t = 1, \ldots, T$, we get

$$\begin{aligned} \ln S^{(jt)}\left(q;\theta_2\right) - \ln S^{(0t)}\left(q;\theta_2\right) =& \delta_{jt}\left(\mathbf{X}_{jt}, p_{jt}, \xi_{jt}; \theta_1\right), \\ =& \beta_0 + \mathbf{X}_{jt}\beta - \alpha p_{jt} + \xi_{jt}, \end{aligned} \tag{10}$$

for all $j = 1, \ldots, J$ and $t = 1, \ldots, T$, where the econometric error term ξ_{jt} enters linearly.

Any GE model may therefore be a nonlinear (in the general case) or linear (in the case in which the terms $\ln S^{(jt)}$ are linear in the nonlinear parameters θ_2) regression model where the error term ξ_{jt} is additive. Let us now consider the case of the GNL model for which the flexible generator is given by Equation (8) and for which the terms $\ln S^{(jt)}$ are linear in the nonlinear parameters θ_2.

Estimating the GNL model. Let $\sigma_c(jt)$ be the set of products that are grouped together with product j in market t on dimension c, and $q_{\sigma_c(jt)} = \sum_{i \in \sigma_c(jt)} q_{it}$, and recall that the GNL model is derived using the following flexible generator:

$$S^{(jt)}\left(q_t\right) = \begin{cases} q_{0t}, & j = 0, \\ q_{jt}^{\mu_0} \prod_{c=1}^{C} q_{\sigma_c(jt)}^{\mu_c}, & j > 0, \end{cases},$$

where $\mu_0 + \sum_{c=1}^{C} \mu_c = 1$ and $\mu_c > 0$ for all $c = 0, \ldots, C$.

Then from Equation (10), we have

$$\mu_0 \ln\left(q_{jt}\right) + \sum_{c=1}^{C} \mu_c \ln\left(q_{\sigma_c(jt)}\right) - \ln\left(q_{0t}\right) = \delta_{jt}\left(\mathbf{X}_{jt}, p_{jt}, \xi_{jt}; \theta\right),$$

and using the constraint on parameters $\mu_0 + \sum_{c=1}^{C} \mu_c = 1$, we get

$$\ln\left(q_{jt}\right) - \ln\left(q_{0t}\right) = \delta_{jt}\left(\mathbf{X}_{jt}, p_{jt}, \xi_{jt}; \theta\right) + \sum_{c=1}^{C} \mu_c \ln\left(\frac{q_{jt}}{q_{\sigma_c(jt)}}\right).$$

This equation is the same as for the logit, the nested logit, and the three-level nested logit, except for the additional term $\sum_{c=1}^{C} \mu_c \ln \left(\frac{q_{jt}}{q_{\sigma_c(j)}} \right)$ (see Berry (1994); Brenkers and Verboven (2006)). This term is equal to the weighted sum of the logarithms of the market shares of products within their respective groups for each dimension c, weighted by the μ_c's.

With parametrization $\delta_{jt}\left(\mathbf{X}_{jt}, p_{jt}, \xi_{jt}; \theta\right) = \beta_0 + \mathbf{X}_{jt}\beta - \alpha p_{jt} + \xi_{jt}$, we have therefore showed that the GNL model is a linear regression model in the sense that it generates (demand) equations in both the linear and nonlinear parameters and the econometric error enter linearly:

Proposition 6. *The GNL model generates the following demand equations*

$$\ln\left(q_{jt}\right) - \ln\left(q_{0t}\right) = \beta_0 + \mathbf{X}_{jt}'\beta - \alpha p_{jt} + \sum_{c=1}^{C} \mu_c \ln\left(\frac{q_{jt}}{q_{\sigma_c(jt)}}\right) + \xi_{jt}, \tag{11}$$

for all $j = 1, \ldots, J$ and all $t = 1, \ldots, T$.

This regression is just the regression of the GNL model in Fosgerau and de Palma (2016), where we explicitly take into account the constraint on parameters. The GNL model is therefore easily estimated using standard linear regression techniques. This is in constrast to the random coefficient logit model that requires simulation techniques, and in turn, is associated with problems of global convergence (Knittel and Metaxoglou, 2014) and of numerical integration (Skrainka and Judd, 2011).

Two comments are in order. First, the prices p_{jt}, as well as, the nesting terms $\ln\left(\frac{q_{jt}}{q_{\sigma_c(jt)}}\right)$ are likely to be endogenous (as discussed in Subsection 5.2), suggesting the need of appropriate instrumental variables. Second, Equation (11) can be rewritten with nest-specific parameters μ_{cn} (for nests n on dimension c) instead of μ_c, where $\mu_{0n} + \sum_{c=1}^{C} \mu_{cn} = 1$ for all j and n and $\mu_{cn} > 0$ for all $c = 0, \ldots, C$ and for all n.

5 Empirical Application: Demand for Cereals

5.1 Data

We use data from the Dominick's Database made available by the James M. Kilts Center, University of Chicago Booth School of Business.[11] They comprise all Dominick's Finer Foods chain stores in the Chicago metropolitan area over the period 1989-1997, and concern 30 categories of packaged products. They are weekly store-level scanner data at the UPC level, and include unit sales, retail price, and information on stores traffic. The data are supplemented by store-specific demographics (including information on income, age, household size, etc.)

For our analysis, we consider the ready-to-eat (RTE) cereal category during the period 1991-1992. This market is known to exhibit segmentation (see e.g, Nevo (2001)). This period describes a stable demand period: before (May 1990), DFF stores introduced several private labels (Meza and Sudhir, 2010) which certainly had a significant effect on brand preferences, and after (January 1993), Post purchased the Nabisco cereal branch (Michel, 2016). We also supplement the data with the nutrient content (fiber, sugar, calories,etc.) of the RTE cereals using the USDA Nutrient Database for Standard Reference made available by the United States Department of Agriculture[12]; and with the price for sugar (which will be used as instrument) from the website www.indexmundi.com.

Following the prevailing literature, we aggregate the UPCs into brands (e.g., Kellogg's Special K), so that different size boxes are considered one brand[13], where a brand is a cereal (e.g., Special K) associated to its brand name (e.g., Kellogg's). We focus attention on the top 50 brands, presented in Table 11, which account for 73 percent of sales in the category for the period and stores considered in the study.

[11] The dataset is available at https://research.chicagobooth.edu/kilts/marketing-databases/dominicks.

[12] This dataset provides the nutrient content of more than 8,500 different foods including ready-to-eat cereals. I used releases SR11 and SR16 (for sugar). It is available at https://www.ars.usda.gov/northeast-area/beltsville-md/beltsville-human-nutrition-research-center/nutrient-data-laboratory/docs.

[13] One potential problem with this aggregation arises if the price/serving is lower for larger than for smaller size boxes.

In our main specifications, a product is defined as a brand, and a market as a store-month combination. We also run other specifications in which a product is defined as a brand-store combination and each month defines a market.

Following Nevo (2001), we define market shares of the (inside) products by converting volume sales into number of servings sold, and then by dividing it by the total potential number of servings at a store in a given month. To compute the potential market size, we assume that the serving weight is 35 grammes, and that an individual in a household consumes around 15 servings per month.[14] The potential month-week market size (in servings) is computed as the average number (per month) of households which visited that store in that given month, times the average household size for that store, times the number of servings an individual consumers in a month.[15] The market share of the outside option is then the difference between one and the sum of the inside products market shares. Following Nevo (2001), we compute the price of a serving weight by dividing the dollar sales by the number of servings sold, where the dollar sales reflect the price consumers paid.

Descriptive Statistics. We consider two dimensions, as reported in Table 11: the segments (family, kids, health, taste enhanced) as defined by the prevalent literature (see e.g., Nevo (2001)), and the brand names (General Mills, Kellogg's, Quakers, Post, Nabisco, Ralston). Combining these two dimensions results in 17 types of products (i.e., pairs brand name-segment).

The sample we use consists of the six biggest companies (General Mills, Kellogg's, Nabisco, Post, Quaker, Ralston). Kellogg's is the biggest one with large market shares in all segments, and General Mills, the second biggest one, is especially present in the family and kids segments. The latter two taken together account for around 80 percent of the market. The biggest segments are the family and kids segments which, together, account for almost 70 percent of the market (in sales).

The following table shows the nutrient content of the cereals by segment and

[14] According to USDA's Economic Research Service, per capita consumption of RTE breakfast cereals was equal to around 14 pounds (that is, about 6350 grammes) in 1992, which is equivalent to serving 15 weights per week.

[15] We use the average number of households instead of the number of households itself to take into account the fact that consumer visits stores several times a month.

brand name:

	Sugar g/serve	Energy Kcal/serve	Fiber g/serve	Lipid g/serve	Sodium mg/serve	N
Segment						
Family	6.46	111.78	1.90	0.85	231.14	17
Health/nutrition	4.31	105.03	2.71	0.46	144.47	9
Kids	11.49	118.07	0.85	1.15	181.18	16
Taste enhanced	8.32	110.81	2.84	1.91	142.65	8
Brand Name						
General Mills	8.50	113.22	1.70	1.30	197.74	17
Kellogg's	8.21	109.28	2.12	0.73	195.85	18
Nabisco	0.22	107.55	2.94	0.50	1.80	2
Post	10.30	112.08	1.79	0.88	181.74	5
Quaker	7.28	119.52	1.94	2.08	137.04	5
Ralston	6.08	118.70	0.50	0.44	261.80	3
Total	7.98	112.42	1.86	1.05	185.39	50

Table 1: Sample statistics: nutrient content by segment and by brand

We observe that cereals for health contain less sugar, more fiber, less lipid, and less sodium, and are less caloric. Cereals for Kids contain more sugar and are more caloric. Nabisco offers cereals with less sugar and are less caloric, while Quaker and Ralston offer more caloric cereals. The two dimensions therefore proxy, at least partially, for the continuous variables (sugar, calories, fiber, lipid, sodium).

