Discussion Papers
Department of Economics
University of Copenhagen

No. 18-02

Demand Models for Differentiated Goods with Complementarity and Substitutability

Mogens Fosgerau, André de Palma and Julien Monardo

Øster Farimagsgade 5, Building 26, DK-1353 Copenhagen K., Denmark
Tel.: +45 35 32 30 01 – Fax: +45 35 32 30 00
http://www.econ.ku.dk

ISSN: 1601-2461 (E)

Demand Models for Differentiated Goods with Complementarity and Substitutability*

Mogens Fosgerau[†] André de Palma[‡] Julien Monardo[§]

March 14, 2018

Abstract

We develop a class of demand models for differentiated products. The new models facilitate the BLP method (Berry et al., 1995) while numerical inversion of the demand system is not required. They can accommodate rich patterns of substitution and complementarity while being easily estimated with standard regression techniques and allowing very large choice sets. We use the new models to describe markets for differentiated products that exhibit segmentation according to several dimensions and illustrate their application by estimating demand for cereals in Chicago.

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

*First Draft: March, 2016. We are grateful for comments from Xavier D'Haultfoeuille, Laurent Linnemer, Yurii Nesterov, Bernard Salanié, Thibaud Vergé, Tatiana Babicheva, Sophie Dantan and Hugo Molina as well as participants at the conference on Advances in Discrete Choice Models in honor of Daniel McFadden at the University of Cergy-Pontoise, and seminars participants at the Tinbergen Institute, the University of Copenhagen, Northwestern University, Université de Montréal, CREST-LEI, KU Leuven and ENS Paris-Saclay. We have received financial support from the Danish Strategic Research Council, iCODE, University Paris-Saclay, and the ARN project (Elitisme). Mogens Fosgerau is supported by the ERC Advanced Grant GEM 740369. This paper is a continuation of the paper "Demand system for market shares". that was first circulated in 2016.

[†]University of Copenhagen; mogens.fosgerau@econ.ku.dk.
[‡]CREST, ENS Paris-Saclay, University Paris-Saclay; andre.depalma@ens-paris-saclay.fr.
[§]CREST, ENS Paris-Saclay, University Paris-Saclay; jmonardo@ens-paris-saclay.fr.

1 Introduction

This paper develops a class of discrete choice demand models, applicable for estimating the demand for differentiated products, using the BLP method (Berry et al., 1995) to handle endogeneity issues, while avoiding the numerical inversion of the demand system. The new models are capable of accommodating rich patterns of substitution and complementarity.[1] Nevertheless, they may be estimated using just standard regression techniques, which means that it is feasible to handle very large choice sets.

The new models build on new insights regarding the relationship between the additive random utility model (ARUM) and the perturbed utility model (PUM). ARUM rely on the single-unit purchase assumption that each consumer buys one unit of the alternative that gives her the highest utility and impose the structure that utilities are the sum of deterministic and random utility terms.[2] In contrast, PUM assume that each consumer chooses the probability distribution over the alternatives that maximizes her utility given by the sum of an expected utility and a perturbation, which is a nonlinear concave function.[3]

Despite this fundamental difference, the two models are closely linked. Hofbauer and Sandholm (2002) showed that the choice probabilities generated by any ARUM can be derived from a PUM with a deterministic perturbation. The concept of entropy plays an important role in this relationship: it is well known that the logit probabilities can be obtained from a PUM when the perturbation is the Shannon entropy (Anderson et al., 1988); and similarly, that the nested logit probabilities can be obtained using an entropy-type perturbation (Verboven, 1996).[4]

In this paper, we develop this relationship, defining a class of generalized entropies (GE) that can serve as perturbations in the PUM. GE generalize the Shannon entropy by

[1] In this respect, they possess the main features making them appealing for merger evaluation and studying vertically markets, as highlighted by Pinkse and Slade (2004).

[2] ARUM have been widely used in the empirical industrial organization literature since the seminal paper of McFadden (1974).

[3] PUM have been used to model optimization with effort (Mattsson and Weibull, 2002), stochastic choices (Fudenberg et al., 2015) and rational inattention (Matejka and McKay, 2015; Fosgerau et al., 2017).

[4] The concept of entropy was invented by Rudolf Clausius in 1865 in the field of thermodynamics and was first introduced in economics by Podolinsky in 1880. Since Shannon (1948), the Shannon entropy and its generalizations has found applications in several fields of economics (e.g., economic growth (Georgesçu-Roegen, 1971), transport economics (Erlander, 1977), income inequality (Shorrocks, 1980), social choice and inequality (Cowell, 2000), decision theory (Mattsson and Weibull, 2002), demographic economics (Edwards and Tuljapurkar, 2005), urban and regional economics (Wilson, 2011), rational inattention (Matejka and McKay, 2015), and international trade (Mrazova et al., 2017)). The concept of entropy also appears in Galichon and Salanié (2015) who study matching models with transferable utility and unobserved heterogeneity, and in Chiong et al. (2016) who study identification and estimation of dynamic ARUM.

relaxing its symmetry property and take the form $\Omega(\mathbf{q}) = -\mathbf{q}^{\intercal} \ln \mathbf{S}(\mathbf{q})$, with $\mathbf{q}$ being a vector of choice probabilities and $\mathbf{S}$ a function that satisfies some mild conditions. GE models (GEM) are thus a special kind of PUM in which the perturbation is specified to be a GE.

The class of GEM is large. We show that we can always find a GEM that leads to the same choice probabilities as any given ARUM. The contrary, however, does not hold: some GEM combine substitutability and complementarity and, therefore, cannot be rationalized by any ARUM.[5] This means that our class of GEM is strictly larger than the class of ARUM.

In their seminal paper, Berry et al. (1995) provide a method for estimating the demand for differentiated products, while accounting for price endogeneity due to the presence of an unobserved characteristics term, which is the structural error of the model.[6] They propose a generalized method-of-moments (GMM) estimator, together with an estimation algorithm to compute it. To construct the GMM objective function, they need to invert the demand system to get the structural error as a function of the data and parameters, which cannot be done analytically in general. They suggest inverting the system numerically using a contraction mapping, which may be time consuming and requires using a tight convergence tolerance and a good starting value for the BLP estimator to produce reliable estimates (see e.g., Dubé et al., 2012; Knittel and Metaxoglou, 2014).

In contrast, with GEM, we obtain the structural error term directly as a known function of the data and parameters, meaning that we can easily implement the BLP method with standard regression techniques. This is because GE models are formulated in the space of consumption, and not in the dual space of indirect utilities, making the inverse demand system directly available.[7] Existence and uniqueness of the inverse system relies on the invertibility of the generator $\mathbf{S}$, which is shown using Gale and Nikaido (1965). Our invertibility result supplements, and in some cases extends, other results on demand invertibility

[5] In this paper, complementarity (resp., substitutability) is defined by a negative (resp., positive) cross derivative of demand with respect to alternative-specific characteristics, which can be the prices or any non-price characteristics (see Allen and Rehbeck, 2016, for more details on complementarity in PUM). When the characteristic is the price, this is the standard definition of complementarity (Samuelson, 1974). Note that there are different ways of defining complementarity and that the definition we use is related but different from the definition based on random utility used by Gentzkow (2007).

[6] We use the wider term "alternative" in the theoretical parts of the paper and the more narrow term "product" in the empirical parts.

[7] In this respect, GE models are alternatives to Dubé et al. (2012)'s and Lee and Seo (2015)'s algorithms. Dubé et al. (2012) transform the BLP's GMM minimization into a mathematical program with equilibrium constraints (MPEC), which minimizes the GMM objective function subject to the constraint that observed market shares be equal to predicted market shares. Lee and Seo (2015) approximate by linearization the non-linear system of market shares for the random coefficient logit model, and, in turn, do inversion analytically.

in different settings (see e.g., Berry, 1994; Beckert and Blundell, 2008; Chiappori and Komunjer, 2009; Berry et al., 2013).

GEM lead to demands with a tractable and familiar form that generalizes the logit demand in a nontrivial way. Different specifications of the generator $\mathbf{S}$ lead to different GEM. We propose a family of generators that lead to models that extend the multi-level nested logit models by allowing the nests to overlap in any way. This allows us to build GEM that are similar in the spirit to existing generalized extreme value (GEV) models that have already proved useful for demand estimation purposes. We show how to build ordered models describing markets having a natural ordering of alternatives (see Small, 1987; Grigolon, 2017) and nested models that generalize multi-level nested logit models in the spirit of Bresnahan et al. (1997).

Specifically, we propose a family of models that extend the nested logit model by allowing nests to overlap in any way. This allows us to build and estimate a generalized nested entropy (GNE) model that describes markets for differentiated products that exhibit segmentation according to several dimension. We illustrate their application by estimating demand for cereals in Chicago in 1991–1992. The GNE model provides rich patterns of substitution and complementarity, while being parsimonious, computationally fast and very easy to estimate. In particular, it can be estimated by a linear regression model of market shares on alternative-specific characteristics and terms related to segmentation.[8] [9]

Section 2 introduces the class of GE demand models and provides general methods for building them. Section 3 studies the linkages between choice models. Section 4 shows how to estimate GEM with aggregate data and discusses identification of GEM. Section 5 introduces the GNE model and demonstrates its use by estimating the demand for cereals in Chicago.

Notation. We use italics for scalar variables and real-valued functions, boldface for vectors, matrices and vector-valued functions, and script for sets. By default, vectors are column vectors.

Let $\mathbf{q} = (q_0, \ldots, q_J)^\intercal \in \mathbb{R}^{J+1}$ and $\boldsymbol{\delta} = (\delta_0, \ldots, \delta_J)^\intercal \in \mathbb{R}^{J+1}$ be two vectors. $|\mathbf{q}| = \sum_{j=0}^{J} |q_j|$ denotes the 1-norm of vector $\mathbf{q}$ and $\boldsymbol{\delta} \cdot \mathbf{q} = \sum_{j=0}^{J} \delta_j q_j$ denotes the vector scalar product.

[8] In this paper, the words "market share" and "demand" are used interchangeably. Note, however, that we use "market shares" in the empirical parts and "demands" in the theoretical parts.

[9] Practically, it is easily implemented using, e.g., the `ivregress` or `ivreg2` commands of the software package STATA.

Let $\Omega : \mathbb{R}^{J+1} \to \mathbb{R}$. Then, $\Omega_j(\mathbf{q}) = \frac{\partial \Omega(\mathbf{q})}{\partial q_j}$ denotes its partial derivative with respect to its jth entry and $\nabla_{\mathbf{q}} \Omega(\mathbf{q})$ denotes its gradient with respect to the vector $\mathbf{q}$. A univariate function $\mathbb{R} \to \mathbb{R}$ applied to a vector is a coordinate-wise application of the function, e.g., $\ln(\mathbf{q}) = (\ln(q_0), \ldots, \ln(q_J))$.

Let $\mathbf{S} : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ be a function composed of functions $S^{(j)} : \mathbb{R}^{J+1} \to \mathbb{R}$: $\mathbf{S}(\mathbf{q}) = \left(S^{(0)}(\mathbf{q}), \ldots, S^{(J)}(\mathbf{q})\right)$. Then, its Jacobian matrix $\mathbf{J_S}(\mathbf{q})$ has elements ij given by $\frac{\partial S^{(i)}(\mathbf{q})}{\partial q_j}$.

$\mathbf{A}^\intercal \in \mathbb{R}^{J \times J}$ denotes the transpose matrix of $\mathbf{A} \in \mathbb{R}^{J \times J}$. $\mathbf{0}_J = (0, \ldots, 0)^\intercal \in \mathbb{R}^J$ and $\mathbf{1}_J = (1, \ldots, 1)^\intercal \in \mathbb{R}^J$ denote the J-dimensional zero and unit vectors, respectively. $\mathbf{I}_J \in \mathbb{R}^{J \times J}$ and $\mathbf{1}_{JJ} \in \mathbb{R}^{J \times J}$ denote the $J \times J$ identity matrix and unit matrix (where every element equals one), respectively.

Let $\mathbb{R}^J_+ = [0, \infty)^J$ and $\mathbb{R}^J_{++} = (0, \infty)^J$. $\Delta = \left\{\mathbf{q} \in \mathbb{R}^{J+1}_+ : \sum_{j=0}^J q_j = 1\right\}$ denotes the J-dimensional unit simplex, with $\operatorname{int}(\Delta) = \Delta \cap \mathbb{R}^{J+1}_{++}$ its interior and $\operatorname{bd}(\Delta) = \Delta \setminus \operatorname{int}(\Delta)$ its boundary.

2 The Class of Generalized Entropy Models

2.1 Definitions

A consumer faces a choice set $\mathscr{J} = \{0, 1, \ldots J\}$ of $J+1$ alternatives. Let $\boldsymbol{\delta} = (\delta_0, \ldots, \delta_J)^\intercal$, where δ_j is the alternative j-specific utility component. The consumer chooses a vector of choice probabilities $\mathbf{q} = (q_0, \ldots, q_J)^\intercal \in \Delta$ to maximize her utility function

$$\sum_{j=0}^{J} \delta_j q_j + \Omega(\mathbf{q}), \tag{1}$$

defined as the sum of an expected utility component, which is linear in $\mathbf{q}$ and $\boldsymbol{\delta}$, and a function Ω, which is a nonlinear and deterministic function of $\mathbf{q}$. When Ω is concave, it is referred to as a perturbation function. This is then a perturbed utility model (hereafter, PUM).[10]

We build the class of generalized entropy models (hereafter, GEM) by specifying a functional form for Ω, which we call *generalized entropy* (hereafter, GE). Specifically, we require that Ω has a specific form defined in terms of a function $\mathbf{S}$, which we call *generator* and define as follows.