5.2 Identification

We have a system of demand equations (11), where ξ_{jt} is the error term. We can include product-specific dummies as product characteristics and exploit the panel features of the data by specifying the error term as $\xi_{jt} = \xi_j + \xi_t + u_{jt}$, where ξ_j is the product fixed effect, which captures market-invariant unobserved characteristics of product j; ξ_t is the market fixed effect, which captures market-specific demand shocks relative to the outside option; and u_{jt} is the residual term, which captures the remaining unobserved product characteristics, varying across products and markets. In our application, the structural error term u_{jt} will for example include advertising, shelf-space, and positioning of the product among others that affect consumer utility that consumers and firms, but not the econometrician, observe.

The presence of unobserved characteristics creates the well-known problem of price endogeneity: if firms and consumers know the values of the unobserved product characteristics u_{jt}, while the econometrician does not, then prices are likely to be correlated with them. In addition, in the GNL model, the C nesting terms $\ln\left(q_{jt}/q_{\sigma_c(jt)}\right)$ are endogenous by construction, since any shocks to u_{jt} that increase the dependent variable $\ln\left(q_{jt}/q_{0t}\right)$ in Equation (11) will also increase the C nesting terms $\ln\left(q_{jt}/q_{\sigma_c(jt)}\right)$

The main identification assumption is the existence of instruments z_t. Exogenous variables can be instrumented by themselves. As a consequence, we need to find at least $(C+1)$ appropriate instruments to estimate the GNL model. They are variables that are correlated with the endogeneous variables (relevance) but are not correlated with the error term u_{jt} (exogeneity). Formally, they have to satisfy the conditional moment restrictions $\mathbb{E}\left[u_{jt}|z_t\right]=0$, which lead to the unconditional moment restrictions $\mathbb{E}\left[z_t u_{jt}\right]=0$.

We use two sets of instrumental variables. First, the characteristics-based instruments, which are valid under the assumption that the vector of characteristics is exogenous, and therefore uncorrelated to the error term u_{jt}. The idea is that characteristics of other products are correlated with price since the markup of each product depends on how close products are into the characteristics space. We use the promotional activity of the firms in a given market, which varies both across stores for a given time and across time for a given store. The idea is that for a given product, promotional activities of a firm affect consumers' choices, and is therefore correlated with the price of that firm, but uncorrelated with the error term u_{jt}.[16] We motivate this assumption by the fact that, at Dominicks, the promotional calendar is known several weeks in advance of the weekly price decisions, so that promotion is an exogenous variable.[17] We use the number of promoted products in the same

[16]We could use the vector of product characteristics, the sum of characteristics of other products of competing firms, and the sum of characteristics of other products of the same firm, as suggested by Berry, Levinsohn and Pakes (1995); and for the NL and the GNL models, we could also use the sum of characteristics over products belonging to the same nests (Verboven, 1996a). However, they are invariant across time and stores

[17]Pricing decisions are done on a chainwide basis (at the corporate headquarters) and on a weekly basis (the chain changes prices no more than once a week). Advertising decisions are done at the chain-level and display decisions are done at the store-level. Promotions decisions (timing and

segment and with the same brand name, which we interact with the corresponding fixed effects. For the GNL model, we also use the number of promoted products of the same type, which we interact we the the corresponding fixed effects.

Second, cost-based instruments, i.e. input prices (Berry, Levinsohn and Pakes, 1995), which are valid under the assumption that input price variations (over time) are correlated with variations in prices, but not with consumers' preferences for unobservable product characteristics. We use input prices for sugar times the sugar content of the cereals, which we interact with fixed effects for segments and for brands, respectively.

A potential serious problem is weak identification, which happens when instruments are only weakly correlated with the endogenous variables. Since there are multiple endogenous variables, the standard first-stage F-statistic is no longer appropriate to test for weak instruments. We therefore use the Sanderson and Windmeijer (2016)'s F-statistic to test whether each particular endogenous variable is weakly identified. A F-statistic higher than 10 suggests that we can be quite confident that instruments are not weak.

5.3 Results

We present now the results from the estimation of the GNL model for different specifications, as well as the ones of the three-level NL models.

5.3.1 Main Specifications

Our main specifications define markets as pairs month-store and products as brands. Following Bresnahan, Stern and Trajtenberg (1997), we model the product fixed effects ξ_j with segment and brand name fixed effects as well as with product characteristics (sugar, fiber, calories, lipid, protein and sodium), and the market fixed effects ξ_t with month fixed effects and store fixed effects.

Table 2 presents parameter estimates of demand. The regressions in columns

format) are done at a chain-level and are globally funded by manufacturers. Since promotional calendar is generally known several weeks in advances, pricing decisions are done conditional on the promotion (that is, promotion is an exogenous variable to pricing).

	(1) GNL	(2) 3NL1	(3) 3NL2
Price ($-\alpha$)	-1.114***	-2.499***	-2.642***
	(0.0896)	(0.118)	(0.130)
Segment/nest (μ_1)	0.608***	0.778***	0.768***
	(0.0102)	(0.00882)	(0.00996)
Company/subnest (μ_2)	0.293***	0.818***	0.807***
	(0.0103)	(0.00715)	(0.00802)
Promotion (β)	0.0704***	0.0924***	0.107***
	(0.00272)	(0.00326)	(0.00348)
Fixed Effects Segments (γ)			
Health/nutrition (γ_H)	-0.647***	-0.876***	-0.0569***
	(0.0110)	(0.00752)	(0.00567)
Kids (γ_K)	-0.435***	-0.554***	0.0336***
	(0.00886)	(0.00868)	(0.00443)
Taste enhanced (γ_T)	-0.683***	-0.926***	-0.0682***
	(0.0114)	(0.00753)	(0.00586)
Fixed Effects Brand Names (θ)			
Kellogg's (θ_K)	0.0541***	-0.0429***	0.160***
	(0.00422)	(0.00340)	(0.00536)
Nabisco (θ_N)	-0.867***	-0.207***	-2.277***
	(0.0275)	(0.0118)	(0.0191)
Post (θ_P)	-0.545***	-0.185***	-1.451***
	(0.0165)	(0.00946)	(0.00858)
Quaker (θ_Q)	-0.573***	-0.308***	-1.511***
	(0.0166)	(0.0150)	(0.00669)
Ralston (θ_R)	-0.871***	-0.228***	-2.382***
	(0.0277)	(0.0131)	(0.0175)
Constant (β_0)	-0.141*	0.221***	-0.102
	(0.0570)	(0.0668)	(0.0678)
Observations	99281	99281	99281
RMSE	0.237	0.267	0.274
F-test for price	464.47	514.32	471.42
F-test for segment/nest	359.01	468.09	467.46
F-test for brand/subnest	326.60	488.31	464.10

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is $\ln(q_{jt}) - \ln(q_{0t})$. Regressions include fixed effects for brand names and segments, months, and stores. Robust standard errors are reported in parentheses. The Sanderson-Windmeijer F statistics are reported for the weak identification test.

Table 2: Parameter Estimates from the Main Specifications

(1) to (3) account for endogeneity using the instruments described in the previous subsection and present parameters estimates from the GNL model, and the three-level NL models with nests for segments on top and with nests for brands on top, respectively.

Consider first the results from the GNL model (hereafter, baseline model). As expected the parameters on the negative of price ($-\alpha$) and on promotion (β) are positive and significant. The estimated nesting parameters ($0 < \mu_i < 1$, $i = 1, 2$), which are consistent with the GE model ($\mu_1 + \mu_2 < 1$), imply that there is indeed market segmentation along both the segment and brand name dimensions. That is, brands with the same (brand) name are closer substitutes than brands with different brand names; and brands within the same segment are closer substitutes than brands within different segments. Overall, products of the same type are closer substitutes.

The advantages provided by the two dimensions are parametrized by the segment and brand name fixed effects (i.e., the γ's and the β's) and by the nesting parameters (i.e., μ_1 and μ_2). The fixed effects measures the extent to which belonging to a nest shifts the demand for the product. The nesting parameters measures the extent to which products within a nest are protected (from competition) from products within different nests along each dimension. In our application, we find that

$$\theta_K > \theta_G = 0 > \gamma_P \simeq \gamma_Q > \gamma_N \simeq \gamma_R,$$
$$\gamma_F = 0 > \gamma_K > \gamma_H \simeq \gamma_T.$$

In other words, the brand-name reputation of the cereals confers a significant advantage on products from General Mills and Kellogg's; and cereals for family also benefit from a significant advantage. In addition, we find that $\mu_1 > \mu_2$, that is, products within the same segment are more protected from products within different segments than products with the same brand name are from products with different brand names.

Turn now to the results from the three-level NL models. They are both consistent with random utility maximization with $\mu_2 > \mu_1$, meaning that the parameters are highly sensitive to the order of nesting and that it is difficult for the researcher to distinguish between them while they give contradictory results. The NL with nests for segments on top implies that consumers perceived the brand name dimension as being more important, while the NL with nests for brand names on top implies the opposite. In constrast, the GNL model by treating the two dimensions symmetrically and independently suggest that the most important dimension from the consumers' point of view is the segment.

5.3.2 Alternative Specifications and Robustness

We consider alternative specifications. First, Table 3 present parameter estimates of demand for several alternative specifications obtained by making slight adjustements compared to the baseline model. The model in column (4) is obtained by removing the store fixed effects, while the model in column (5) is obtained by replacing the segment and brand name fixed effects with the brand fixed effects.[18] In column (6), the model considers products as brand-store combination and months define markets (we end up with 24 markets with more than 4,000 products each).[19] The results from the three alternative specifications do not change qualitatively with respect to the baseline model. However, for model (4), removing the store fixed effects lowers the precision of the estimates; and for model (5), the brand fixed effects reduces the strenght of the instruments. For model (6), this is appealing in the sense that this means that market definition does not drive our results. Running this specification also shows the ability of the GNL model to deal with a high number of products (which is relevant for supermarket and for vertical relations studies).

Second, our baseline model constraints the nesting parameters μ_1 and μ_2 to be equal for all four segments and for all six brand names, respectively. Table 4 presents for the GNL model a more flexible specification in which the parameters are allowed to vary: the parameters μ_1 can take three values (μ_{1F} for family, μ_{1K} for kids, and μ_{1O} for others); the parameters μ_2 can also take three values (μ_{2G} for General Mills, μ_{2K} for Kellogg's, and μ_{2O} for others).

[18] For this specifications, the characteristics-based instruments are interacted with the brand fixed effects instead of with the segment and brand name fixed effects.

[19] We also model the product fixed effects ξ_j with segment, brand name, and store fixed effects as well as with product characteristics (sugar, fiber, calories, lipid, protein and sodium), and the market fixed effects ξ_t with month fixed effects.

	(4) GNL	(5) GNL	(6) GNL
Price ($-\alpha$)	-1.024***	-0.542***	-0.853***
	(0.214)	(0.119)	(0.0308)
Promotion (β)	0.0630***	0.0783***	0.0528***
	(0.00327)	(0.00541)	(0.00101)
Segment/nest (μ_1)	0.623***	0.650***	0.669***
	(0.00947)	(0.0189)	(0.00361)
Company/subnest (μ_2)	0.254***	0.268***	0.258***
	(0.0103)	(0.0192)	(0.00362)
Fixed Effects Segments (γ)			
Health/nutrition (γ_H)		-0.682***	-0.692***
		(0.0205)	(0.00377)
Kids (γ_K)		-0.488***	-0.472***
		(0.0154)	(0.00294)
Taste enhanced (γ_T)		-0.714***	-0.738***
		(0.0212)	(0.00401)
Fixed Effects Brand Names (θ)			
Kellogg's (θ_K)		0.0532***	0.0485***
		(0.00771)	(0.00139)
Nabisco (θ_N)		-0.781***	-0.742***
		(0.0511)	(0.00959)
Post (θ_P)		-0.495***	-0.475***
		(0.0309)	(0.00580)
Quaker (θ_Q)		-0.478***	-0.485***
		(0.0300)	(0.00580)
Ralston (θ_R)		-0.801***	-0.765***
		(0.0512)	(0.0101)
Constant (β_0)	-0.738***	-0.852***	0.177***
	(0.0528)	(0.0251)	(0.0170)
Observations	99281	99281	99281
RMSE	0.441	0.236	0.0831
Stores FE	No	Yes	Yes
Brands FE	No	Yes	No
Products brand-store	No	No	Yes
F-test for price	637.32	35.14	404.09
F-test for segment/nest	361.90	72.90	928.91
F-test for brand/subnest	333.31	67.89	1132.55

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is $\ln(q_{jt}) - \ln(q_{0t})$. Regressions include fixed effects for months. Robust standard errors are reported in parentheses. The Sanderson-Windmeijer F statistics are reported for the weak identification test.

Table 3: Parameter Estimates from the Alternative Specifications

	(7) Unconstrained GNL		
Price ($-\alpha$)	-1.257***	Fixed Effects Segments (γ)	
	(0.115)	Health/nutrition (γ_H)	-0.574***
Promotion (β)	0.0929***		(0.0277)
	(0.00366)	Kids (γ_K)	-0.488***
Nesting Parameters Segments (μ_1)			(0.0284)
Family (μ_{1F})	0.606***	Taste enhanced (γ_T)	-0.610***
	(0.0118)		(0.0270)
Kids (μ_{1K})	0.590***	Fixed Effects Brands (θ)	
	(0.0124)	Kellogg's (θ_K)	0.0278
Others (μ_{1O})	0.644***		(0.0431)
	(0.0122)	Nabisco (θ_N)	-0.605***
Nesting Parameters Brands (μ_2)			(0.0471)
General Mills (μ_{2G})	0.237***	Post (θ_P)	-0.301***
	(0.0128)		(0.0414)
Kellogg's (μ_{2K})	0.231***	Quaker (θ_Q)	-0.333***
	(0.0135)		(0.0419)
Others (μ_{2O})	0.328***	Ralston (θ_R)	-0.648***
	(0.0129)		(0.0426)
Constant (β_0)	0.342***		
	(0.0871)		
Observations		99281	
RMSE		0.243	
Instruments		Yes	
F-test for price	417.45	–	–
F-test for segment	384.28	418.66	463.38
F-test for brand	390.18	306.49	339.69

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: The dependent variable is $\ln(q_{jt}) - \ln(q_{0t})$. The regression includes fixed effects for brand names, segments, months, and stores. Robust standard errors are reported in parentheses. The Sanderson-Windmeijer F statistics are reported for the weak identification test.

Table 4: Parameter Estimates from the Unconstrained GNL Model

The estimated nesting parameters are consistent with the GE model, for all relevant combinations ($\mu_{1F} + \mu_{2G} < 1$, etc.). We find that

$$\theta_G \simeq \theta_K = 0 \simeq \gamma_P \simeq \gamma_Q > \gamma_N \simeq \gamma_R,$$
$$\gamma_F = 0 > \gamma_K > \gamma_H \simeq \gamma_T,$$

which is quite similar to the restricted specification. As for the restricted specification, we find for each product that products within the same segment are more protected from products within different segments than products with the same brand name are from products with different brand names (i.e. $\mu_{1F} > \mu_{2G}$, etc.). Interestingly, we also find that

$$\mu_{1F} \simeq \mu_{1K} \simeq \mu_{1O},$$
$$\mu_{2O} > \mu_{2G} \simeq \mu_{2K},$$

meaning that products from Kellogg's or General Mills are less protected from competition than are other products – this result may reflect the fact that μ_{2O} is associated with four brands while μ_{2G} and μ_{2K}. We note that the parameters are less precisely estimated than in the restricted specification.

5.3.3 Substitution Patterns

Tables 5 and 6 give the estimated average own- and cross-price elasticities of demands for the main specifications of the GNL and three-level NL models.

Our estimated own-price elasticities are in line with the prevalent literature (see e.g., Nevo (2001)). The higher flexibility of the GNL model with respect to the NL models is illustrated by the fact that, for a given product, cross-price elasticities takes as many values as there exists different types of products, while they can take only three values for the three-level NL models – one for products within the same subgroup, one for products within the same group (but not within the same subgroup), and one for products from different groups). In addition, while the three-level NL models assume a priori that products are substitutable, the GNL model allows products to be also complementary. For example, we find, in line with the

Type	Brand	Segment	Own	Cross							
				General Mills				Kellogg's			
				Family 1	Health 2	Kids 3	Taste 4	Family 5	Health 6	Kids 7	Taste 8
1	General Mills	Family	-2.7524	0.2396	0.1179	0.0896	0.0990	0.0792	-0.0425	-0.0708	-0.0613
2	General Mills	Health/nutrition	-2.7123	0.0775	0.3844	0.0760	0.0815	-0.0455	0.2615	-0.0469	-0.0414
3	General Mills	Kids	-2.9885	0.0664	0.0858	0.2999	0.0799	-0.0286	-0.0092	0.2049	-0.0150
4	General Mills	Taste enhanced	-2.5691	0.0688	0.0864	0.0749	0.3592	-0.0375	-0.0199	-0.0313	0.2529
5	Kelloggs	Family	-2.2305	0.0810	-0.0717	-0.0396	-0.0554	0.2035	0.0509	0.0829	0.0671
6	Kelloggs	Health/nutrition	-2.3392	-0.0312	0.2917	-0.0091	-0.0207	0.0365	0.3594	0.0586	0.0470
7	Kelloggs	Kids	-2.9261	-0.0411	-0.0416	0.1615	-0.0267	0.0475	0.0469	0.2500	0.0618
8	Kelloggs	Taste enhanced	-2.1892	-0.0454	-0.0468	-0.0156	0.2704	0.0488	0.0474	0.0786	0.3643
9	Nabisco	Health/nutrition	-1.5646	0.0011	0.0469	0.0007	0.0013	-0.0022	0.0436	-0.0025	-0.0020
10	Post	Health/nutrition	-1.2850	0.0085	0.1959	-0.0528	-0.0300	-0.0104	0.1770	-0.0718	-0.0489
11	Post	Kids	-2.7172	-0.0054	-0.0516	0.0824	-0.0310	0.0054	-0.0407	0.0932	-0.0201
12	Post	Taste enhanced	-1.6185	-0.0019	-0.0532	-0.0503	0.1695	-0.0012	-0.0525	-0.0496	0.1701
13	Quaker	Family	-1.9753	0.0486	-0.0024	-0.0566	-0.0314	0.0426	-0.0084	-0.0626	-0.0374
14	Quaker	Kids	-1.9466	-0.0111	-0.0033	0.0750	-0.0320	-0.0041	0.0037	0.0820	-0.0249
15	Quaker	Taste enhanced	-1.5042	-0.0073	0.0009	-0.0609	0.1574	-0.0118	-0.0035	-0.0653	0.1529
16	Ralston	Family	-2.1511	0.0211	-0.0016	-0.0209	-0.0018	0.0188	-0.0039	-0.0232	-0.0041
17	Ralston	Kids	-2.8539	-0.0224	-0.0018	0.0649	0.0008	-0.0158	0.0048	0.0715	0.0075