[10] See Hofbauer and Sandholm (2002), McFadden and Fosgerau (2012) and Fudenberg et al. (2015).

Definition 1 (Generator)**.** The function $\mathbf{S} = \left(S^{(0)}, \ldots, S^{(J)}\right) : \mathbb{R}_+^{J+1} \to \mathbb{R}_+^{J+1}$ is a generator if it is twice continuously differentiable and linearly homogeneous, and the Jacobian of $\ln \mathbf{S}$, $\mathbf{J}_{\ln \mathbf{S}}$, is positive definite and symmetric on $\mathrm{int}\,(\Delta)$.

A GE $\Omega : \mathbb{R}_+^{J+1} \to \mathbb{R} \cup \{-\infty\}$ is defined in terms of a generator $\mathbf{S}$ by

$$\Omega\left(\mathbf{q}\right) = -\sum_{j=0}^{J} q_j \ln S^{(j)}\left(\mathbf{q}\right), \ \mathbf{q} \in \Delta, \tag{2}$$

with $\Omega\left(\mathbf{q}\right) = -\infty$ when $\mathbf{q} \notin \Delta$.

A GEM is then defined as follows.

Definition 2 (GEM)**.** A GEM is a demand system that maximizes a utility of the form (1) over the unit simplex Δ, where $\Omega\left(\mathbf{q}\right)$ is a GE (2) and $\mathbf{S}$ is a generator.

The characterization of GEM in Definition 2 does not rule out zero demands in general. These are situations in which some alternatives are inferior to others so that they are never consumed. The following additional condition on $\mathbf{S}$ does rule out zero demands.[11] We retain Assumption 1 in the remainder of the paper, except when otherwise stated.

Assumption 1 (Positivity)**.** $|\ln \mathbf{S}(\mathbf{q})|$ approaches infinity as $\mathbf{q}$ approaches $\mathrm{bd}\,(\Delta)$.

We show below that the GE (2) is concave, which implies that any GEM is also a PUM. The converse, however, does not hold: there are PUM that are not GEM.[12] Nevertheless, the class of GEM remains large: we show in Section 3 that it incorporates all ARUM as duals. As we will show, the GEM structure turns out to be very useful in applications. It allows us to implement the BLP method with standard regression techniques, without having to invert demand numerically. At the same time, it allows us to tailor models to specific applications and accommodates rich patterns of substitution and complementarity.

[11] Hofbauer and Sandholm (2002) and Fudenberg et al. (2015) require similar conditions. Hofbauer and Sandholm (2002) assume that their perturbation function $\mathbf{V} : \mathrm{int}\,(\Delta) \to \mathbb{R}$ has a positive definite Hessian for all $\mathbf{q}$ that $|\nabla \mathbf{V}\left(\mathbf{q}\right)|$ approaches infinity as $\mathbf{q}$ approaches the boundary of Δ. Their perturbation $\mathbf{V}$ plays the same role as the negative of our GE $-\Omega$. Similarly, Fudenberg et al. (2015) assume that their cost function $c : [0,1] \to \mathbb{R} \cup \{\infty\}$ is strictly convex and continuously differentiable over $(0,1)$ and $\lim_{q\to 0} c'\left(q\right) = -\infty$.

[12] Hofbauer and Sandholm (2002) discuss the concave perturbation function $\sum_{j=0}^{J} \ln q_j$. The corresponding candidate generator $S^{(j)}\left(\mathbf{q}\right) = q_j^{1/q_j}$ is not linearly homogeneous and is hence not a generator.

2.2 Demand

The following lemma shows that a GE is indeed a concave function, such that a GEM is actually a PUM.

Lemma 1. Assume that $\mathbf{S}$ is a generator. Then $\mathbf{S}$ is invertible on $\mathrm{int}\,(\Delta)$ and satisfies the modified generalized Euler equation

$$\sum_{j=0}^{J} q_j \frac{\partial \ln S^{(j)}(\mathbf{q})}{\partial q_k} = 1, \quad k \in \mathscr{J},\ \mathbf{q} \in \mathrm{int}\,(\Delta), \tag{3}$$

and its corresponding GE Ω is strictly concave on $\mathrm{int}\,(\Delta)$.

The utility maximizing demand in the GEM exists, since the utility function is continuous on the compact set Δ. The strict concavity of Ω ensures that demand is unique and Assumption 1 ensures that it is interior. The modified generalized Euler equation (3), together with the invertibility of $\mathbf{S}$, allow us to derive a tractable and familiar demand form in Theorem 1. We denote the inverse of $\mathbf{S}$ by $\mathbf{H} = \mathbf{S}^{-1}$.

Theorem 1. Let $\mathbf{S}$ be a generator. Under Assumption 1, GEM lead to non-zero GE demands

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

where $H^{(i)}\left(e^{\boldsymbol{\delta}}\right) = S^{-1(i)}\left(e^{\boldsymbol{\delta}}\right)$.

Utility $\boldsymbol{\delta}$ and demand $\mathbf{q}$ are related through the generator $\mathbf{S}$ and its inverse $\mathbf{H}$ by

$$\delta_i = \ln S^{(i)}(\mathbf{q}) + \ln\left(\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)\right), \quad i \in \mathscr{J},\ \mathbf{q} \in \mathrm{int}\,(\Delta). \tag{5}$$

Equation (4) gives the mapping from demands $\mathbf{q}$ to utility $\boldsymbol{\delta}$, which can also be obtained using Roy's identity (see Proposition 1 below). This equation shows that GE demands have a tractable and familiar form that generalizes the logit demand in a nontrivial way.

Conversely, Equation (5) gives the inverse mapping from utility $\boldsymbol{\delta}$ to demands $\mathbf{q}$, which is unique up to a constant. This shows that GEM generate demands with an explicit inverse which, after specifying the functional form of the generator $\mathbf{S}$, can be used as basis for demand estimation.

For example, in the simplest possible case, the generator is the identity $\mathbf{S}(\mathbf{q}) = \mathbf{q}$ which implies that the inverse generator is also the identity $\mathbf{H}\left(e^{\boldsymbol{\delta}}\right) = e^{\boldsymbol{\delta}}$. In this case, the GE reduces to the Shannon entropy $\Omega(\mathbf{q}) = -\sum_{j=0}^{J} q_j \ln(q_j)$ and we obtain the logit demand (see Anderson et al., 1988):

$$q_i(\boldsymbol{\delta}) = \frac{e^{\delta_i}}{\sum_{j=0}^{J} e^{\delta_j}}. \tag{6}$$

In accordance with (5), utility $\boldsymbol{\delta}$ and demand $\mathbf{q}$ satisfy the relations

$$\delta_i = \ln(q_i) + \ln\left(\sum_{j=0}^{J} e^{\delta_j}\right), \quad i \in \mathscr{J}.$$

Let $G(\boldsymbol{\delta}) = \sum_{j=0}^{J} \delta_j q_j(\boldsymbol{\delta}) + \Omega(\mathbf{q}(\boldsymbol{\delta}))$ be the consumer's surplus, or indirect utility, associated with the perturbed utility (1). Proposition 1 shows that, as in the logit model, the consumer's surplus is simply the log of the denominator of GE demands.

Proposition 1. The consumer's surplus is given by

$$G(\boldsymbol{\delta}) = \ln\left(\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)\right). \tag{7}$$

GE demands (4) are consistent with Roy's identity, i.e., $q_i = \frac{\partial G(\boldsymbol{\delta})}{\partial \delta_i}$ for all $i \in \mathscr{J}$.

GE demands q_j given by (4) are increasing in their own utility component δ_j.[13] Proposition 2 provides an expression for the whole matrix of demand derivatives.

Proposition 2. The matrix of demand derivatives $\partial q_j / \partial \delta_i$ is given by

$$\mathbf{J}_{\mathbf{q}} = \left[\mathbf{J}_{\ln \mathbf{S}}(\mathbf{q})\right]^{-1} \left[\mathbf{I} - \mathbf{1}\mathbf{q}^{\intercal}\right], \tag{8}$$

where $\mathbf{q} = \mathbf{q}(\boldsymbol{\delta})$ given by Equation (4).

Since GEM are defined without explicit reference to income, complementarity (resp., substitutability) between alternatives is just understood as a negative (resp., positive) cross derivative of GE demands. Proposition 2 does not allow to know whether complementarity may or may not arise in GEM. Example 3 below exhibits a GEM in which alternatives are sometimes complements.

[13]This property holds for all PUM (see McFadden and Fosgerau, 2012) and is due to the concavity of the perturbation function, hence it also holds for all GEM. When the δ_j are decreasing functions of prices p_j, this is equivalent to stating that demands are decreasing in their own prices.

2.3 Construction of GEM

To construct a GEM, it suffices to construct a generator that satisfies Definition 1.[14] We here propose a family of generators that lead to models that extend the nested logit (NL) model in a very intuitive way as follows.[15]

We first observe that the nested logit (NL) model can be cast as a GEM. Suppose that the choice set is partitioned into non-overlapping sets, usually called nests. Let g_j be the nest that contains alternative j. Then the generator that leads to the NL demands is given by

$$S^{(j)}(\mathbf{q}) = q_j^{\mu}\left(\sum_{i\in g_j} q_i\right)^{1-\mu},$$

where $\mu \in (0,1)$ is the nesting parameter.[16]

The multi-level NL models generalize the NL model. They are obtained by partitioning the choice set into nests and then further partitioning each nest into subnests, and so on (see e.g., Goldberg, 1995; Verboven, 1996). This hierarchical structure implies that each choice alternative belongs to only one (sub)nest at each level, meaning that nests are not allowed to overlap by construction.

The following proposition generalizes the NL model by giving a construction of generators through a nesting operation that allows the nests to overlap in any way.

Proposition 3 (General nesting)**.** Let $\mathscr{G} \subseteq 2^{\mathscr{J}}$ be a finite set of nests with associated nesting parameters μ_g , where $\mu_0 + \sum_{\{g\in\mathscr{G}|j\in g\}}\mu_g = 1$ for all $j \in \mathscr{J}$ with $\mu_g \geq 0$ for all $g \in \mathscr{G}$ and $\mu_0 > 0$. Let $\mathbf{S}$ be given by

$$S^{(j)}(\mathbf{q}) = q_j^{\mu_0}\prod_{\{g\in\mathscr{G}|j\in g\}} q_g^{\mu_g}, \tag{9}$$

where $q_g = \sum_{i\in g} q_i$. Then $\mathbf{S}$ is a generator.[17]

[14]Similarly, different GEV models (see McFadden, 1981) are obtained from different specifications of a choice probability function (Fosgerau et al., 2013).

[15]In Appendix D, we provide a range of general methods for building generators along with illustrative examples.

[16]The corresponding GE is given by the sum of two Shannon entropies since $\Omega(\mathbf{q}) = -\mu\sum_{j\in\mathscr{J}} q_j \ln(q_j) - (1-\mu)\sum_{g=1}^{G} q_g \ln(q_g)$.

[17]Without the term $q_j^{\mu_0}$, $\mathbf{S}$ is twice continuously differentiable and linearly homogeneous, and $\mathbf{J}_{\ln \mathbf{S}}$ is symmetric, but not necessarily positive definite. The general nesting operation leads to the following GE $\Omega(\mathbf{q}) = -\mu_0\sum_{j\in\mathscr{J}} q_j \ln(q_j) - \sum_{\{g\in\mathscr{G}|j\in g\}}\left[\mu_g \sum_{j\in\mathscr{J}} q_j \ln(q_g)\right]$, where the first term is the Shannon

Proposition 3 allows building GEM that are similar in spirit to the well-known GEV models based on nesting (see e.g., Train, 2009, Chapter 4 for details).

As an example, we construct here a model describing a market having a natural ordering of alternatives, where alternatives that are nearer each other in the ordering are closer substitutes. This is true, for example, for hotels that can be ordered according to their number of stars and for breakfast cereals according to sugar content. The example below provides GEM that is similar to the GEV ordered models of Small (1987) and Grigolon (2017).

Example 1 (Ordered model)**.** Let alternative 0 be the outside option, and alternatives $1, \ldots, J$ be ordered in ascending sequence. We make the ordering circular, letting alternative 1 follow alternative J. Let $\mu_0 > 0$ and $\mu_1, \mu_2, \mu_3 \geq 0$ with $\mu_0 + \mu_1 + \mu_2 + \mu_3 = 1$. The function $\mathbf{S}$ given by

$$S^{(j)}(\mathbf{q}) = \begin{cases} q_0, & j = 0 \\ q_j^{\mu_0} q_{\sigma_1(j)}^{\mu_1} q_{\sigma_2(j)}^{\mu_2} q_{\sigma_3(j)}^{\mu_3}, & j > 0, \end{cases}$$

with $q_{\sigma_1(j)} = q_{j-2} + q_{j-1} + q_j$, $q_{\sigma_2(j)} = q_{j-1} + q_j + q_{j+1}$, $q_{\sigma_3(j)} = q_j + q_{j+1} + q_{j+2}$, is a generator.