Type	Brand	Segment	Cross								
			Nabisco	Post			Quaker			Ralston	
			Health 9	Health 10	Kids 11	Taste 12	Family 13	Kids 14	Taste 15	Family 16	Kids 17
1	General Mills	Family	0.0027	0.0154	-0.0129	-0.0029	0.1250	-0.0249	-0.0156	0.0618	-0.0881
2	General Mills	Health/nutrition	0.0743	0.2293	-0.0791	-0.0692	-0.0031	-0.0046	0.0009	-0.0027	-0.0042
3	General Mills	Kids	0.0013	-0.0692	0.1449	-0.0716	-0.1108	0.1227	-0.0973	-0.0477	0.1857
4	General Mills	Taste enhanced	0.0021	-0.0401	-0.0515	0.2211	-0.0549	-0.0488	0.2355	-0.0035	0.0027
5	Kelloggs	Family	-0.0053	-0.0190	0.0131	-0.0029	0.1109	-0.0096	-0.0255	0.0558	-0.0646
6	Kelloggs	Health/nutrition	0.0772	0.2305	-0.0703	-0.0772	-0.0160	0.0061	-0.0055	-0.0085	0.0136
7	Kelloggs	Kids	-0.0035	-0.0744	0.1288	-0.0557	-0.0972	0.1054	-0.0828	-0.0408	0.1615
8	Kelloggs	Taste enhanced	-0.0036	-0.0660	-0.0348	0.2403	-0.0707	-0.0410	0.2450	-0.0090	0.0208
9	Nabisco	Health/nutrition	0.9458	0.0387	-0.0074	-0.0061	0.0043	0.0040	0.0045	0.0042	0.0039
10	Post	Health/nutrition	0.0529	–	0.3943	0.3936	0.0396	-0.0217	0.0011	0.0173	-0.0439
11	Post	Kids	-0.0074	0.2934	0.4274	0.2982	-0.0313	0.0565	-0.0569	-0.0125	0.0752
12	Post	Taste enhanced	-0.0071	0.3723	0.3752	–	-0.0006	-0.0490	0.1708	0.0151	-0.0333
13	Quaker	Family	0.0040	0.0270	-0.0273	-0.0023	–	0.2631	0.2884	0.0363	-0.0688
14	Quaker	Kids	0.0044	-0.0169	0.0614	-0.0437	0.3035	0.3896	0.2828	-0.0138	0.0722
15	Quaker	Taste enhanced	0.0050	-0.0004	-0.0622	0.1498	0.3413	0.2878	–	0.0133	-0.0402
16	Ralston	Family	0.0036	0.0103	-0.0090	0.0098	0.0323	-0.0097	0.0094	0.8488	0.8055
17	Ralston	Kids	0.0026	-0.0205	0.0462	-0.0170	-0.0461	0.0412	-0.0228	0.6218	–

Notes: Elasticities are averaged over product types and over markets.

Table 5: Average price elasticities for the GNL model

Type	Brand	Segment	3NL1				3NL2			
			Own	Cross			Own	Cross		
				Same sub.	Same grp.	Diff. grp.		Same sub.	Same grp.	Diff. grp.
1	General Mills	Family	-3.4258	0.2134	0.0070	0.1260	-3.3940	0.2392	0.1576	0.0072
2	General Mills	Health/nutrition	-3.1944	0.5721	0.0050	0.2438	-3.3434	0.4168	0.1094	0.0050
3	General Mills	Kids	-3.6561	0.3437	0.0057	0.2045	-3.7271	0.2661	0.1359	0.0062
4	General Mills	Taste enhanced	-3.0961	0.4657	0.0053	0.2680	-3.2491	0.3068	0.1219	0.0056
5	Kelloggs	Family	-2.7657	0.1949	0.0073	0.1307	-2.7648	0.1908	0.1308	0.0077
6	Kelloggs	Health/nutrition	-2.8935	0.3889	0.0051	0.2504	-3.0540	0.2230	0.0935	0.0055
7	Kelloggs	Kids	-3.5288	0.3345	0.0045	0.1626	-3.6096	0.2472	0.0866	0.0051
8	Kelloggs	Taste enhanced	-2.6616	0.4447	0.0052	0.2682	-2.8378	0.2633	0.0983	0.0057
9	Nabisco	Health/nutrition	-2.6312	0.4222	–	0.1514	-1.8272	1.2211	–	0.0032
10	Post	Health/nutrition	-1.7092	–	0.0044	0.2158	-1.4056	–	0.5432	0.0049
11	Post	Kids	-3.4965	0.3283	0.0028	0.1009	-3.2540	0.5644	0.3520	0.0032
12	Post	Taste enhanced	-2.0293	–	0.0036	0.1807	-1.8308	–	0.4065	0.0037
13	Quaker	Family	-2.2777	–	0.0035	0.0623	-2.0227	–	0.3462	0.0035
14	Quaker	Kids	-2.5682	0.2735	0.0031	0.1111	-2.3527	0.4843	0.3323	0.0034
15	Quaker	Taste enhanced	-1.8215	–	0.0036	0.1856	-1.6575	–	0.3742	0.0038
16	Ralston	Family	-3.2704	0.3784	0.0029	0.0518	-2.4825	1.1602	0.8537	0.0026
17	Ralston	Kids	-3.4772	–	0.0022	0.0801	-2.9040	–	0.6971	0.0021

Notes: Elasticities are averaged over product types and over markets.

Table 6: Average price elasticities for the three-level NL models

intuition and as confirmed by the simulations, that products of the same type are always substitutable. We also find that products from General Mills that belong to the family segment are substitutable those from General Mills that belong to the kids segment but complementary with products from Kellogg's that belong to the taste enhanced segment.

References

Anderson, Simon P, André De Palma, and J-F Thisse, "A representative consumer theory of the logit model," *International Economic Review*, 1988, pp. 461–466.

Armstrong, Mark, "A more general theory of commodity bundling," *Journal of Economic Theory*, 2013, *148* (2), 448–472.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile prices in market equilibrium," *Econometrica: Journal of the Econometric Society*, 1995, pp. 841–890.

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

Brenkers, Randy and Frank Verboven, "Liberalizing a distribution system: the European car market," *Journal of the European Economic Association*, 2006, *4* (1), 216–251.

Bresnahan, Timothy F, Scott Stern, and Manuel Trajtenberg, "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s," *RAND Journal of Economics*, 1997, pp. S17–S44.

Davis, Peter and Pasquale Schiraldi, "The flexible coefficient multinomial logit (FC-MNL) model of demand for differentiated products," *The RAND Journal of Economics*, 2014, *45* (1), 32–63.

Deaton, Angus and John Muellbauer, "An almost ideal demand system," *The American economic review*, 1980, *70* (3), 312–326.

_ and _ , *Economics and consumer behavior*, Cambridge university press, 1980.

Fosgerau, Mogens and André de Palma, "Generalized entropy models," *Working Paper HAL Id: hal-01291347*, 2016.

_ and Daniel L McFadden, "A theory of the perturbed consumer with general budgets," Technical Report, National Bureau of Economic Research 2012.

_ , Emerson Melo, André de Palma, and Matthew Shum, "Discrete Choice and Rational Inattention: A General Equivalence Result," *Working Paper HAL Id: hal-01501313*, 2017.

Fudenberg, Drew, Ryota Iijima, and Tomasz Strzalecki, "Stochastic choice and revealed perturbed utility," *Econometrica*, 2015, *83* (6), 2371–2409.

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

Hofbauer, Josef and William H Sandholm, "On the global convergence of stochastic fictitious play," *Econometrica*, 2002, *70* (6), 2265–2294.

Knittel, Christopher R and Konstantinos Metaxoglou, "Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective," *Review of Economics and Statistics*, 2014, *96* (1), 34–59.

Lancaster, Kelvin J, "A new approach to consumer theory," *Journal of political economy*, 1966, *74* (2), 132–157.

Lewbel, Arthur, "Demand Systems with and without Errors," *American Economic Review*, 2001, pp. 611–618.

Mas-Colell, Andreu, Michael Dennis Whinston, Jerry R Green et al., *Microeconomic theory*, Vol. 1, Oxford university press New York, 1995.

Matejka, Filip and Alisdair McKay, "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *The American Economic Review*, 2015, *105* (1), 272–298.

Mattsson, Lars-Göran and Jörgen W Weibull, "Probabilistic choice and procedurally bounded rationality," *Games and Economic Behavior*, 2002, *41* (1), 61–78.

McElroy, Frederick W, "Returns to scale, Euler's theorem, and the form of production functions," *Econometrica: Journal of Econometric Society*, 1969, pp. 275–279.

McFadden, D, "Conditional logit analysis of qualitative choice behavior," *Frontiers in Econometrics*, 1974, pp. 105–142.

McFadden, Daniel, "Modeling the choice of residential location," *Transportation Research Record*, 1978, (673).

— **and Kenneth Train**, "Mixed MNL models for discrete response," *Journal of applied Econometrics*, 2000, pp. 447–470.

Meza, Sergio and K Sudhir, "Do private labels increase retailer bargaining power?," *Quantitative Marketing and Economics*, 2010, *8* (3), 333–363.

Michel, Christian, "Identification and Estimation of Intra-Firm and Industry Competition via Ownership Change," 2016.

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

Ruzhansky, Michael and Mitsuru Sugimoto, "On global inversion of homogeneous maps," *Bulletin of Mathematical Sciences*, 2015, *5* (1), 13–18.

Samuelson, Paul A, "Complementarity: An essay on the 40th anniversary of the Hicks-Allen revolution in demand theory," *Journal of Economic literature*, 1974, *12* (4), 1255–1289.

Sanderson, Eleanor and Frank Windmeijer, "A weak instrument F-test in linear IV models with multiple endogenous variables," *Journal of Econometrics*, 2016, *190* (2), 212–221.

Skrainka, Benjamin S and Kenneth L Judd, "High performance quadrature rules: How numerical integration affects a popular model of product differentiation," 2011.

Verboven, Frank, "International price discrimination in the European car market," *The RAND Journal of Economics*, 1996, pp. 240–268.

_ , "The nested logit model and representative consumer theory," *Economics Letters*, 1996, *50* (1), 57–63.