In Example 2, similarly to the Product-Differentiation Logit (PDL) model of Bresnahan et al. (1997), we build a nested model describing markets that exhibit product segmentation along several dimensions (see Section 5 for more details).

Example 2 (Nested model)**.** Let $\mu_0 > 0$ and $\mu_1, \mu_2 \geq 0$ with $\mu_0 + \mu_1 + \mu_2 = 1$. Let $\sigma_c(j)$ be the set of alternatives that are grouped together with alternative j on dimension $c = 1, 2$ and $q_{\sigma_c(j)} = \sum_{i \in \sigma_c(j)} q_i$. The function $\mathbf{S}$ given by

$$S^{(j)}(\mathbf{q}) = \begin{cases} q_0, & j = 0 \\ q_j^{\mu_0} q_{\sigma_1(j)}^{\mu_1} q_{\sigma_2(j)}^{\mu_2}, & j > 0. \end{cases}$$

is a generator.

The next example shows that GEM allow alternatives to be complements.

entropy that expresses consumer's taste for variety over all alternatives and the second term expresses consumer's taste for variety over alternatives belonging to group g (see Verboven, 1996).

Example 3. Let $\mathbf{S}$ be defined by

$$\mathbf{S}(\mathbf{q}) = \begin{cases} q_0^{\mu}\left(q_0 + \frac{1}{2}q_1\right)^{1-\mu}, \\ q_1^{\mu}\left(q_0 + \frac{1}{2}q_1\right)^{\frac{1-\mu}{2}}\left(\frac{1}{2}q_1 + q_2\right)^{\frac{1-\mu}{2}}, \\ q_2^{\mu}\left(\frac{1}{2}q_1 + q_2\right)^{1-\mu}, \end{cases}$$

with $\mu \in (0,1)$. Then $\mathbf{S}$ is a generator.

Differentiating the first-order conditions of the utility maximization problem with respect to δ_0, we find that $\partial q_2/\partial \delta_0 > 0$ if and only if

$$\mu < \frac{q_1}{4q_0q_2 + 3q_1q_2 + 2q_1^2 + 3q_0q_1}.$$

At $\boldsymbol{\delta}$ such that $q_0 = q_1 = q_2 = 1/3$, the condition becomes $\mu < 1/4$, thereby showing that there exists combinations of parameters μ and utilities $\boldsymbol{\delta}$ at which some alternatives are complements.

As all alternatives are substitutes in an ARUM, Example 3 proves the following result.

Proposition 4. Some GEM lead to demand systems that cannot be rationalized by any ARUM.

We show in Subsection 3.1 that any ARUM has a GEM counterpart that leads to the same choice probabilities. Combining this with Proposition 4 shows that the class of GEM is strictly larger than the class of ARUM.

3 Linkages between Choice Models

In this section, we study first the relation between GEM and ARUM, finding that the choice probabilities of any ARUM can be obtained as the demand of some GEM. Then we introduce income and prices to bridge between representative consumer models and GEM.

3.1 ARUM as GEM

We begin by setting up the additive random utility model (ARUM). Consider a consumer who faces a choice set $\mathscr{J} = \{0, 1, \ldots, J\}$ of $J+1$ alternatives and chooses the alternative that gives her the highest (indirect) utility $u_j = \delta_j + \varepsilon_j$, $j \in \mathscr{J}$, where δ_j is a deterministic utility term and ε_j is a random utility term. The following assumption on $\boldsymbol{\varepsilon}$ is standard in the discrete choice literature.

Assumption 2. The random vector $\boldsymbol{\varepsilon} = (\varepsilon_0, \ldots, \varepsilon_J)$ follows a joint distribution with finite means that is absolutely continuous, independent of $\boldsymbol{\delta} = (\delta_0, \ldots, \delta_J)$, and has full support on $\mathbb{R}^{J+1}$.

Assumption 2 implies that utility ties $u_i = u_j$, $i \neq j$, occur with probability 0 (because the joint distribution of $\boldsymbol{\varepsilon}$ is absolutely continuous), meaning that the argmax set of the ARUM is almost surely a singleton, the choice probabilities are all everywhere positive (because $\boldsymbol{\varepsilon}$ has full support) and that random coefficients are not allowed (because the joint distribution of $\boldsymbol{\varepsilon}$ is independent of $\boldsymbol{\delta}$).

Let $\overline{G} : \mathbb{R}^{J+1} \to \mathbb{R}$ given by

$$\overline{G}(\boldsymbol{\delta}) = \mathbb{E}\left(\max_{j \in \mathscr{J}} u_j\right) \tag{10}$$

be the expected maximum utility. Let $\mathbf{P} = (P_0(\boldsymbol{\delta}), \ldots, P_J(\boldsymbol{\delta})) : \mathbb{R}^{J+1} \to \Delta$ be the vector of choice probabilities with $P_j(\boldsymbol{\delta})$ being the probability of choosing alternative j.

From the Williams-Daly-Zachary theorem (McFadden, 1981), the choice probabilities and the derivatives of $\overline{G}(\boldsymbol{\delta})$ coincide, i.e.,

$$P_j(\boldsymbol{\delta}) = \frac{\partial \overline{G}(\boldsymbol{\delta})}{\partial \delta_j}. \quad j \in \mathscr{J}, \tag{11}$$

Let $\overline{\mathbf{H}} = \left(\overline{H}^{(0)}, \ldots, \overline{H}^{(J)}\right)$, with $\overline{H}^{(i)} : \mathbb{R}^{J+1}_{++} \to \mathbb{R}_{++}$ defined as the derivative of the exponentiated surplus with respect to its ith component, i.e.,

$$\overline{H}^{(i)}\left(e^{\boldsymbol{\delta}}\right) = \frac{\partial e^{\overline{G}(\boldsymbol{\delta})}}{\partial \delta_i}. \tag{12}$$

Note that $\sum_{j \in \mathscr{J}} \overline{H}^{(i)}\left(e^{\boldsymbol{\delta}}\right) = e^{\overline{G}(\boldsymbol{\delta})}$.[18] Then the ARUM choice probabilities may be written as

$$P_i(\boldsymbol{\delta}) = \frac{\overline{H}^{(i)}\left(e^{\boldsymbol{\delta}}\right)}{\sum_{j=0}^{J} \overline{H}^{(j)}\left(e^{\boldsymbol{\delta}}\right)}, \quad i \in \mathscr{J}, \tag{13}$$

[18]This follows since (12) may be written as $\overline{H}^{(i)}\left(e^{\boldsymbol{\delta}}\right) = \frac{\partial \overline{G}(\boldsymbol{\delta})}{\partial \delta_i} e^{\overline{G}(\boldsymbol{\delta})}$ for all $j \in \mathscr{J}$. Then by (11), $\overline{H}^{(i)}\left(e^{\boldsymbol{\delta}}\right) = P_j(\boldsymbol{\delta})\, e^{\overline{G}(\boldsymbol{\delta})}$ for all $j \in \mathscr{J}$. Finally, sum over $j \in \mathscr{J}$ and use that choice probabilities sum to one.

which is exactly the same form as the GEM demand (4) if $\mathbf{H} = \overline{\mathbf{H}}$. To establish that the ARUM choice probabilities (13) can be generated by a GEM, it then only remains to show that $\overline{\mathbf{H}}$ has an inverse $\overline{\mathbf{S}} = \overline{\mathbf{H}}^{-1}$ and that this inverse is a generator. This is established in the following lemma.

Lemma 2. The function $\overline{\mathbf{H}}$ is invertible, and its inverse $\overline{\mathbf{S}} = \overline{\mathbf{H}}^{-1}$ is a generator.

Then the function $-\overline{G}^*$ given by

$$-\overline{G}^*(\mathbf{q}) = -\sum_{j=0}^{J} q_j \ln \overline{S}^{(j)}(\mathbf{q}), \ \mathbf{q} \in \Delta, \tag{14}$$

and $-\overline{G}^*(\mathbf{q}) = +\infty$ when $\mathbf{q} \notin \Delta$ is a GE. Fosgerau et al. (2017) show that $-\overline{G}^*$ is the convex conjugate of $\overline{G}$.[19]

Theorem 2 below summarizes the results as follows.

Theorem 2. The ARUM choice probabilities (13) with surplus function $\overline{G}$ given by (10) coincide with the GE demand system (4) with GE function $-\overline{G}^*$, where $\overline{G}^*$ is the convex conjugate of $\overline{G}$ given by (14).

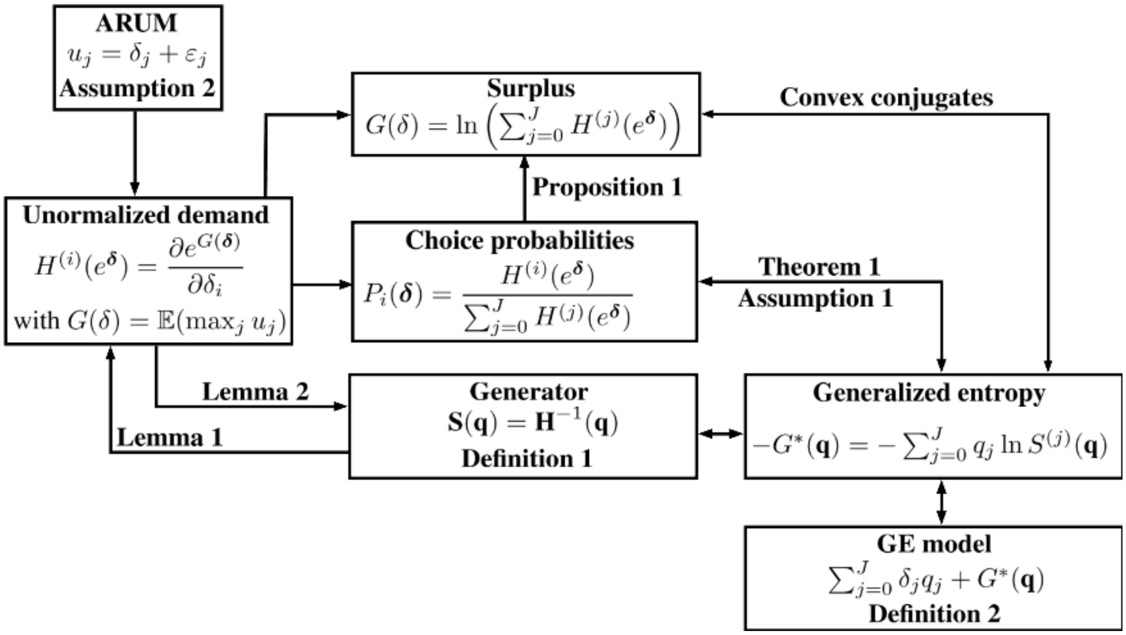
According to Theorem 2, all ARUM have a GEM as counterpart that leads to the same demand. However, as shown in Example 3, the converse is not true: the class of GEM is strictly larger than the class of ARUM. When a GEM corresponds to an ARUM, the surplus function (7) and the maximum expected utility (10) coincide, i.e., $G = \overline{G}$; and similarly for their generators, i.e., $\mathbf{S} = \overline{\mathbf{S}}$. Figure 1 illustrates how ARUM and GEM are linked and shows how, beginning with some ARUM, we can determine a GEM with a corresponding demand that is equal to the ARUM choice probabilities.

3.2 Link to standard consumer theory

In this section we discuss briefly how a PUM and hence also a GEM may be specified as a standard consumer demand model including income and prices. Distinguishing price from quality in the utility associated with a product will be useful in empirical applications in

[19]The latter result is well-known in the special case of the logit model, i.e. that the convex conjugate of the negative entropy $f(\mathbf{q}) = \sum_j q_j \ln(q_j)$ is the log-sum $f^*(\boldsymbol{\delta}) = \ln\left(\sum_j e^{\delta_j}\right)$ (see e.g., Boyd and Vandenberghe, 2004).

Figure 1: LINKAGES BETWEEN GE MODELS AND ARUM



industrial organization, since prices must generally be considered as endogenous (Berry, 1994; Berry et al., 1995).

Consider a variety-seeking consumer facing choice set of $J+1$ differentiated products, $\mathscr{J} = \{0, 1, \dots J\}$, and a homogeneous numeraire good, with demands for the differentiated products summing to one. Let p_j and v_j be the price and the quality of product $j \in \mathscr{J}$, respectively. We normalize the price of numeraire good to 1 and assume that the consumer's income y is sufficiently high, $y > \max_{j \in \mathscr{J}} p_j$, that consumption of the numeraire good is strictly positive (i.e., not all income is spent on the differentiated products).