Appendices

A Appendix for Section 2

A.1 Proof of Lemma 1

To prove Lemma 1, we require the following proposition on homogeneous and homothetic functions:

Proposition. *Let $\phi : \mathbb{R}^n_+ \to \mathbb{R}$ and $F : \mathbb{R} \to \mathbb{R}$ be two continuous and differentiable functions. Define $f : \mathbb{R}^n_+ \to \mathbb{R}$ by $f(x) = F(\phi(x))$, with $x = (x_1, \ldots, x_n)$, and $h : \mathbb{R} \to \mathbb{R}$ by $h(\cdot) = F^{-1}(\cdot)$.*

1. *Euler equation for homogeneous functions. If ϕ is a differentiable function that is homogeneous of degree 1, then*

$$\phi(x) = \sum_{i=1}^{n} \frac{\partial \phi(x)}{\partial x_i} x_i.$$

2. *Generalized Euler equation for homothetic functions (McElroy, 1969). If ϕ is homogeneous of degree 1 and F is non-decreasing, then f is homothetic, and*

$$\sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i} x_i = \frac{h(y)}{h'(y)} \equiv \epsilon(y)\, y.$$

Proof. Let $y = f(x) = F(\phi(x))$ and $h(y) = \phi(x)$, i.e. $h(\cdot) = F^{-1}(\cdot)$, with $x = (x_1, \ldots, x_n)$.

Part 1. See e.g., proof of Theorem M.B.2. in Mas-Colell, Whinston, Green et al. (1995, p. 929).

Part 2. Consider $h(y) = \phi(x)$. Differentiating both side with respect to x_i and rearranging terms, we get

$$\frac{\partial y}{\partial x_i} = \frac{1}{h'(y)} \frac{\partial \phi}{\partial x_i}.$$

Then,

$$\begin{aligned}\epsilon(y) \equiv \sum_{i=1}^{n} \frac{\partial f(x)}{\partial x_i}\frac{x_i}{y} =& \sum_{i=1}^{n} \frac{\partial y}{\partial x_i}\frac{x_i}{y}, \\ =& \sum_{i=1}^{n} \frac{1}{h'(y)}\frac{\partial \phi}{\partial x_i}\frac{x_i}{y}, \\ =& \frac{1}{h'(y)\,y}\sum_{i=1}^{n} \frac{\partial \phi}{\partial x_i}x_i, \\ =& \frac{h(y)}{h'(y)\,y},\end{aligned}$$

where the last equality uses Euler equation for the homogeneous function ϕ. Then, dividing both side by y, we get the required equality. $\square$

We prove Lemma 1 as follows:

Proof of Lemma 1. Define $\phi : \mathbb{R}_{+}^{J+1} \to \mathbb{R}$ and $F : \mathbb{R} \to \mathbb{R}$ by $\phi(q) = S^{(k)}(q)$ and $F(q) = \ln(q)$. Then, $h(\delta) = \exp(\delta)$. Define $\delta = f(q) = F(\phi(q)) = \ln\left(S^{(k)}(q)\right)$ and $h(\delta) = \phi(q) = \exp(\delta)$.

Assume that $\ln \mathbf{J}_{\ln S}$ is symmetric. $S^{(k)}(q)$ is homogeneous of degree 1, then $\ln S^{(k)}(q)$ is homothetic and satisfies

$$\sum_{j=0}^{J} q_j \frac{\partial \ln S^{(k)}(q)}{\partial q_j} = \frac{\exp(\delta)}{\exp(\delta)\,\delta}\delta = 1.$$

By symmetry of $\ln \mathbf{J}_{\ln S}$, we end up with

$$\sum_{j=0}^{J} q_j \frac{\partial \ln S^{(j)}(q)}{\partial q_k} = 1.$$

Assume that $\sum_{j=0}^{J} q_j \frac{\partial \ln S^{(j)}(q)}{\partial q_k} = 1$ holds. Then $\frac{\partial \Omega}{\partial q_j} = -\ln\left(S^{(j)}(q)\right) - 1$ and $\frac{\partial \Omega}{\partial q_k} = -\ln\left(S^{(k)}(q)\right) - 1$, so that $\frac{\partial^2 \Omega}{\partial q_j \partial q_k} = -\frac{\partial \ln\left(S^{(j)}(q)\right)}{\partial q_k}$ and $\frac{\partial^2 \Omega}{\partial q_k \partial q_j} = -\frac{\partial \ln\left(S^{(k)}(q)\right)}{\partial q_j}$. However, Ω is twice continuously differentiable, then by Schwarz's theorem, $\frac{\partial^2 \Omega}{\partial q_j \partial q_k} = \frac{\partial^2 \Omega}{\partial q_k \partial q_j}$, i.e. $\frac{\partial \ln\left(S^{(j)}(q)\right)}{\partial q_k} = \frac{\partial \ln\left(S^{(k)}(q)\right)}{\partial q_j}$, i.e., $\mathbf{J}_{\ln S}$ is symmetric. $\square$

A.2 Proof of Proposition 1

Proof of Proposition 1. Adding-up holds because the budget constraint $\sum_{j=0}^{J} p_j q_j + z \leq y$ is binding.

Homogeneity holds because the prices and income are in real terms, since they are implicitly divided by the price of the *numéraire* good.

Slutsky symmetry follows from Schwarz's theorem. By Roy's identity $q_i = \frac{1}{\alpha}\frac{\partial CS}{\partial p_i}$, then $\frac{\partial q_i}{\partial p_j} = \frac{1}{\alpha^2}\frac{\partial^2 CS}{\partial p_i \partial p_j}$. Similarly, $\frac{\partial q_j}{\partial p_i} = \frac{1}{\alpha^2}\frac{\partial^2 CS}{\partial p_j \partial p_i}$. In addition, the consumer surplus CS being twice continuously differentiable, by Schwarz's theorem $\frac{\partial^2 CS}{\partial p_i \partial p_j} = \frac{\partial^2 CS}{\partial p_j \partial p_i}$, thus showing that $\frac{\partial q_i}{\partial p_j} = \frac{\partial q_j}{\partial p_i}$.

Negativity holds because $(\partial q_i / \partial p_j)$ is the matrix of second derivatives of the concave function CS and so is negative semidefinite. $\square$

A.3 Proof of Proposition 2

Proof of Proposition 2. Consider the Lagrangian of the consumer's program is given by

$$\mathcal{L}(q, \lambda, \mu) = \alpha y + \sum_{j=0}^{J} \delta_j q_j + \Omega(q) + \lambda \left(1 - \sum_{j=0}^{J} q_j\right) + \sum_{j=0}^{J} \lambda_j q_j,$$

where $\lambda \geq 0$ and $\mu_j \geq 0$. Since we are at an interior point, so that $\lambda_j = 0$ for all $j = 0, \ldots, J$, the first-order conditions are given by

$$\begin{aligned} \delta_k + \Omega_k(q) - \lambda &= 0 \quad \text{for } j = 1, \ldots, J \\ \sum_{j=0}^{J} q_j &= 1, \end{aligned}$$

i.e., using Condition 3 in Axiom 1,

$$\begin{aligned}
\delta_k - 1 - \ln S^{(k)}(q) - \lambda =& 0 \quad \text{for } j = 1, \ldots, J \\
\sum_{j=0}^{J} q_j =& 1,
\end{aligned}$$

Let $q_j = q_j(p_1, \ldots, p_J)$ and $\lambda = \lambda(p_1, \ldots, p_J)$ be the optimal quantities and Lagrangian multiplier, respectively. Consider a change in p_1. Totally differentiating the system of first-order conditions given by Equations (**??**) and (**??**) with respect to p_1, we get:

$$\begin{aligned}
\frac{\partial \delta_1}{\partial p_1} - \sum_{j=0}^{J} \frac{\partial \ln S^{(1)}(q)}{\partial q_j} \frac{\partial q_j}{\partial p_1} - \frac{\partial \lambda}{\partial p_1} =& 0, \\
- \sum_{j=0}^{J} \frac{\partial \ln S^{(k)}(q)}{\partial q_j} \frac{\partial q_j}{\partial p_1} - \frac{\partial \lambda}{\partial p_1} =& 0, \quad \text{for } k \neq 1 \\
\sum_{j=0}^{J} \frac{\partial q_j}{\partial p_1} =& 0.
\end{aligned}$$

Let $\mathbf{1}$ be the $(J+1)$ column vector given by $\mathbf{1} = (1, \ldots, 1)^{\intercal}$, and M the $(J+2) \times (J+2)$ matrix given by

$$M = - \begin{bmatrix} \mathbf{J}_{\ln S}(q) & \mathbf{1} \\ \mathbf{1}^{\intercal} & 0 \end{bmatrix}.$$

Then, the system of equations becomes

$$M \begin{bmatrix} \frac{\partial q_1}{\partial p_1} & \cdots & \frac{\partial q_J}{\partial p_1} & \frac{\partial q_0}{\partial p_1} & \frac{\partial \lambda}{\partial p_1} \end{bmatrix}^{\intercal} = \begin{bmatrix} \alpha & \cdots & 0 & 0 & 0 \end{bmatrix}^{\intercal}.$$

Let $\det(M_{ij})$ be the minor of the ith and jth column, i.e., the determinant of the matrix that results from deleting the ith row and jth column of M. Using Cramer's rule, we find that the derivatives of demand are given by

$$\frac{\partial q_j}{\partial p_1} = \alpha \frac{(-1)^{j+1} \det(M_{1j})}{\det(M)}.$$

We want to determine the sign of this derivative. From the second-order conditions of the utility maximization problem, the matrix M is negative definite. Let M^{ts} be the $(t \times s)$ submatrix of M where the first $t \leq J+2$ rows and $s \leq J+2$ columns are retained. Then, by negative definiteness M, we have $(-1)^r \det(M_\pi^{rr}) > 0$ for every $r = 1, \ldots, J+2$ and for every permutation of the indices $\{1, \ldots, J+2\}$.

Then, the quantities$(-1)^{J+1} \det(M_{11})$ and $(-1)^{J+2} \det(M)$ are positive, so that ratio $\det(M_{11}) / \det(M_1)$, is negative. As a consequence, the own-price derivative $\frac{\partial q_1}{\partial p_1} = \alpha \frac{\det(M_{1,1})}{\det(M)}$ is always negative. For the cross-price derivatives, the sign is ambiguous in general, depending on the magnitude of the derivatives of the (logarithm of the) flexible generator S.[20]

Consider in general a change in p_i. Using Laplace expansion, and fixing column j. Therefore

$$\frac{\partial q_j}{\partial p_i} = \alpha \frac{(-1)^{i+j} \det(M_{ij})}{\det(M)}.$$

For $i = j$, $\frac{\partial q_j}{\partial p_j} = \alpha \frac{(-1)^{2j} \det(M_{ij})}{\det(M)} < 0$, while $\frac{\partial q_j}{\partial p_i}$ has an ambiguous sign. □

A.4 Proof of Proposition 3

Let us first give some notations:

- $\sigma_c(j)$ is the set of products that are grouped together with product j on criterion c.

[20]Alternatively, we have $\det(M) = (-1)^{J+1}(J+1)\det(\mathbf{J}_{\ln S})$ and $\det(M_{11}) = (-1)^J J \det((\mathbf{J}_{\ln S})_{11})$. By positive definiteness of $\mathbf{J}_{\ln S}$ (Condition 2), we have $\det(\mathbf{J}_{\ln S}) > 0$ anddet$((\mathbf{J}_{\ln S})_{11}) > 0$. However, $\det(M_{1j})$ has an ambiguous sign, depending on the magnitude of the derivatives of the (logarithm of the) flexible generator S. As a consequence, the own-price derivative $\frac{\partial q_1}{\partial p_1} = \alpha \frac{\det(M_{1,1})}{\det(M)}$ is always negative, while the cross-price derivatives $\frac{\partial q_j}{\partial p_1} = \alpha \frac{(-1)^{j+1}\det(M_{1j})}{\det(M)}$ can be either positive or negative, meaning that the model allows for substitutability and complementarity.

- The $J \times J$ matrix Θ_c is the nesting structure matrix for criterion c, i.e.,

$$(\Theta_c)_{ij} = \begin{cases} 1, & \text{if } i \in \sigma_c(j), \\ 0, & \text{otherwise.} \end{cases}$$

The logit model arises when for all c, $\Theta_c = I$, and the nested logit model arises when for all c, but one, $\Theta_c = I$.

- The $J \times J$ matrix $\mathbf{J}_F(x)$ is the Jacobian of F.
- The symbol $\circ$ denotes the Hadamard product.
- I_J denotes the $J \times J$ identity matrix, and J_J denotes the $J \times J$ unit matrix (where every element is equal to one).

Proof of Proposition 3. Consider $\varepsilon = \Sigma^\intercal$, the transpose matrix of the own- and cross-price elasticities $\Sigma_{ij} = \frac{\partial q_j}{\partial p_i} \frac{p_i}{q_j}$ (recall that the outside option is just a good whose net utility is normalized to zero, so that it is not necessary to deal with it directly in the proof). Its diagonal elements $(\varepsilon)_{jj} = \varepsilon_{jj}$ are the own-price elasticities given by

$$\varepsilon_{jj} = \frac{\partial q_j}{\partial p_j} \frac{p_j}{q_j} = \left[\frac{H_j^{(j)}\left(e^\delta\right)}{H^{(j)}\left(e^\delta\right)} - \sum_{i=1}^{J} \frac{H_j^{(i)}\left(e^\delta\right)}{\sum_{i=1}^{J} H^{(i)}\left(e^\delta\right)} \right] \frac{\partial e^{\delta_j}}{\partial \ln p_j},$$

and its off-diagonal elements $(\varepsilon)_{jk} = \varepsilon_{jk}$ are the cross-price elasticities given by

$$\varepsilon_{jk} = \frac{\partial q_j}{\partial p_k} \frac{p_k}{q_j} = \left[\frac{H_k^{(j)}\left(e^\delta\right)}{H^{(j)}\left(e^\delta\right)} - \sum_{i=1}^{J} \frac{H_k^{(i)}\left(e^\delta\right)}{\sum_{i=1}^{J} H^{(i)}\left(e^\delta\right)} \right] \frac{\partial e^{\delta_k}}{\partial \ln p_k},$$

and equivalently, using $q_j = H^{(j)}\left(e^\delta\right) / \sum_{i=1}^{J} H^{(i)}\left(e^\delta\right)$ and $\frac{\partial e^{\delta_j}}{\partial \ln p_j} = -\alpha p_j e^{\delta_j}$,

$$\begin{aligned} \varepsilon_{jj} &= -\alpha p_j e^{\delta_j} \left[\frac{H_j^{(j)}\left(e^\delta\right)}{H^{(j)}\left(e^\delta\right)} - \sum_{i=1}^{J} q_i \frac{H_j^{(i)}\left(e^\delta\right)}{H^{(i)}\left(e^\delta\right)} \right], \\ \varepsilon_{jk} &= -\alpha p_k e^{\delta_k} \left[\frac{H_k^{(j)}\left(e^\delta\right)}{H^{(j)}\left(e^\delta\right)} - \sum_{i=1}^{J} q_i \frac{H_k^{(i)}\left(e^\delta\right)}{H^{(i)}\left(e^\delta\right)} \right]. \end{aligned}$$

After some computations, we find that

$$\varepsilon = -\alpha \left(\text{diag}\left(1/q\right) - J_J\right) \cdot \text{diag}\left(q\right) \cdot \mathbf{J}_{\ln H}\left(e^{\delta}\right) \cdot \text{diag}\left(pe^{\delta}\right).$$

Now, using the inverse function theorem, we get

$$\begin{aligned}
\mathbf{J}_{\ln H}\left(e^{\delta}\right) \equiv \mathbf{J}_{\ln S^{-1}}\left(e^{\delta}\right) &= \mathbf{J}_{\ln}\left(S^{-1}\left(e^{\delta}\right)\right) \cdot \mathbf{J}_{S^{-1}}\left(e^{\delta}\right), \\
&= \mathbf{J}_{\ln}\left(S^{-1}\left(e^{\delta}\right)\right) \cdot \left[\mathbf{J}_S\left(S^{-1}\left(e^{\delta}\right)\right)\right]^{-1}, \\
&= \mathbf{J}_{\ln}\left(H\left(e^{\delta}\right)\right) \cdot \left[\mathbf{J}_S\left(H\left(e^{\delta}\right)\right)\right]^{-1},
\end{aligned}$$

where $\mathbf{J}_{\ln}\left(H\left(e^{\delta}\right)\right) = \text{diag}\left(H\left(e^{\delta}\right)\right)^{-1} = \text{diag}\left(1/q\right) \cdot \text{diag}\left(1/\mathbf{H}\right)$. By homogeneity of S, $\mathbf{H}$ can be considered as a positive constant; and since S is homogeneous of degree 1, its derivatives are of degree 0, so that $\mathbf{J}_S\left(H\left(e^{\delta}\right)\right) = \mathbf{J}_S\left(q\right)$. In addition, from Theorem 1, we have $e^{\delta}/\mathbf{H} = S\left(q\right)$, so that

$$\begin{aligned}
\Sigma \equiv \varepsilon^{\intercal} &= \left(-\alpha \left(\text{diag}\left(1/q\right) - J_J\right) \cdot \left[\mathbf{J}_S\left(q\right)\right]^{-1} \cdot \text{diag}\left(p \cdot S\left(q\right)\right)\right)^{\intercal}, \\
&= -\alpha \text{diag}\left(p \cdot S\left(q\right)\right) \cdot \left[\mathbf{J}_S\left(q\right)^{\intercal}\right]^{-1} \cdot \left(\text{diag}\left(1/q\right) - J_J\right).
\end{aligned}$$

Note that the matrix of own- and cross-price derivatives is given by

$$\nabla_q\left(p\right) = \text{diag}\left(1/p\right) \cdot \Sigma \cdot \text{diag}\left(q\right).$$

$\square$

B Appendix for Section 3

B.1 Proof of Proposition 4

Proof of Proposition 4. *Part 1* shows that the three-level NL can be written as a GNL model, with $S^{(j)}\left(q\right) = q_j^{\mu_0} q_{\sigma_1(j)}^{\mu_1} q_{\sigma_2(j)}^{\mu_2} = q_j^{\mu_0} q_g^{\mu_1} q_{hg}^{\mu_2}$, for $j = 0, \ldots, J$, with $\mu_0 + \mu_1 + \mu_2 = 1$, where the $\sigma_{2(j)}$'s partition the $\sigma_{1(j)}$'s, i.e., the choice set is partitioned into G nests, where each nest g is partitioned into H_g subnests, where each subnests h contains J_{hg} products.