Let $\mathbf{q} = (q_0, \dots, q_J)^\intercal$ be the vector of quantities consumed of the differentiated products and z be the quantity consumed of the numeraire good. The consumer's direct utility function u, which is quasi-linear in the numeraire, is given by

$$u(\mathbf{q}, z) = \alpha z + \sum_{j=0}^{J} v_j q_j + \Omega(\mathbf{q}), \tag{15}$$

where $\alpha > 0$ is the marginal utility of income, and Ω is a GE defined by (2).

The utility in (15) consists of three components: the first describes the utility derived from the consumption of the numeraire good, the second describes the net utility derived from the consumption of the products in the absence of interaction among them, and the third expresses the consumer's taste for variety in terms of a GE.

The consumer chooses $\mathbf{q} \in \Delta$ and $z \in \mathbb{R}_+$ so as to maximize her utility (15) subject to her budget constraint. She solves

$$\max_{(\mathbf{q},z)\in\Delta\times\mathbb{R}_+} u(\mathbf{q}, z), \quad \text{subject to} \quad \sum_{j=0}^{J} p_j q_j + z \leq y. \tag{16}$$

The budget constraint is binding,[20] so that (16) can be rewritten as follows

$$\max_{\mathbf{q}\in\Delta} \left\{ \alpha y + \sum_{j=0}^{J} \delta_j q_j + \Omega(\mathbf{q}) \right\}, \tag{17}$$

where $\delta_j = v_j - \alpha p_j$ is the net utility that the consumer derives from consuming one unit of product j.[21]

This shows that such a model can be cast as a GEM with prices entering the perturbed utility linearly. Then, under Assumption 1, the implied demand system is given by (4) and the indirect utility by $w(\boldsymbol{\delta}, y) = \alpha y + G(\boldsymbol{\delta})$, where G is the surplus function (7). Note that the demand system is consistent with Roy's identity, i.e., $q_i = -\frac{\partial w(\boldsymbol{\delta},y)}{\partial p_i} / \frac{\partial w(\boldsymbol{\delta},y)}{\partial y}$, $i \in \mathscr{J}$.

The quasi-linearity of the direct utility function 15 has two implications. First, GEM demands for the differentiated products are independent of income, so that all the income effects are captured by the numeraire.

Second, in the GEM, as in any model with quasi-linear direct utility (see e.g., Vives, 2001), the assumption of a representative consumer is not restrictive. Indeed, consider a population of utility-maximizing consumers all with quasi-linear direct utility of the form (15), and assume that they all have the same constant marginal utility of income $\alpha > 0$. Then individual indirect utilities have the Gorman form and can thus be aggregated across consumers, meaning that consumers can be treated as if they were a single consumer, regardless of the distribution of unobserved consumer heterogeneity or of income.[22]

[20] This is because $\alpha > 0$ and $y > \max_{j\in\mathscr{J}} p_j$.

[21] In the empirical industrial organization literature, δ_j is referred to as the mean utility of product j.

[22] Consider a population of N consumers. Suppose that each consumer n's direct utility function takes the form $u_n(\mathbf{q}_n, z_n) = \alpha z_n + \sum_{j=0}^{J} v_j q_{jn} + \Omega_n(\mathbf{q}_n)$. Then consumer n's indirect utility is given by $v_n(\boldsymbol{\delta}, y_n) = \alpha y_n + \ln\left(\sum_{j=0}^{J} H_n^{(j)}\left(e^{\boldsymbol{\delta}}\right)\right)$ and has the Gorman form $v_n(\mathbf{p}, y_n) = b(\mathbf{p}) y_n + a_n(\mathbf{p})$ with $b(\mathbf{p}) = \alpha$ that is identical for all consumers and $a_n(\mathbf{p}) = \ln\left(\sum_{j=0}^{J} H_n^{(j)}\left(e^{\boldsymbol{\delta}}\right)\right)$ that differs from consumer to consumer.

4 Estimation of GEM

We are now able to estimate GEM. In Subsection 2.3, we proposed some general methods for constructing generators. The generators thus constructed can be written as functions of the data and some parameters to be estimated with individual-level or aggregate data. In this section, we show how to estimate GEM with aggregate data. The main finding is that GEM are easily estimated by standard regression techniques.

4.1 Econometric Model

The aggregate data required to estimate GEM consists of the market shares, prices and characteristics for each product in each market (see e.g., Nevo, 2001).

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 characteristics term of product j in market t: this take into account the fact that the product characteristics used in estimation do not include all the product characteristics that consumers care about. We fix $\xi_{0t} = 0$ for all markets t.

In addition, assume that net utility is parametrized as

$$\delta_{jt}\left(\mathbf{X}_{jt}, p_{jt}, \xi_{jt}; \boldsymbol{\theta}_1\right) = \beta_0 + \mathbf{X}_{jt}\boldsymbol{\beta} - \alpha p_{jt} + \xi_{jt},$$

where $\boldsymbol{\theta}_1 = (\alpha, \beta_0, \boldsymbol{\beta})$ is the vector of the *linear parameters* that enter the linear part of the utility, and p_{jt} and $\mathbf{X}_{jt}$ are the price and (any function of) the characteristics of each product j in each market t that vary both across products and across markets, respectively. The intercept β_0 captures the value of consuming an inside product instead of the outside option; the parameter vector $\boldsymbol{\beta}$ represents the consumers' taste for the $\mathbf{X}_{jt}$'s; and the parameter $\alpha > 0$ is consumers' price sensitivity (i.e. the marginal utility of income).

Let $\boldsymbol{\theta}_2$ be a parameter vector that enters the nonlinear part of the utility and parametrize the generator as $S^{(j)}\left(\mathbf{q}_t; \boldsymbol{\theta}_2\right)$. Then using (5) we have

$$\ln S^{(j)}\left(\mathbf{q}_t; \boldsymbol{\theta}_2\right) = \delta_{jt}\left(\mathbf{X}_{jt}, p_{jt}, \xi_{jt}; \boldsymbol{\theta}_1\right) + c_t,\ j = 0 \ldots J,\ t = 1 \ldots T,$$

where $c_t \in \mathbb{R}$ denotes the log-sum term that is common across products on the same market, and $\mathbf{q}_t = (q_{0t}, \ldots, q_{Jt})^{\intercal}$.

We impose the normalization that $\delta_{0t} = 0$ for all $t = 1 \ldots T$, ensuring that $\boldsymbol{\delta}_t$ is uniquely determined on each market t.

Subtracting the equations for the outside good, we end up with the $J \times T$ demand equations $\xi_{jt} = \xi_{jt}(\boldsymbol{\theta})$ where the market-specific constant terms c_t have dropped out, and with

$$\xi_{jt}(\boldsymbol{\theta}) = \ln S^{(j)}(\mathbf{q}_t; \boldsymbol{\theta}_2) - \ln S^{(0)}(\mathbf{q}_t; \boldsymbol{\theta}_2) - (\beta_0 + \mathbf{X}_{jt}\boldsymbol{\beta} - \alpha p_{jt}), \tag{18}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ are the demand parameters to be estimated.

After transformation, GEM are nonlinear regression models, where the error is non-additive. Such models can be estimated using standard regression techniques. Thus, there is no need to use numerical inversion of market shares and simulation techniques that are associated with problems of global convergence (Knittel and Metaxoglou, 2014), of numerical integration (Skrainka and Judd, 2011), and of accuracy of BLP's contraction mapping (Dubé et al., 2012).

4.2 Identification

Prices and market shares form two different sets of endogenous variables and require different sources of exogenous variation for the model be identified. Prices are endogenous due to the presence of the unobserved product characteristics ξ_{jt}. Indeed, price competition models with differentiated products typically assume that firms consider both observed and unobserved product characteristics when setting prices, and that they make prices a function of marginal costs and a markup term. Since the markup term is a function of the (entire vector of) unobserved product characteristics, which constitute the error terms in Equations (18), prices are likely to be correlated with the error terms. Market shares are endogenous because demands are defined by a system of equations, where each demand depends on the entire vectors of endogenous prices and of unobserved product characteristics.

GE models provide a system of demand equations (18) where each equation has one unobservable ξ_{jt} and, under the standard assumption that products characteristics are exogenous, depends on $(J+1)$ endogenous variables, namely all the market shares $\mathbf{q}_t$ and one price p_{jt}. The main identification assumption is the existence of as many excluded (from the demand equations) instruments z_t as there are endogenous variables. Instruments are variables that are correlated with the endogenous variables (relevance) but are not correlated with the error term ξ_{jt} (exogeneity). We propose a GMM estimator based on the conditional moment restrictions $\mathbb{E}[\xi_{jt}(\boldsymbol{\theta}) | z_t] = 0$, which lead to the unconditional moment restrictions $\mathbb{E}[z_t \xi_{jt}(\boldsymbol{\theta})] = 0$.

We require instruments for prices and for some (functions of) market shares, where the need for instruments for market shares depends on the structure of the generator. For example, in the case of the NL model,

$$\xi_{jt} = \ln\left(\frac{q_{jt}}{q_{0t}}\right) - \mu \ln\left(q_{jt|g_j}\right) + \alpha p_{jt} - \left(\mathbf{X}_{jt}\boldsymbol{\beta} + \beta_0\right),$$

where $q_{jt|g_j}$ is the share of product j within its corresponding nest g_j. This requires only two instruments, one for price p_{jt} and one for the share $q_{jt|g_j}$.

Following the prevailing literature (Berry and Haile, 2014; Reynaert and Verboven, 2014; Armstrong, 2016), both cost shifters and BLP instruments are required. Cost shifters (i.e., input prices) separate exogenous variation in prices due to exogenous cost changes from endogenous variation in prices from unobserved product characteristics changes. They are valid under the assumption that input price variations are correlated with price variations, but not with changes in unobservable product characteristics. However, they are not sufficient on their own, because costs affect the endogenous market shares only through prices.

BLP instruments are functions of the characteristics of competing products and are valid instruments under the assumption that $\mathbf{X}_{jt}$ is exogenous (i.e., ξ_{jt} is independent of $\mathbf{X}_{jt}$). They separate exogenous variation in prices due to changes in $\mathbf{X}_{jt}$ from endogenous variation in prices from unobserved product characteristics changes. They are commonly used to instrument prices with the idea that characteristics of competing products are correlated with prices since the (equilibrium) markup of each product depends on how close products are in characteristics space (products with close substitutes will tend to have low markups and thus low prices relative to cost). They are also appropriate instruments for market shares on the RHS of (18).[23] BLP instruments can suffice for identification but cost shifters are useful in practice (see e.g., Reynaert and Verboven, 2014).

4.3 Relation to Berry Inversion

Berry inversion consists in inverting the system that equates observed market shares to predicted market shares, in which the terms ξ_{jt} enter non-linearly in general, to get a system of equations in which the terms ξ_{jt} enter linearly. Inversion can be done analytically or

[23]This is because identifying the effects of markets shares in the inverse demand system amounts to identifying the effects of $\mathbf{v}$ on market shares and that BLP instruments directly shifts $\mathbf{v}$.

numerically, depending on whether the system has a closed form or not.[24] The inverse system thus obtained serves as a basis for demand estimation.

Berry et al. (2013) generalize Berry (1994)'s invertibility result and show that their "connected substitutes" structure is sufficient for invertibility. They require that (i) products be weak gross substitutes (i.e., everything else equal, an increase in δ_j weakly decreases demand q_i for all other products) and (ii) the "connected strict substitution" condition hold (i.e., there is sufficient strict substitution between products to treat them in one demand system). Their structure can accommodate models with complementary products, but the first requirement is not always satisfied in GEM, meaning that Berry et al. (2013)'s results are not applicable.

GEM provide the system (5) which is just the inverse system obtained by Berry inversion. This is because GEM are formulated in the space of market shares and not in the space of indirect utilities. In GEM, the inverse system is thus directly available and has a known and analytic formula. In turn, getting GEM demands (4) requires inverting the system (5) and amounts to performing Berry inversion but in the opposite direction.

5 Empirical Application: Demand for Cereals

In this section, we apply a GEM to estimate the demand for cereals in Chicago in 1991 – 1992. The cereals market is known to exhibit product segmentation. To take into account this feature, we build a generalized nested entropy (GNE) model by application of Corollary 3. As it will become clear, the GNE model is convenient for describing markets that exhibit product segmentation along several dimensions. It is closely related to the Product-Differentiation Logit (PDL) model of Bresnahan et al. (1997), which is an instance of a cross-nested logit model. We find that our GNE is simple and fast to estimate using standard linear regression techniques.

5.1 Product Segmentation on the Cereals Market

Data. We use data from the Dominick's Database made available by the James M. Kilts Center, University of Chicago Booth School of Business. We consider the ready-to-eat (RTE) cereal category during the period 1991–1992; and we supplement the data with the

[24] Berry (1994) and Brenkers and Verboven (2006) show that there is such a closed form for the logit and NL models, and the three-level NL model, respectively. Berry, Levinsohn, and Pakes (1995) show that there is no longer a closed form for the RCL model, but that the inverse exists.

nutrient content of the RTE cereals using the USDA Nutrient Database for Standard Reference (fiber, sugar, lipid, protein, energy, and sodium), and with the sugar monthly price from the website www.indexmundi.com. Following the prevailing literature, we aggregate UPCs into brands (e.g., Kellogg's Special K), so that different size boxes are considered one brand, 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, which account for 73 percent of sales of the category in the sample we use. We define a product as a brand, and a market as a store-month pair. Market shares and prices are computed following Nevo (2001) (see Appendix E.3 for more details).

Product segmentation. Formulated in general, we consider a market for differentiated products that exhibits product segmentation according to C *dimensions*, indexed c. Each dimension c taken separately potentially provides a source of segmentation and defines a finite number of *nests*. Each product belongs to exactly C nests, one for each dimension, and the nesting structure is exogenous. The dimensions taken together define product *types*. Products of the same type are those that are grouped together according to all the dimensions. Each dimension defines a concept of product closeness (or distance), so that products of the same type will be closer substitutes than products of different types.

For the application, we focus on two dimensions that form 17 product types: one measures the substitutability between products within the same market segment, where segments are family, kids, health, and taste enhanced (see e.g., Nevo, 2001); and the other measures the advantages the brand-name reputation provides to the products, where brand names are General Mills, Kellogg's, Quaker, Post, Nabisco, and Ralston.

5.2 The GNE Model

Let $\sigma_c(j)$ be the set of products that are grouped with product j according to dimension c (i.e., a nest), and $q_{\sigma_c(j),t} = \sum_{i\in\sigma_c(j)} q_{it}$ be the market share of nest $\sigma_c(j)$ in market t. Let $\mathbf{\Theta}_c$ be the nesting structure matrix for dimension c, having elements

$$(\mathbf{\Theta}_c)_{ij} = \begin{cases} 1, & \text{if } i \in \sigma_c(j), \\ 0, & \text{otherwise,} \end{cases} \tag{19}$$

and let $\mathbf{\Theta} = (\mathbf{\Theta}_1, \ldots, \mathbf{\Theta}_C)$ denote the array of nesting structure matrices.

Based on Corollary 3, we define the generalized nesting entropy (GNE) model as follows.

Definition 3 (GNE model)**.** The GNE model is a GEM with generator given by

$$S^{(j)}(\mathbf{q}) = \begin{cases} q_0, & j = 0, \\ q_j^{\mu_0} \prod_{c=1}^{C} q_{\sigma_c(j)}^{\mu_c}, & j > 0, \end{cases} \tag{20}$$

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

The GNE model satisfies Assumption 1, so that zero demands never arise. Product $j = 0$ is the outside option, which defines itself a product type and is the only product of its type. Let $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_C)$ be the vector of nesting parameters. The parameter μ_0 measures the consumers' taste for variety over all products and each μ_c, $c \geq 1$, measures the consumers' taste for variety over products of the same nest according to dimension c (see Verboven, 1996).

The following proposition is useful for understanding the behavior of the GNE model.

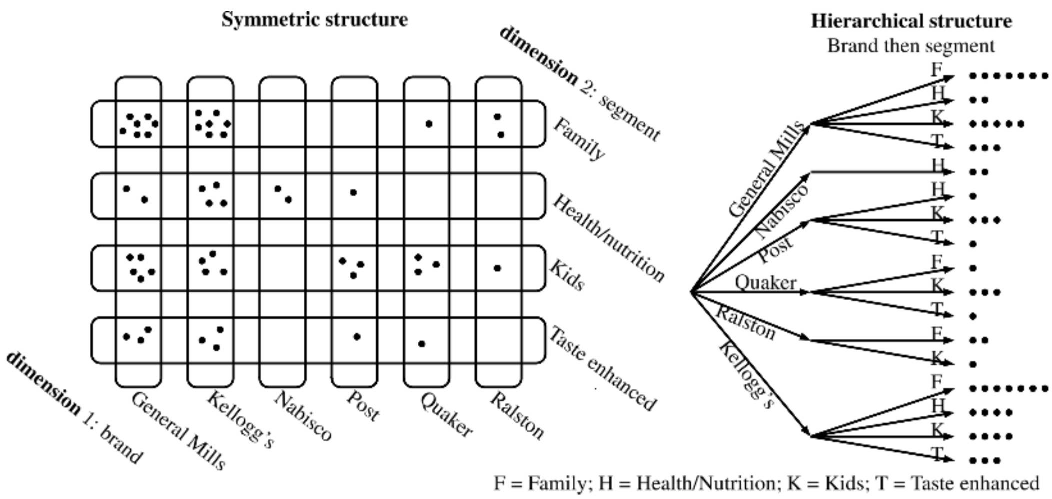
Proposition 5. In the GNE model, the IIA holds for products of the same type; but does not hold in general for products of different types.

Appendix E.2 provides some simulation results investigating the patterns of substitution and complementarity as the nesting structure and market shares change. In summary, we find that (i) products of the same type are never complementary, while products of different types may or may not be complementary; (ii) a larger outside option (in terms of market shares), except if it is extremely large, does not generate complementarity; and (iii) the size of the cross-elasticities depends on the degree of closeness between products as measured by the value of the nesting parameters and by the proximity of the products in the characteristics space used to form product types.

The structure of the GNE model in the present application to cereals is illustrated in the left panel of Figure 2. Each dot illustrates the location of a product in the nesting structure and there are 17 non-empty types. The two segmentation dimensions are treated symmetrically in this model.

The right panel of Figure 2 illustrates one of the two NL models that are possible with the same two segmentations. The NL models have a hierarchical nesting structure, in which the second layer of nesting is a partitioning of the first. Both NL models can be represented as GNE models and we estimate both for comparison. This is easily done using the same regression setup while changing only the nesting structure.

Figure 2: PRODUCT SEGMENTATION ON THE CEREALS MARKET



5.3 Estimation

For the GNE model, Equation (18) can be written as

$$\xi_{jt} = \mu_0 \ln (q_{jt}) + \sum_{c=1}^{C} \mu_c \ln \left(q_{\sigma_c(j),t}\right) - \ln (q_{0t}) - (\beta_0 + \mathbf{X}_{jt}\boldsymbol{\beta} - \alpha p_{jt}),$$

and using the parameter constraint $\mu_0 + \sum_{c=1}^{C} \mu_c = 1$, we obtain

$$\ln \left(\frac{q_{jt}}{q_{0t}}\right) = \beta_0 + \mathbf{X}_{jt}\boldsymbol{\beta} - \alpha p_{jt} + \sum_{c=1}^{C} \mu_c \ln \left(\frac{q_{jt}}{q_{\sigma_c(j),t}}\right) + \xi_{jt}, \tag{21}$$

for $j = 1 \ldots J$ and $t = 1 \ldots T$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, with $\boldsymbol{\theta}_1 = (\beta_0, \boldsymbol{\beta}, \alpha)$ and $\boldsymbol{\theta}_2 = (\mu_1, \ldots, \mu_C)$, are the parameters to be estimated.

Equation (21) is the same as the logit and NL equations (see Berry, 1994; Brenkers and Verboven, 2006), except for the terms $\mu_c \ln \left(q_{jt}/q_{\sigma_c(j)}\right)$; this suggests estimating the GNE model by a linear instrumental variables regression of market shares on product characteristics and terms related to market segmentation.

Price p_{jt} is endogenous due to the presence of the unobserved product characteristics

ξ_{jt}. In addition, the C nesting terms $\ln\left(q_{jt}/q_{\sigma_c(jt)}\right)$ are endogenous by construction: any shock to ξ_{jt} that increases the dependent variable $\ln(q_{jt}/q_{0t})$ will also increase the nesting terms $\ln\left(q_{jt}/q_{\sigma_c(jt)}\right)$. Assuming that product characteristics are exogenous, identification requires finding at least one instrument for price and each of the C nesting terms.

We define markets as month-store pairs and products as brands. Following Bresnahan et al. (1997), we include brand name and segment fixed effects, ξ_s and ξ_b, and market-invariant continuous product characteristics $\mathbf{x}_j$ (i.e., fiber, sugar, lipid, protein, energy, and sodium). The fixed effects, ξ_b and ξ_s, capture market-invariant observed and unobserved brand name (i.e. company) and segment-specific characteristics. We also include month and store fixed effects, ξ_m and ξ_s, that capture monthly unobserved determinants of demand and time-invariant store characteristics, respectively. The structural error that remains in ξ_{jt} therefore captures the unobserved product characteristics varying across products and markets (e.g., changes in shelf-space, positioning of the products among others) that affect consumers utility and that consumers and firms (but not the modeller) observe so that they are likely to be correlated with prices.

We use two sets of instruments as sources of exogenous variations in prices and market shares. First, as cost shifters, we use the (market-level) price of sugar times the sugar content of the cereals, interacted with segment and brand name fixed effects, respectively. Multiplying the price of sugar by the sugar content allows the instrument to vary by product; and interacting this with fixed effects allows the price of sugar to enter the production function of each firm differently.

Second, we form BLP instruments by using other products' promotional activity in a given month, which varies both across stores for a given month and across months for a given store: for a given product, other products' promotional activity affects consumers' choices, and is thus correlated with the price of that product, but uncorrelated with the error term.[25] We use the number of other promoted products of rival firms and the number of other promoted products of the same firm, which we interact with brand names fixed effects. We also use these numbers over products belonging to the same segment, which we interact with segment fixed effects. We distinguish between products of the same firm and of rival firms, and interact instruments with brand name fixed effects with the idea that (equilibrium) markup is a function of the ownership structure since multi-product firms set

[25]The promotional is treated as an exogenous variable since, at Dominick's, the promotional calendar is known several weeks in advance of the weekly price decisions. In addition, we do not use functions of the continuous product characteristics as instruments since by construction of the data, for each product, they are invariant across markets (see Nevo, 2001).

prices so as to maximize their total profits. Interaction with segment fixed effects accounts for within-segment competitive conditions.

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

5.4 Related Approaches

The logit and NL models are special cases of the GNE model. They are attractive since they have analytic formulae for both their market shares and their inverse (Berry, 1994). However, they are restricted in their ability to generate different substitution patterns. In contrast, the GNE model may generate far richer substitution patterns, it is very easily estimated; but does not have an analytic formula for its market shares.

The GNE model is similar in spirit to some existing GEV models with (non)overlapping nests,[26] but has two key advantages over these models. First, the GNE has an analytic formula for the inverse market shares, which means the BLP method can be implemented while avoiding numerical inversion in the estimation process. Second, while GEV models restrict products to be substitutes; the GNE model allows also complementarity to occur.

The GNE model is linked to two other approaches. First, Hausman et al. (1994) build a three-stage demand model. Their model also requires classification of products into groups, and is estimated by linear regression. However, they treat consumers' choice as a sequence of separate but related choices, need a large number of instruments that may be difficult to find, and cannot handle large choice sets.

Second, Pinkse and Slade (2004) construct a continuous-choice demand model. Their model is also simple to estimate, allows rich substitution patterns, handles large choice sets, and makes cross-price elasticities functions of the distance of products in characteristics space. However, their model is not as parsimonious as the GNE model and requires a large number of instruments to be used.

[26] See e.g., the ordered logit models (Small, 1987; Grigolon, 2017); the PDL model (Bresnahan et al., 1997); and the flexible coefficient MNL model (Davis and Schiraldi, 2014).

5.5 Empirical results

Demand parameters. Table 1 presents 2SLS estimates of demand parameters from the GNE model and the three-level NL models with nests for segment on top and with nests for brand on top, respectively.

Table 1: PARAMETER ESTIMATES OF DEMAND

	(1) GNE		(2) 3NL1		(3) 3NL2	
Price ($-\alpha$)	-1.831	(0.116)	-2.908	(0.118)	-4.101	(0.156)
Promotion (β)	0.0882	(0.00278)	0.102	(0.00305)	0.144	(0.00365)
Constant (β_0)	-0.697	(0.0593)	-0.379	(0.0645)	-0.195	(0.0755)
Nesting Parameters (μ)						
Segment/nest (μ_1)	0.626	(0.00931)	0.771	(0.00818)	0.668	(0.0109)
Brand/subnest (μ_2)	0.232	(0.00944)	0.792	(0.00725)	0.709	(0.00961)
FE Segments (γ)						
Health/nutrition (γ_H)	-0.672	(0.00990)	-0.855	(0.00751)	-0.0693	(0.00538)
Kids (γ_K)	-0.433	(0.00875)	-0.529	(0.00869)	0.0705	(0.00522)
Taste enhanced (γ_T)	-0.710	(0.0102)	-0.903	(0.00747)	-0.0877	(0.00558)
FE Brand Names (θ)						
Kellogg's (θ_K)	0.0243	(0.00460)	-0.0563	(0.00344)	0.104	(0.00635)
Nabisco (θ_N)	-0.754	(0.0242)	-0.218	(0.0109)	-2.105	(0.0201)
Post (θ_P)	-0.485	(0.0144)	-0.187	(0.00830)	-1.364	(0.00931)
Quaker (θ_Q)	-0.553	(0.0150)	-0.329	(0.0137)	-1.508	(0.00653)
Ralston (θ_R)	-0.732	(0.0249)	-0.200	(0.0111)	-2.131	(0.0211)
Observations	99281		99281		99281	
RMSE	0.210		0.242		0.270	

Notes: The dependent variable is $\ln(q_{jt}/q_{0t})$. Regressions include fixed effects (FE) for brand names and segments, months, and stores, as well as the market-invariant continuous product characteristics (fiber, sugar, lipid, protein, energy, and sodium). Robust standard errors are reported in parentheses. The values of the F-statistics in the first stages suggest that weak instruments are not a problem.