The equation $S(\tilde{q}) = e^{\delta}$ has solution

$$\tilde{q}_j = \exp(\delta_j/\mu_0) \left[\sum_{j=1}^{J_{hg}} \exp(\delta_j/\mu_0)\right]^{-\frac{\mu_1}{1-\mu_2}} \left[\sum_{h=1}^{H_g} \left(\sum_{j=1}^{J_{hg}} \exp(\delta_j/\mu_0)\right)\right]^{-\mu_1}. \quad (12)$$

Indeed, $S(\tilde{q}) = e^{\delta}$ writes $\exp(\delta_j/\mu_0) = \tilde{q}_j \tilde{q}_g^{\mu_1/\mu_0} \tilde{q}_{hg}^{\mu_2/\mu_0}$. Summing over products j in a subnest h of nest g, we get

$$\sum_{j=1}^{J_{hg}} \exp(\delta_j/\mu_0) = \tilde{q}_g^{\mu_1/\mu_0} \tilde{q}_{hg}^{1+\mu_2/\mu_0} = \tilde{q}_g^{\mu_1/\mu_0} \tilde{q}_{hg}^{(1-\mu_1)/\mu_0},$$

where the last equality comes from the constraint $\mu_0 + \mu_1 + \mu_2 = 1$. Rearranging the last equation, we obtain

$$\left(\sum_{j=1}^{J_{hg}} \exp(\delta_j/\mu_0)\right)^{\mu_0/(1-\mu_1)} = \tilde{q}_g^{\mu_1/(1-\mu_1)} \tilde{q}_{hg},$$

and summing over subnests h of g, we get

$$\sum_{h=1}^{H_g} \left(\sum_{j=1}^{J_{hg}} \exp(\delta_j/\mu_0)\right)^{\mu_0/(1-\mu_1)} = \tilde{q}_g^{1+\mu_1/(1-\mu_1)} = \tilde{q}_g^{1/(1-\mu_1)}.$$

Therefore, we have

$$\tilde{q}_{hg} = \tilde{q}_g^{-\mu_1/(1-\mu_1)} \left(\sum_{j=1}^{J_{hg}} \exp(\delta_j/\mu_0)\right)^{\mu_0/(1-\mu_1)},$$

$$\tilde{q}_g = \left[\sum_{h=1}^{H_g} \left(\sum_{j=1}^{J_{hg}} \exp(\delta_j/\mu_0)\right)^{\mu_0/(1-\mu_1)}\right]^{1-\mu_1}.$$

Plugging these formula into $\tilde{q}_j = e^{\delta_j/\mu_0} \tilde{q}_g^{-\mu_1/\mu_0} \tilde{q}_{hg}^{-\mu_2/\mu_0}$, we get the solution (12). Then, normalizing the sum of demands to 1, i.e., $\sum_{g=1}^{G} q_g = 1$, and after rearranging terms, the normalized demand $q_j = \tilde{q}_j / \sum_i \tilde{q}_i$ is given by

$$q_j = \frac{\exp\left(\delta_j/\mu_0\right)}{\exp\left(I_{hg}/\mu_0\right)} \frac{\exp\left(I_{hg}/\left(1-\mu_1\right)\right)}{\exp\left(I_g/\left(1-\mu_1\right)\right)} \frac{\exp\left(I_g\right)}{\exp\left(I\right)},$$

where the inclusives values I_{hg}, I_g and I are defined as follows

$$\begin{aligned}
I_{hg} =& \mu_0 \ln\left(\sum_{j=1}^{J_{hg}} \exp\left(\delta_j/\mu_0\right)\right), \\
I_g =& \left(1-\mu_1\right) \ln\left(\sum_{h=1}^{H_g} \exp\left(I_{hg}/\left(1-\mu_1\right)\right)\right), \\
I =& \ln\left(\sum_{g=1}^{G} \exp\left(I_g\right)\right).
\end{aligned}$$

Part 3 shows that the GNL model allows products to be complementary, independent, or substitutable. To do so, we consider a particular case with 3 inside products and one outside option. Inside products are grouped according two dimensions. For the first dimension, product 1 is in one nest, and products 2 and 3 are in a second nest. For the second dimension, products 1 and 2 are in one nest, and product 3 is in a second nest.

Totally differentiating the first-order conditions of the utility maximization program with respect to price p_3 gives following the system in matrix form:

$$M \begin{bmatrix} \frac{\partial q_1^*}{\partial p_3} & \frac{\partial q_2^*}{\partial p_3} & \frac{\partial q_3^*}{\partial p_3} & \frac{\partial q_0^*}{\partial p_3} & \frac{\partial \lambda^*}{\partial p_3} \end{bmatrix}^\intercal = \begin{bmatrix} 0 & 0 & \alpha & 0 & 0 \end{bmatrix}^\intercal,$$

where the matrix M is given by

$$M = \begin{bmatrix}
0 & 0 & 0 & -1/q_0 & -1 \\
-\frac{\mu_0}{q_1} - \frac{\mu_1}{q_{\sigma_1(1)}} - \frac{\mu_2}{q_{\sigma_2(1)}} & -\frac{\mu_2}{q_{\sigma_2(1)}} & 0 & 0 & -1 \\
-\frac{\mu_2}{q_{\sigma_2(2)}} & -\frac{\mu_0}{q_2} - \frac{\mu_1}{q_{\sigma_1(2)}} - \frac{\mu_2}{q_{\sigma_2(2)}} & -\frac{\mu_1}{q_{\sigma_1(2)}} & 0 & -1 \\
\frac{\mu_1}{q_{\sigma_1(3)}} & -\frac{\mu_0}{q_3} - \frac{\mu_1}{q_{\sigma_1(3)}} - \frac{\mu_2}{q_{\sigma_2(3)}} 0 & 0 & 0 & -1 \\
-1 & -1 & -1 & -1 & 0
\end{bmatrix}.$$

After some matrix manipulations, we find that the cross-derivative $\frac{\partial q_1}{\partial p_3} = \frac{\partial q_3}{\partial p_1} = \alpha \frac{\det(M_{13})}{\det(M)}$ is negative for δ that satisfies

$$\mu_0 \left(q_1 (\delta) + q_2 (\delta)\right) \left(q_2 (\delta) + q_3 (\delta)\right) - \mu_1 \mu_2 q_0 (\delta) \, q_2 (\delta) < 0.$$

. $\square$

B.2 Proof of Proposition 5

Recall that $\sigma_c (j)$ is the set of products that are grouped together with product j on dimension c. The $(J + 1) \times (J + 1)$ matrix Θ_c is the nesting structure matrix for dimension c, i.e.,

$$(\Theta_c)_{i+1,j+1} = \begin{cases} 1, & \text{if } i \in \sigma_c (j) \, , \\ 0, & \text{otherwise.} \end{cases}$$

In the GNL model, the outside option has the special property that $(\Theta_c)_{i+1,1} = 0$ and $(\Theta_c)_{1,j+1} = 0$. To ease exposition, we focus on the matrix of own- and cross-price elasticities of the inside goods alone in the text. However, we consider the inside good in the proof.

Proof of Proposition 5 . Let q_0 be the market share for the outside option and $q = (q_1, \ldots, q_J)$ the vector of the market shares for the inside goods, and recall that $q_{\sigma_c(j)} = \sum_{i \in \sigma_c(j)} q_j$. Then, for all $j > 0$,

$$\frac{\partial S^{(j)} (q)}{\partial q_j} = S^{(j)} (q) \left(\frac{\mu_0}{q_j} + \sum_{c=1}^{C} \frac{\mu_c}{q_{\sigma_c(j)}} \right), \; \frac{\partial S^{(j)} (q)}{\partial q_k} = S^{(j)} (q) \sum_{c=1}^{C} \frac{\mathbf{1}_{\{k \in \sigma_c(j)\}} \mu_c}{q_{\sigma_c(j)}},$$

and

$$\frac{\partial S^{(0)} (q)}{\partial q_j} = 0, \; \frac{\partial S^{(j)} (q)}{\partial q_0} = 0, \; \frac{\partial S^{(0)} (q)}{\partial q_0} = 1,$$

so that the Jacobian matrix of S is given by

$$\mathbf{J}_S (q) = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0}' & \tilde{\mathbf{J}}_S (q) \end{pmatrix},$$

with the $(1 \times J)$ row vector $\mathbf{0} = (0, \ldots, 0)$ and $\tilde{\mathbf{J}}_S(q) = \left[\frac{\partial S^{(i)}(q)}{\partial q_j}\right]_{i>0, j>0}$ is the $(J \times J)$ matrix given by

$$\tilde{\mathbf{J}}_S(q) = \text{diag}\left(\frac{S(q)}{q}\right) \cdot \left[\mu_0 I_J + \sum_{c=1}^{C} \mu_c Q_{\sigma_c} \cdot \Theta_c\right],$$

where Q_{σ_c} is the diagonal matrix of the market shares of the inside goods in their nest σ_c, i.e. $(Q_{\sigma_c})_{jj} = \frac{q_j}{q_{\sigma_c(j)}}$.

Using block matrix results and using the derived formula of ε in Appendix A.4,

$$\begin{aligned}
\varepsilon = & -\alpha \begin{pmatrix} \frac{1-q_0}{q_0} & -\mathbf{1} \\ -\mathbf{1}^\intercal & \text{diag}(1/q) - J_J \end{pmatrix} \cdot \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0}^\intercal & \tilde{\mathbf{J}}_S(q)^{-1} \end{pmatrix} \cdot \begin{pmatrix} p_0 q_0 & \mathbf{0} \\ \mathbf{0}^\intercal & \text{diag}(p \cdot S(q)) \end{pmatrix}, \\
= & -\alpha \begin{pmatrix} \frac{1-q_0}{q_0} & -\mathbf{1} \cdot \tilde{\mathbf{J}}_S(q)^{-1} \\ -\mathbf{1}^\intercal \cdot q_0 & (\text{diag}(1/q) - J_J) \cdot \tilde{\mathbf{J}}_S(q)^{-1} \end{pmatrix} \cdot \begin{pmatrix} p_0 q_0 & \mathbf{0} \\ \mathbf{0}^\intercal & \text{diag}(p \cdot S(q)) \end{pmatrix}, \\
= & -\alpha \begin{pmatrix} p_0(1-q_0) & -\mathbf{1} \cdot \tilde{\mathbf{J}}_S(q)^{-1} \cdot \text{diag}(p \cdot S(q)) \\ -\mathbf{1}^\intercal \cdot p_0 q_0 & (\text{diag}(1/q) - J_J) \cdot \tilde{\mathbf{J}}_S(q)^{-1} \cdot \text{diag}(p \cdot S(q)) \end{pmatrix}.
\end{aligned}$$

Then focussing on the inside goods, we have

$$\begin{aligned}
\Sigma \equiv \varepsilon^\intercal = & \left(-\alpha(\text{diag}(1/q) - J_J) \cdot \tilde{\mathbf{J}}_S(q)^{-1} \cdot \text{diag}(p \cdot S(q))\right)^\intercal, \\
= & \left(-\alpha(\text{diag}(1/q) - J_J) \cdot \left[\mu_0 I_J + \sum_{c=1}^{C} \mu_c Q_{\sigma_c} \cdot \Theta_c\right]^{-1} \cdot \text{diag}(p \cdot q)\right)^\intercal, \\
& -\alpha \text{diag}(p \cdot q) \cdot \left[\mu_0 I_J + \sum_{c=1}^{C} \mu_c Q_{\sigma_c} \cdot \Theta_c\right]^{-1} \cdot (\text{diag}(1/q) - J_J).
\end{aligned}$$