Consider first the results from the GNE model. The estimated parameters on the negative of price (α) and on promotion (β) are significantly positive. The estimated nesting parameters ($0 < \mu_2 < \mu_1 < 1$) are consistent with the GEM ($\mu_1 + \mu_2 < 1$); this provides an empirical check on the appropriateness of the GEM as the constraint was not imposed on the estimates. The parameter estimates imply that there is product segmentation along both dimensions: products with the same brand name are closer substitutes than products with different brand names; and products within the same segment are closer substitutes than products from 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 (the γ's and θ's) and the nesting parameters (μ_1 and μ_2). The fixed effects measure the extent to which belonging to a nest shifts the demand for the product, and the nesting parameters measure the extent to which products within a nest are protected from competition from products from different nests along each dimension.

We find that the brand-name reputation of the cereals confers a significant advantage on products from General Mills and Kellogg's ($\theta_K > \theta_G = 0 > \theta_P > \theta_Q > \theta_R > \theta_N$); and cereals for family also benefit from a significant advantage ($\gamma_F = 0 > \gamma_K > \gamma_H > \gamma_T$). In addition, we find that $\mu_1 > \mu_2$, i.e., the segments confer more protection from competition than brand-name reputation does (products within the same segment are more protected from products from 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 ($\mu_2 > \mu_1$), which means that it is not possible to decide between them based on this criterion. However, the Rivers and Vuong (2002) test very clearly reject both NL models in favor of the GNE. [27]

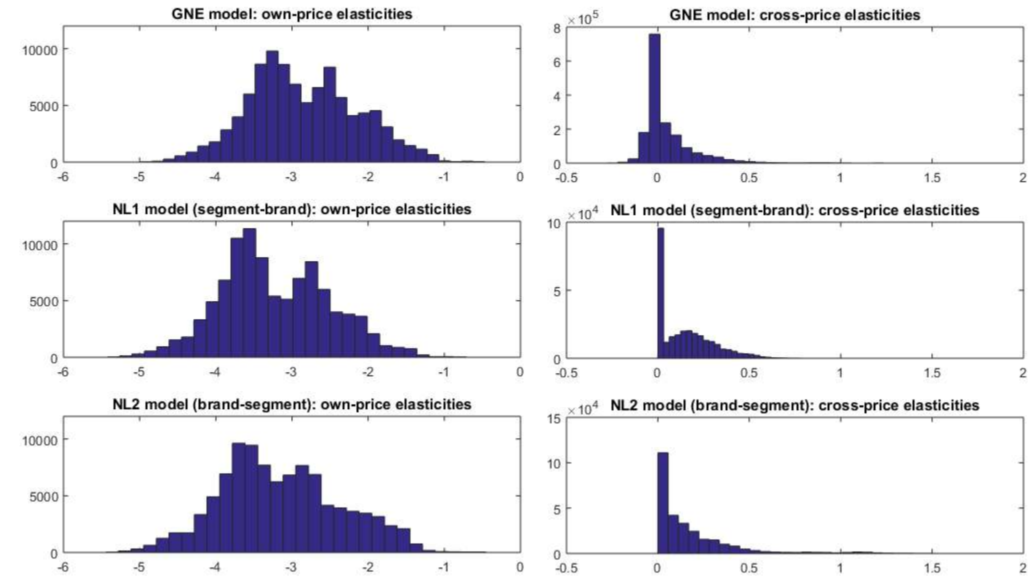
Alternative specification with very large choice set. We have estimated an alternative model in which all brand-store combinations are considered as products while markets are taken to be months. The resulting model has more than 4,000 products, but was estimated very quickly without any issues. This shows the ability of the GNE model to deal with large choice sets.

The parameter estimates were not significantly affected by this change in specification, which indicates that the results are fairly robust.

Substitution patterns. Figure 3 presents the estimated density of the own- and cross-price elasticities of demands of the GNE and NL models (see Tables 9 and 10 for the estimated own- and cross-price elasticities of demands, averaged over markets and product types).

[27] The statistics of the test of the two NL models (model 1) against the GNE model (model 2), 2891.97 and 4879.82, are evaluated against the standard normal distribution. Each statistic is given by $T_N = \frac{\sqrt{N}}{\hat{\sigma}}\left(\hat{Q}_1 - \hat{Q}_2\right)$, where N is the number of observations, $\hat{Q}_i$ is the value of the estimated RMSE of model i, and $\hat{\sigma}^2$ is the estimated value of the variance of the difference between $\hat{Q}_i$'s. The variance $\hat{\sigma}^2$ was estimated using 500 bootstrap replications.

Figure 3: ESTIMATED ELASTICITIES



The estimated own-price elasticities are in line with the literature (see e.g., Nevo, 2001). On average, the estimated own-price elasticity of demands is -2.815 for the GNE model. However, there is an important variation in price responsiveness across product types: demands for cereals for kids produced by General Mills exhibit a much higher own-price elasticity than cereals for health/nutrition produced by Post (-3.427 vs. -1.524).

Consider the cross-price elasticities. Among the 17×50 different cross-price elasticities in the GNE model, 48.5 percent (resp., 51.5 percent) are negative (resp., positive), meaning that some cereals are substitutes, while others are complements. For example, cereals for families produced by General Mills are complementary with those with taste enhanced produced by Kellogg's; but are substitutable with those for kids produced by General Mills.

6 Conclusion

We have formulated the class of generalized utility models (GEM) and shown that have a number of useful properties. The GEM class belongs to the class of perturbed utility models and incorporates all additive random utility discrete choice models. GEM demands have a tractable and familiar form that generalizes the logit demand.

We have shown how GEM can be specified in terms of a generator function. Different GE models can be obtained from different specifications of the generator. We have proposed a general nesting operation for constructing generators and demonstrated its use in an empirical application.

GE models are useful for implementing the BLP method, exploiting its advantages for handling endogeneity issues, while avoiding the issues involved in the numerical inversion of demand during estimation. This is because GE models give the structural error term directly as a known function of the data and parameters, so that only standard regression techniques are required. At the same time, GEM are able to accommodate a large variety substitution patterns. In contrast to ARUM and its generalizations such as the mixed logit model, the class of GEM comprises models that allow alternatives to be complements.

We have built the GNE model, a special case of a GEM, for describing markets with product segmentation along several dimensions and we employ it to estimate a GNE using a large real-world dataset. The GNE model is a serious competitor to the multi-level NL models and the PDL model of Bresnahan et al. (1997): it improves on these models by allowing rich patterns of substitution and complementarity, while being parsimonious and computationally fast, and easily estimated by linear regression. In fact, we are able estimate a version of the GNE model with 4,000 products at small computational effort and encountering no numerical issues.

With the GEM, we have opened the door to a new universe of models. It is is larger than the universe of ARUM and much remains to be explored. The further development of GEM provides many opportunities for research, the main two are perhaps the following. First, it would be desirable to develop ways to estimate GEM using individual-level discrete choice data. Second, it would be very desirable to develop dynamic GEM parallelling the dynamic discrete choice model Rust (1987). This would be useful for describing situations involving forward looking behavior.

References

ALLEN, R. AND J. REHBECK (2016): "Complementarity in perturbed utility models," University of California-San Diego.

ANDERSON, S. P., A. DE PALMA, AND J. F. THISSE (1988): "A representative consumer theory of the logit model," *International Economic Review*, 461–466.

ARMSTRONG, T. B. (2016): "Large market asymptotics for differentiated product demand estimators with economic models of supply," *Econometrica*, 84, 1961–1980.

BECKERT, W. AND R. BLUNDELL (2008): "Heterogeneity and the non-parametric analysis of consumer choice: conditions for invertibility," *The Review of Economic Studies*, 75, 1069–1080.

BERRY, S., A. GANDHI, AND P. HAILE (2013): "Connected substitutes and invertibility of demand," *Econometrica*, 81, 2087–2111.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): "Automobile prices in market equilibrium," *Econometrica*, 841–890.

BERRY, S. T. (1994): "Estimating discrete-choice models of product differentiation," *The RAND Journal of Economics*, 242–262.

BERRY, S. T. AND P. A. HAILE (2014): "Identification in differentiated products markets using market level data," *Econometrica*, 82, 1749–1797.

BOYD, S. AND L. VANDENBERGHE (2004): *Convex optimization*, Cambridge university press.

BRENKERS, R. AND F. VERBOVEN (2006): "Liberalizing a distribution system: the European car market," *Journal of the European Economic Association*, 4, 216–251.

BRESNAHAN, T. F., S. STERN, AND M. TRAJTENBERG (1997): "Market Segmentation and the Sources of Rents from Innovation: Personal Computers in the Late 1980s," *RAND Journal of Economics*, S17–S44.

CHIAPPORI, P. AND I. KOMUNJER (2009): "On the nonparametric identification of multiple choice models," Columbia University.

CHIONG, K. X., A. GALICHON, AND M. SHUM (2016): "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7, 83–115.

COWELL, F. A. (2000): "Measurement of inequality," *Handbook of income distribution*, 1, 87–166.

DAVIS, P. AND P. SCHIRALDI (2014): "The flexible coefficient multinomial logit (FC-MNL) model of demand for differentiated products," *The RAND Journal of Economics*, 45, 32–63.

DUBÉ, J.-P., J. T. FOX, AND C.-L. SU (2012): "Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation," *Econometrica*, 80, 2231–2267.

EDWARDS, R. D. AND S. TULJAPURKAR (2005): "Inequality in life spans and a new perspective on mortality convergence across industrialized countries," *Population and Development Review*, 31, 645–674.

ERLANDER, S. (1977): "Accessibility, entropy and the distribution and assignment of traffic," *Transportation Research*, 11, 149–153.

FOSGERAU, M., D. MCFADDEN, AND M. BIERLAIRE (2013): "Choice probability generating functions," *Journal of Choice Modelling*, 8, 1–18.

FOSGERAU, M., E. MELO, A. DE PALMA, AND M. SHUM (2017): "Discrete Choice and Rational Inattention: A General Equivalence Result," *Working Paper HAL.*

FUDENBERG, D., R. IIJIMA, AND T. STRZALECKI (2015): "Stochastic choice and revealed perturbed utility," *Econometrica*, 83, 2371–2409.

GALE, D. AND H. NIKAIDO (1965): "The Jacobian matrix and global univalence of mappings," *Mathematische Annalen*, 159, 81–93.

GALICHON, A. AND B. SALANIÉ (2015): "Cupid's invisible hand: Social surplus and identification in matching models," .

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

GEORGESÇU-ROEGEN, N. (1971): *The entropy law and the economic process*, A Harvard paperback: Economics, Harvard University Press.

GOLDBERG, P. K. (1995): "Product differentiation and oligopoly in international markets: The case of the US automobile industry," *Econometrica*, 891–951.

GRIGOLON, L. (2017): "Blurred boundaries: a flexible approach for segmentation applied to the car market," .

HAUSMAN, J., G. LEONARD, AND J. D. ZONA (1994): "Competitive analysis with differenciated products," *Annales d'Economie et de Statistique*, 159–180.

HOFBAUER, J. AND W. H. SANDHOLM (2002): "On the global convergence of stochastic fictitious play," *Econometrica*, 70, 2265–2294.

KNITTEL, C. R. AND K. METAXOGLOU (2014): "Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective," *Review of Economics and Statistics*, 96, 34–59.

LEE, J. AND K. SEO (2015): "A computationally fast estimator for random coefficients logit demand models using aggregate data," *The RAND Journal of Economics*, 46, 86–102.

MAS-COLELL, A., M. D. WHINSTON, AND J. R. GREEN (1995): *Microeconomic theory*, vol. 1, Oxford university press New York.

MATEJKA, F. AND A. MCKAY (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *The American Economic Review*, 105, 272–298.

MATTSSON, L.-G. AND J. W. WEIBULL (2002): "Probabilistic choice and procedurally bounded rationality," *Games and Economic Behavior*, 41, 61–78.

MCELROY, F. W. (1969): "Returns to scale, Euler's theorem, and the form of production functions," *Econometrica*, 275–279.

MCFADDEN, D. (1974): "Conditional logit analysis of qualitative choice behavior," *Frontiers in Econometrics*, 105–142.

——— (1981): "Econometric models of probabilistic choice," in *Structural Analysis of Discrete Data with Econometric Applications*, ed. by C. F. Manski and D. McFadden. Cambridge: MIT Press.

MCFADDEN, D. L. AND M. FOSGERAU (2012): "A theory of the perturbed consumer with general budgets," Tech. rep., National Bureau of Economic Research.

MRAZOVA, M., J. P. NEARY, AND M. PARENTI (2017): "Sales and Markup Dispersion: Theory and Empirics," .

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

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

REYNAERT, M. AND F. VERBOVEN (2014): "Improving the performance of random coefficients demand models: the role of optimal instruments," *Journal of Econometrics*, 179, 83–98.

RIVERS, D. AND Q. VUONG (2002): "Model selection tests for nonlinear dynamic models," *The Econometrics Journal*, 5, 1–39.

RUST, J. (1987): "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica*, 999–1033.

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

SANDERSON, E. AND F. WINDMEIJER (2016): "A weak instrument F-test in linear IV models with multiple endogenous variables," *Journal of Econometrics*, 190, 212–221.

SHANNON, C. E. (1948): "A Mathematical Theory of Communication," *Bell System Technical Journal*, 27, 379–423.

SHORROCKS, A. F. (1980): "The class of additively decomposable inequality measures," *Econometrica*, 613–625.

SIMON, C. P. AND L. BLUME (1994): *Mathematics for economists*, vol. 7, Norton New York.

SKRAINKA, B. S. AND K. L. JUDD (2011): "High performance quadrature rules: How numerical integration affects a popular model of product differentiation," .

SMALL, K. A. (1987): "A discrete choice model for ordered alternatives," *Econometrica*, 409–424.

TRAIN, K. E. (2009): *Discrete choice methods with simulation*, Cambridge university press.

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

VIVES, X. (2001): *Oligopoly pricing: old ideas and new tools*, MIT press.

WILSON, A. G. (2011): *Entropy in urban and regional modelling*, vol. 1, Routledge.

For Online Publication

A Preliminaries

Lemma 3. 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 $f(\mathbf{x}) = F(\phi(\mathbf{x}))$, with $\mathbf{x} = (x_1, \ldots, x_n)$, and $h : \mathbb{R} \to \mathbb{R}$ by $h = F^{-1}$. Assume that ϕ is linearly homogeneous. Then,

a. (Euler equation for homogeneous functions)

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

b. (Generalized Euler equation for homothetic functions (McElroy, 1969)) If F is non-decreasing, then f is homothetic, and

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

Proof. **a.** See e.g., proof of Theorem M.B.2. in Mas-Colell et al. (1995).
b. Consider $h(y) = \phi(\mathbf{x})$. Differentiate with respect to x_i and rearrange terms to get

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

Then

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

where the last equality uses a. applied to the homogeneous function ϕ. Multiplying both side by y yields the required equality. □

A matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ is said to be positive quasi-definite if its symmetric part $\frac{1}{2}(\mathbf{A} + \mathbf{A}^\intercal)$ is positive definite.

Lemma 4 (Gale and Nikaido 1965, Theorem 6)**.** If a differentiable mapping $\mathbf{F} : \Theta \to \mathbb{R}^n$, where Θ is a convex region (either closed or non-closed) of $\mathbb{R}^n$, has a Jacobian matrix that is everywhere quasi-definite in Θ, then $\mathbf{F}$ is injective on Θ.

Lemma 5 (Simon and Blume, 1994, Theorem 14.4)**.** Let $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$ and $\mathbf{G} : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable functions. Let $\mathbf{y} \in \mathbb{R}^n$ and $\mathbf{x} = \mathbf{G}(\mathbf{y}) \in \mathbb{R}^n$. Consider the composite function

$$\mathbf{C} = \mathbf{F} \circ \mathbf{G} : \mathbb{R}^n \to \mathbb{R}^n.$$

Let $\mathbf{J_F}(\mathbf{x}) \in \mathbb{R}^{n \times n}$ be the Jacobian matrix of the partial derivatives of $\mathbf{F}$ at $\mathbf{x}$, and let $\mathbf{J_G}(\mathbf{y}) \in \mathbb{R}^{n \times n}$ be the Jacobian matrix of the partial derivatives of $\mathbf{G}$ at $\mathbf{y}$. Then the Jacobian matrix $\mathbf{J_C}(\mathbf{y})$ is given by the matrix product of the Jacobians:

$$\mathbf{J_C}(\mathbf{y}) = \mathbf{J_{F \circ G}}(\mathbf{y}) = \mathbf{J_F}(\mathbf{x})\,\mathbf{J_G}(\mathbf{y}).$$

Lemma 6 (Chain rule – inverse function)**.** Let $\mathbf{F} = \mathbf{K}^{-1}$ and $\mathbf{G} = \mathbf{K}$, then $\mathbf{C} = \mathbf{F} \circ \mathbf{G}$ is the identity function, whose Jacobian matrix is the identity matrix. In this special case, we have

$$\mathbf{J_C}(\mathbf{y}) = \mathbf{J_{F \circ G}}(\mathbf{y}) = \mathbf{I}_n = \mathbf{J}_{\mathbf{K}^{-1}}(\mathbf{K}(\mathbf{y}))\,\mathbf{J_K}(\mathbf{y}),$$

which solving for $\mathbf{J}_{\mathbf{K}^{-1}}(\mathbf{K}(\mathbf{y}))$ gives

$$\mathbf{J}_{\mathbf{K}^{-1}}(\mathbf{K}(\mathbf{y})) = [\mathbf{J_K}(\mathbf{y})]^{-1},$$

or equivalently solving for $\mathbf{J_K}(\mathbf{y})$ gives

$$\mathbf{J_K}(\mathbf{y}) = [\mathbf{J}_{\mathbf{K}^{-1}}(\mathbf{K}(\mathbf{y}))]^{-1}.$$

B Proofs for Section 2

Proof of Lemma 1. Lemma 1 is implied by Lemma 7 below. □

Lemma 7. Assume that $\mathbf{S}$ is twice continuously differentiable and linearly homogeneous. Then,

a. $\mathbf{J}_{\ln \mathbf{S}}$ is symmetric on $\mathrm{int}(\Delta)$ if and only if

$$\sum_{j=0}^{J} q_j \frac{\partial \ln S^{(j)}(\mathbf{q})}{\partial q_k} = 1,\ k \in \mathscr{J},\ \forall \mathbf{q} \in \operatorname{int}(\Delta). \tag{22}$$

b. If $\mathbf{J}_{\ln \mathbf{S}}$ is symmetric and positive definite on $\operatorname{int}(\Delta)$, then Ω is strictly concave on $\operatorname{int}(\Delta)$.

c. If $\mathbf{J}_{\ln \mathbf{S}}$ is positive definite, then $\mathbf{S}$ is invertible on $\operatorname{int}(\Delta)$.

Proof of Lemma 7. a. Assume that $\mathbf{J}_{\ln \mathbf{S}}$ is symmetric. $S^{(k)}$ is linearly homogeneous, then $\ln S^{(k)}$ is homothetic.

Let $\phi(\mathbf{q}) = S^{(k)}(\mathbf{q})$ and $F(\mathbf{q}) = \ln(\mathbf{q})$, then $h(\delta) = \exp(\delta)$. Define $\delta = f(\mathbf{q}) = F(\phi(\mathbf{q})) = \ln\left(S^{(k)}(\mathbf{q})\right)$ and $h(\delta) = \phi(\mathbf{q}) = \exp(\delta)$. Then, by Lemma 3, $S^{(k)}$ satisfies

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

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

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

Assume now that $\sum_{j=0}^{J} q_j \frac{\partial \ln S^{(j)}(q)}{\partial q_k} = 1$. Then, for each $j, k \in \mathscr{J}$,

$$\frac{\partial \Omega(\mathbf{q})}{\partial q_j} = -\ln S^{(j)}(\mathbf{q}) - 1; \quad \frac{\partial \Omega(\mathbf{q})}{\partial q_k} = -\ln S^{(k)}(\mathbf{q}) - 1,$$

so that

$$\frac{\partial^2 \Omega(\mathbf{q})}{\partial q_j \partial q_k} = -\frac{\partial \ln S^{(j)}(\mathbf{q})}{\partial q_k}; \quad \frac{\partial^2 \Omega(\mathbf{q})}{\partial q_k \partial q_j} = -\frac{\partial \ln S^{(k)}(\mathbf{q})}{\partial q_j}.$$

Since Ω is twice continuously differentiable, then by Schwarz's theorem,

$$\frac{\partial^2 \Omega(\mathbf{q})}{\partial q_j \partial q_k} = \frac{\partial^2 \Omega(\mathbf{q})}{\partial q_k \partial q_j},$$

i.e.,

$$\frac{\partial \ln S^{(j)}(\mathbf{q})}{\partial q_k} = \frac{\partial \ln S^{(k)}(\mathbf{q})}{\partial q_j},$$

Then $\mathbf{J}_{\ln \mathbf{S}}$ is symmetric as required.

b. From Part 1, we find that $\mathbf{J}_{\ln \mathbf{S}}(\mathbf{q}) = \nabla^2_{\mathbf{q}}(-\Omega(\mathbf{q}))$, for all $\mathbf{q} \in \text{int}(\Delta)$. Then Ω is strictly concave by positive definiteness of $\mathbf{J}_{\ln \mathbf{S}}$.

c. The function $\ln \mathbf{S}$ is differentiable on the convex region $\text{int}(\Delta)$ of $\mathbb{R}^{J+1}$. In addition, $\mathbf{J}_{\ln \mathbf{S}}$ is positive quasi-definite on $\text{int}(\Delta)$, since its symmetric part $\frac{1}{2}\left(\mathbf{J}_{\ln \mathbf{S}} + (\mathbf{J}_{\ln \mathbf{S}})^{\intercal}\right) = \mathbf{J}_{\ln \mathbf{S}}$ is positive definite on $\text{int}(\Delta)$. Then, by Lemma 4, $\ln \mathbf{S}$ is injective, implying that $\mathbf{S}$ is injective. $\square$

Proof of Theorem 1. ***GE Demands (4).*** The Lagrangian of the GEM is

$$\mathcal{L}(\mathbf{q}, \lambda, \lambda_0, \ldots, \lambda_J) = \alpha y + \sum_{j=0}^{J} \delta_j q_j - \sum_{j=0}^{J} q_j \ln S^{(j)}(\mathbf{q}) + \lambda \left(1 - \sum_{j=0}^{J} q_j\right) + \sum_{j=0}^{J} \lambda_j q_j,$$

where $\lambda \geq 0$ and $\lambda_j \geq 0$ for all $j \in \mathscr{J}$.

The first-order conditions are

$$\delta_i - \ln S^{(i)}(\mathbf{q}) - \sum_{j=0}^{J} q_j \frac{\partial \ln S^{(j)}(\mathbf{q})}{\partial q_k} - \lambda + \lambda_i = 0, \quad i \in \mathscr{J},$$
$$\sum_{j=0}^{J} q_j = 1.$$

Using Lemma 1, we get

$$\delta_i - \ln S^{(i)}(\mathbf{q}) - 1 - \lambda + \lambda_i = 0, \quad i \in \mathscr{J},$$
$$\sum_{j=0}^{J} q_j = 1.$$

Observe that if $\mathbf{q} \in \text{bd}(\Delta)$, then $|\ln \mathbf{S}(\mathbf{q})| = +\infty$, by Assumption 1. Hence, $\mathbf{q}$ cannot solve the first-order conditions, since the λ_i's must be finite. Therefore the solution must be interior with $\lambda_i = 0$ for all $i \in \mathscr{J}$. Then the first-order conditions reduce to

$$\mathbf{S}(\mathbf{q}) = e^{\boldsymbol{\delta} - 1 - \lambda} > 0, \tag{23}$$

$$\sum_{j=0}^{J} q_j = 1. \tag{24}$$

The linear homogeneity of $\mathbf{S}$ implies that also $\mathbf{H} = \mathbf{S}^{-1}$ is linearly homogeneous. Then

(23) yields

$$\mathbf{q} = \mathbf{S}^{-1}\left(e^{\boldsymbol{\delta}-1-\lambda}\right) = \mathbf{H}\left(e^{\boldsymbol{\delta}-1-\lambda}\right) = e^{-(1+\lambda)}\mathbf{H}\left(e^{\boldsymbol{\delta}}\right).$$

Lastly, (24) implies that $e^{1+\lambda} = \sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)$ such that any solution to the first-order conditions satisfies

$$q_i = \frac{H^{(i)}\left(e^{\boldsymbol{\delta}}\right)}{\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)}, \quad i \in \mathscr{J}. \tag{25}$$

The strict concavity of the utility u on $\operatorname{int}(\Delta)$ implies that this solution is unique and is the argmax to the utility maximization problem.

***Relation (5) between $\boldsymbol{\delta}$ and* $\mathbf{q}$.** Note that if $\mathbf{q}$ is an interior solution to the utility maximization problem then it satisfies Equation (4), which, by invertibility and linear homogeneity of $\mathbf{S}$ implies that

$$\ln S^{(i)}(\mathbf{q}) + \ln\left(\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)\right) = \delta_i, \quad i \in \mathscr{J}.$$

Conversely, if $\forall i \in \mathscr{J}$, we have $\delta_i = \ln S^{(i)}(\mathbf{q}) + \ln\left(\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)\right)$, then $\mathbf{q}$ solves (4). $\square$

Proof of Proposition 1. The surplus function G is defined by

$$G(\boldsymbol{\delta}) = \sum_{j=0}^{J} \delta_j q_j(\boldsymbol{\delta}) + \Omega(\mathbf{q}(\boldsymbol{\delta})),$$

with $q_j(\boldsymbol{\delta})$ given by (4). The log-sum (7) results substituting $q_j(\boldsymbol{\delta})$ by (4).