□

B.3 Numerical Properties of the GNL model

We cannot obtain an analytic formula for each element of the matrix of own- and cross-price elasticities independently. We therefore use Monte Carlo experiments to understand substitution patterns. To do so, we simulate NS different nesting structures (i.e. allocations of products in nests) along C dimensions (with M modalities per dimension), NS different vectors of nesting parameters $\mu = (\mu_0, \mu_1, \ldots, \mu_C)$, and NS different vectors of market shares $q = (q_0, \ldots, q_J)$. We set $NS = 20$, $C = 6$, $M = 3$, and $J = 30$, and by combining these dimensions, we end up with 8,000 market structures. We obtain

- a nesting structure by simulating a $NS \times C$ matrix of binomial random numbers;
- a vector of nestings parameters by simulating a $(C + 1) \times 1$ vector of uniformly distributed random numbers, where the first element thus obtained is μ_0, and by normalizing the vector consisting of the other nesting parameters to get a unit vector μ;
- a vector of market shares by simulating a $(J + 1) \times 1$ vector of uniformly distributed random numbers, where the first element thus obtained is q_0, and by normalizing the vector of market shares of inside products to get a unit vector q.[21]

The following table gives summary statistics on the simulated data:

Variable	Mean	Min	Max
q_0	0.5249	0.0000	0.9906
q	0.0158	3e-06	0.0697
μ_0	0.4662	0.0697	0.9532
μ_1	0.1040	0.0062	0.3298
μ_2	0.0736	0.0080	0.2515
μ_3	0.0965	0.0023	0.2691
μ_4	0.1009	0.0119	0.2303
μ_5	0.0766	0.0011	0.2557
μ_6	0.0822	5e-05	0.2588

Table 7: Summary statistics on the simulated data

[21] We do this normalization to obtain market structures with very low and very high values for μ_0 and q_0.

Nesting structure. Table 8 shows the distribution of the own- and cross-price derivatives for the simulated data.

Same nests	Deriv > 0	Median	Min	Max	Percent
Own-price derivatives					
–	0.00%	-0.0219	-0.6760	-3e-06	100.00%
Cross-price derivatives					
0 (None)	39.22%	-9e-06	-0.1113	0.0104	5.70%
1	66.95%	2e-05	-0.0869	0.1230	21.80%
2	88.88%	0.0002	-0.07832	0.1787	32.03%
3	98.05%	0.0005	-0.06220	0.1711	25.41%
4	99.86%	0.0007	-0.01383	0.2435	12.00%
5	100.00%	0.0009	3e-09	0.2517	2.74%
6 (All)	100.00%	0.0011	1e-08	0.2841	0.31%
Total	85.26%	0.0002	-0.1113	0.2841	100.00%

Table 8: Distribution of price derivatives according to the number of common nests

Own-price elasticities are always negative, while cross-price elasticities can be either negative (complementarity) or positive (substitutability). Products of the same type are always substitutable. As products become different, the share of substitutability decreases. Products that are very similar (that are grouped together according to all the dimensions, but one) are always substitutable too. However, products that are completely different can be either substitutable or complementary. To summarize, complementarity may or may not arise for products that are different according to their dimensions, while products of the same type are always substitutable.

Size of the outside option (q_0). Table 9 shows the percentage of positive cross-price derivatives (substitutes) according to the size of the outside option.

q_0	Deriv > 0	q_0	Deriv > 0
0.00%	88.19%	53.42%	86.12%
4.80%	87.82%	63.39%	85.74%
8.99%	87.82%	77.40%	84.59%
13.37%	87.66%	78.09%	84.45%
14.22%	87.65%	79.78%	84.33%
16.65%	87.55%	88.89%	83.42%
17.97%	87.39%	94.35%	82.15%
47.75%	86.40%	96.94%	80.73%
48.13%	86.32%	98.33%	80.36%
48.33%	86.38%	99.06%	80.02%

Table 9: Percentage of substitutable products according to the size of the outside option

One concern is that a large outside option could generate complementarity. Table 9 shows the size of the outside option does not matter for the purpose considered here, at least for intermediate values: with the simulated data, the percentage of substitutable products is equal to 87% for $q_0 = 23\%$, to 86% for $q_0 = 56\%$, and to 82% for $q_0 = 84\%$. However, at the extremes, a higher value value of the outside option is associated with a higher proportion of positive derivatives, that is, of complementary products.

Nesting parameters. Table 10 shows the distribution of cross-price derivative according to the level of the closeness of products, as measured by the sum of nesting parameters $\mu_{jk} = \sum_{c=1}^{6} \mu_i I \{j \in \sigma_c (k)\}$ for two products j and k.

μ_{jk}	Deriv > 0	Median	Min	Max
$[0, 0.1[$	77.93%	4e-05	-0.1113	0.0669
$[0.1, 0.2[$	85.78%	0.0001	-0.0869	0.1009
$[0.2, 0.3[$	84.64%	0.0003	-0.0864	0.1230
$[0.3, 0.4[$	92.11%	0.0008	-0.0622	0.1787
$[0.4, 0.5[$	97.08%	0.0016	-0.0394	0.1583
$[0.5, 0.6[$	99.52%	0.0028	-0.0237	0.1711
$[0.6, 0.7[$	99.81%	0.0039	-0.0121	0.2095
$[0.7, 0.8[$	100.00%	0.0059	2e-07	0.2435
$[0.8, 0.9[$	100.00%	0.0086	2e-07	0.2517
$[0.9, 1[$	100.00%	0.0187	8e-06	0.2841

Table 10: Distribution of cross-price derivatives according to the value of μ_{jk}

As the parameter μ_{jk} increases, we observe first that the size of the derivatives decreases in their negatives values, and increases in their positive values; then that the share of substitutable products increases. This comes from the fact that a higher value of μ_{ik} indicates that products j and k are perceived as more similar.

C Appendix for Section 5

C.1 Data

Nb.	Brand	Brand name	Segment	Shares	(%)
				Dollars	Volume
1	Apple Cinnamon Cheerios	General Mills	Family	2.23	2.02
2	Cheerios	General Mills	Family	7.67	6.76
3	Clusters	General Mills	Family	1.03	0.89
4	Golden Grahams	General Mills	Family	2.28	2.12
5	Honey Nut Cheerios	General Mills	Family	4.82	4.47
6	Total Corn Flakes	General Mills	Family	0.87	0.59
7	Wheaties	General Mills	Family	2.59	2.75
8	Total	General Mills	Health/nutrition	1.29	1.00
9	Total Raisin Bran	General Mills	Health/nutrition	1.61	1.49
10	Cinnamon Toast Crunch	General Mills	Kids	2.16	1.94
11	Cocoa Puffs	General Mills	Kids	1.22	0.98
12	Kix	General Mills	Kids	1.68	1.29
13	Lucky Charms	General Mills	Kids	2.35	1.94
14	Trix	General Mills	Kids	2.43	1.75
15	Oatmeal (Raisin) Crisp	General Mills	Taste enhanced	2.05	2.09
16	Raisin Nut	General Mills	Taste enhanced	1.60	1.60
17	Whole Grain Total	General Mills	Taste enhanced	1.77	1.29
18	All Bran	Kellogg's	Family	0.97	1.11
19	Common Sense Oat Bran	Kellogg's	Family	0.49	0.46
20	Corn Flakes	Kellogg's	Family	4.12	6.96
21	Crispix	Kellogg's	Family	1.88	1.70
22	Frosted Flakes	Kellogg's	Family	6.01	6.77
23	Honey Smacks	Kellogg's	Family	0.85	0.84
24	Rice Krispies	Kellogg's	Family	5.58	6.06
25	Bran Flakes	Kellogg's	Health/nutrition	0.90	1.16
26	Frosted Mini-Wheats	Kellogg's	Health/nutrition	3.35	3.69
27	Product 19	Kellogg's	Health/nutrition	1.06	0.86
28	Special K	Kellogg's	Health/nutrition	3.07	2.53
29	Apple Jacks	Kellogg's	Kids	1.67	1.32
30	Cocoa Krispies	Kellogg's	Kids	0.99	0.85
31	Corn Pops	Kellogg's	Kids	1.80	1.52
32	Froot Loops	Kellogg's	Kids	2.66	2.22
33	Cracklin' Oat Bran	Kellogg's	Taste enhanced	1.91	1.66
34	Just Right	Kellogg's	Taste enhanced	1.07	1.12
35	Raisin Bran	Kellogg's	Taste enhanced	3.96	4.83
36	Shredded Wheat	Nabisco	Health/nutrition	0.77	0.88
37	Spoon Size Shredded Wheat	Nabisco	Health/nutrition	1.59	1.63
38	Grape Nuts	Post	Health/nutrition	2.27	3.06
39	Cocoa Pebbles	Post	Kids	1.11	0.92
40	Fruity Pebbles	Post	Kids	1.14	0.94
41	Honey-Comb	Post	Kids	1.05	0.90
42	Raisin Bran	Post	Taste enhanced	0.93	1.10
43	Oat Squares	Quaker	Family	0.91	1.02
44	CapNCrunch	Quaker	Kids	1.00	1.10
45	Jumbo Crunch (Cap'n Crunch)	Quaker	Kids	1.27	1.35
46	Life	Quaker	Kids	1.73	2.24
47	100% Cereal	Quaker	Taste enhanced	1.42	1.84
48	Corn Chex	Ralston	Family	0.81	0.72
49	Rice Chex	Ralston	Family	1.15	1.03
50	Cookie-Crisp	Ralston	Kids	0.89	0.68

Table 11: Top 50 Brands