We now show that demands (4) satisfy Roy's identity, i.e.,

$$q_j(\boldsymbol{\delta}) = \frac{\partial G(\boldsymbol{\delta})}{\partial \delta_j}.$$

Let $\boldsymbol{\delta} = \ln \mathbf{S}(\mathbf{q})$, so that $(\ln \mathbf{S})^{-1}(\boldsymbol{\delta}) = \mathbf{H} \circ \exp(\boldsymbol{\delta}) = \mathbf{q}$. Then, by Lemma 6

$$\mathbf{J}_{\ln \mathbf{S}}(\mathbf{q}) = \left[\mathbf{J}_{(\ln \mathbf{S})^{-1}}(\ln \mathbf{S}(\mathbf{q}))\right]^{-1} = \left[\mathbf{J}_{\mathbf{H}\circ\exp}(\boldsymbol{\delta})\right]^{-1}. \tag{26}$$

Since $\mathbf{J}_{\ln \mathbf{S}}(\mathbf{q})$ is symmetric, $\mathbf{J}_{\mathbf{H}\circ\exp}$ is also symmetric, i.e.,

$$\frac{\partial H^{(i)}\left(e^{\boldsymbol{\delta}}\right)}{\partial \delta_j} = \frac{\partial H^{(j)}\left(e^{\boldsymbol{\delta}}\right)}{\partial \delta_i}. \tag{27}$$

This is because every positive definite matrix is invertible, and the inverse of a symmetric matrix is also a symmetric matrix. Then,

$$\begin{aligned}\frac{\partial G\left(e^{\boldsymbol{\delta}}\right)}{\partial \delta_i} &= \frac{\sum_{k=0}^{J} \frac{\partial H^{(k)}\left(e^{\boldsymbol{\delta}}\right)}{\partial \delta_i}}{\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)} = \frac{\sum_{k=0}^{J} \frac{\partial H^{(i)}\left(e^{\boldsymbol{\delta}}\right)}{\partial \delta_k}}{\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)}, \\ &= \frac{\sum_{k=0}^{J} \frac{\partial H^{(i)}\left(e^{\boldsymbol{\delta}}\right)}{\partial e^{\delta_k}} e^{\delta_k}}{\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)} = \frac{H^{(i)}\left(e^{\boldsymbol{\delta}}\right)}{\sum_{j=0}^{J} H^{(j)}\left(e^{\boldsymbol{\delta}}\right)},\end{aligned}$$

where the second equality comes from the symmetry of $\mathbf{J}_{\mathbf{H}\circ\exp}$, and the last equality comes from the Euler equation (in Lemma 3) applied to the linearly homogeneous function $H^{(i)}$.

□

Proof of Proposition 2. From Theorem, 1 (5) we obtain

$$\mathbf{I} = \mathbf{J}_{\ln \mathbf{S}}(\mathbf{q})\mathbf{J}_{\mathbf{q}} + \mathbf{1}\mathbf{q}^{\intercal}.$$

This can be solved to obtain the desired result since $\mathbf{J}_{\ln \mathbf{S}}(\mathbf{q})$ is invertible. □

Proof of Proposition 4. Consider the generator $\mathbf{S}$ in Example 3 and write the corresponding first-order conditions (23) and (24). Differentiating them with respect to δ_0, we obtain the following system of equations:

$$\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{\mu}{q_0} + \frac{1-\mu}{q_0+q_1/2} & \frac{(1-\mu)/2}{q_0+q_1/2} & 0 & 1 \\ \frac{(1-\mu)/2}{q_0+q_1/2} & \frac{\mu}{q_1} + \frac{(1-\mu)/4}{q_0+q_1/2} + \frac{(1-\mu)/4}{q_1/2+q_2} & \frac{(1-\mu)/2}{q_1/2+q_2} & 1 \\ 0 & \frac{(1-\mu)/2}{q_1/2+q_2} & \frac{\mu}{q_2} + \frac{(1-\mu)}{q_1/2+q_2} & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \frac{\partial q_0}{\partial \delta_0} \\ \frac{\partial q_1}{\partial \delta_0} \\ \frac{\partial q_2}{\partial \delta_0} \\ \frac{\partial \lambda}{\partial \delta_0} \end{pmatrix}$$

This can be solved to find that $\partial q_2/\partial \delta_0 > 0$ if and only if

$$\mu < \frac{q_1^2 + q_0 q_1 + q_1 q_2}{4q_0q_2 + 3q_1q_2 + 2q_1^2 + 3q_0q_1},$$

and noting that $q_0 + q_1 + q_2 = 1$, if and only if

$$\mu < \frac{q_1}{4q_0q_2 + 3q_1q_2 + 2q_1^2 + 3q_0q_1}.$$

At $\boldsymbol{\delta}$ such that $q_0 = q_1 = q_2 = 1/3$, the condition becomes $\mu < 1/4$, thus showing that there exists combinations of parameters μ and utilities δ at which some alternatives are complements. $\square$

C Appendix for Section 3

Define $\Lambda = \left\{\boldsymbol{\delta} : \sum_j \delta_j = 0\right\}$ as the tangent space of Δ. The following lemma collects some properties of the expected maximum utility $\overline{G}$.

Lemma 8. The surplus $\overline{G}$ has the following properties.

a. $\overline{G}$ is twice continuously differentiable, convex and finite everywhere.

b. $\overline{G}\left(\boldsymbol{\delta} + c\mathbf{1}\right) = \overline{G}\left(\boldsymbol{\delta}\right) + c$ for any $c \in \mathbb{R}$.

c. The Hessian of $\overline{G}$ is positive definite on Λ.

d. $\overline{G}$ is given in terms of the expected residual of the maximum utility alternative by

$$\overline{G}\left(\boldsymbol{\delta}\right) = \sum_{j=0}^{J} P_j\left(\boldsymbol{\delta}\right)\delta_j + \mathbb{E}\left(\varepsilon_{j^*}|\boldsymbol{\delta}\right).$$

Proof of Lemma 8. McFadden (1981) establishes convexity and finiteness of $\overline{G}$ as well as the homogeneity property (b.) and the existence of all mixed partial derivatives up to order J. This also implies that all second order mixed partial derivatives are continuous, since $J \geq 2$. Hofbauer and Sandholm (2002) show that the Hessian of $\overline{G}$ is positive definite on Λ (see the proof of their Theorem 2.1).

Let j^* be the index of the chosen alternative. The last statement of the lemma follows using the law of iterated expectations:

$$\begin{aligned}
\overline{G}(\boldsymbol{\delta}) &= \sum_{j=0}^{J} \mathbb{E}\left(\max_{j\in\mathscr{J}}\{\delta_j+\varepsilon_j\}\,|j^*=j,\boldsymbol{\delta}\right)P_j(\boldsymbol{\delta}),\\
&= \sum_{j=0}^{J}\left(\delta_j+\mathbb{E}\left(\varepsilon_{j^*}|j^*=j,\boldsymbol{\delta}\right)P_j(\boldsymbol{\delta})\right),\\
&= \sum_{j=0}^{J} P_j(\boldsymbol{\delta})\,\delta_j+\mathbb{E}\left(\varepsilon_{j^*}|\boldsymbol{\delta}\right).
\end{aligned}$$

□

Proof of Lemma 2. *Invertibility of* $\overline{\mathbf{H}}$. Note first that $\overline{\mathbf{H}}$ is differentiable.

In addition, the Jacobian of $\boldsymbol{\delta}\to\overline{\mathbf{H}}\left(e^{\boldsymbol{\delta}}\right)$, labeled $\mathbf{J}_{\overline{\mathbf{H}}}$, is positive quasi-definite on Λ. The Jacobian $\mathbf{J}_{\overline{\mathbf{H}}}$ has elements ij given by

$$\left\{e^{\overline{G}(\boldsymbol{\delta})}\overline{G}_i(\boldsymbol{\delta})\,\overline{G}_j(\boldsymbol{\delta})\right\}+\left\{e^{\overline{G}(\boldsymbol{\delta})}\overline{G}_{ij}(\boldsymbol{\delta})\right\}.$$

The first matrix is positive semi-definite. By part d. of Lemma 8, the second matrix is positive definite on Λ. The Jacobian is therefore positive definite on Λ. Lastly, since $\mathbf{J}_{\overline{\mathbf{H}}}$ is symmetric, its symmetric part is itself, and thus positive quasi-definiteness of $\mathbf{J}_{\overline{\mathbf{H}}}$ is equivalent to its positive definiteness. Then, by Lemma 4, $\overline{\mathbf{H}}$ is invertible on the range $\overline{\mathbf{H}}(e^{\Lambda})$. Global invertibility follows, since by the homogeneity property we have for $\boldsymbol{\delta}\in\mathbb{R}^{J+1}$ that

$$\overline{\mathbf{H}}(e^{\boldsymbol{\delta}})=e^{\mathbf{1}^\intercal\boldsymbol{\delta}}e^{\overline{G}(\boldsymbol{\delta}-\mathbf{1}_{JJ}\boldsymbol{\delta})}\mathbf{P}(\boldsymbol{\delta}-\mathbf{1}_{JJ}\boldsymbol{\delta}).$$

The range of $\overline{\mathbf{H}}$ is $\mathbb{R}^{J+1}_{++}$ since the range of $\mathbf{P}$ is the interior of Δ. To conform to the definition of a generator, we need to extend $\overline{\mathbf{H}}$ continuously to have domain and range $\mathbb{R}^{J+1}_{+}$. Fosgerau et al. (2017, Proposition 2) show that $\overline{\mathbf{H}}$ does in fact have such a continuous and invertible extension $\overline{\overline{\mathbf{H}}}$.[28] We may therefore define a candidate generator $\overline{\mathbf{S}}:\mathbb{R}^{J+1}_{+}\to\mathbb{R}^{J+1}_{+}$ as the inverse of $\overline{\overline{\mathbf{H}}}$.

Generator $\overline{\mathbf{S}}$. The function $\overline{\mathbf{S}}$ is twice continuously differentiable and linearly homogeneous. As shown above, the Jacobian of $\overline{\mathbf{H}}$ is symmetric and positive definite. Then the same is true of the Jacobian of $\ln\overline{\mathbf{S}}$ (see Lemma 6). □

[28] The argument is fairly long, so we do not repeat it here.

D Construction of GE models

In this section, we provide a range general methods for building generators along with illustrative examples. According to Definition 1, candidate generators must be shown to be twice continuously differentiable, linearly homogeneous, and with a Jacobian of their logarithm that is symmetric and positive definite.

Constructing generators, we will encounter many instances where it is possible to construct a candidate generator that satisfies all the requirements for being a generator except the Jacobian of the log generator may be only positive semi-definite. We call such a candidate an almost generator. The first result in this section shows that averaging such an almost generator with a generator produces a new generator.

Proposition 6 (Averaging)**.** Let $\mathbf{T}_k : \mathbb{R}_+^{J+1} \to \mathbb{R}_+^{J+1}, k \in \{1, \ldots, K\}$, be almost generators with at least one being a generator. Let $(\alpha_1, \ldots, \alpha_K) \in \operatorname{int}(\Delta)$. Then $\mathbf{S} : \mathbb{R}_+^{J+1} \to \mathbb{R}_+^{J+1}$ given by

$$\mathbf{S}(\mathbf{q}) = \prod_{k=1}^{K} \mathbf{T}_k(\mathbf{q})^{\alpha_k} \tag{28}$$

is a generator.

Proof of Proposition 6. S given by (28) is twice continuously differentiable. It is also linearly homogeneous since for $\lambda > 0$

$$\begin{aligned}
\mathbf{S}(\lambda \mathbf{q}) &= \prod_{k=1}^{K} \mathbf{T}_k(\lambda \mathbf{q})^{\alpha_k} = \prod_{k=1}^{K} \lambda^{\alpha_k} \mathbf{T}_k(\mathbf{q})^{\alpha_k}, \\
&= \left(\prod_{k=1}^{K} \lambda^{\alpha_k}\right)\left(\prod_{k=1}^{K} \mathbf{T}_k(\mathbf{q})^{\alpha_k}\right), \\
&= \left(\lambda^{\sum_{k=1}^{K} \alpha_k}\right)\left(\prod_{k=1}^{K} \mathbf{T}_k(\mathbf{q})^{\alpha_k}\right) = \lambda \mathbf{S}(\mathbf{q}),
\end{aligned}$$

where the second equality comes from the linear homogeneity of the functions $\mathbf{T}_k$ and the fourth equality comes from the restrictions on parameters $\sum_{k=1}^{K} \alpha_k = 1$.

The Jacobian of $\ln \mathbf{S}$, given by $\mathbf{J}_{\ln \mathbf{S}} = \sum_{k=1}^{K} \alpha_k \mathbf{J}_{\ln \mathbf{T}_k}$, is symmetric as the linear combination of symmetric matrices; and positive definite as the linear combination of at most $K-1$ positive semi-definite matrices and at least one positive definite matrix. □

Proposition 6 has two corollaries: Proposition 3 stated in the main text and Corollary 1 given below.

Proof of Proposition 3. For each $g \in \mathscr{G}$, let $\mathbf{T}_g = \left(T_g^{(1)}, \ldots, T_g^{(J)}\right)$ with $T_g^{(j)}(\mathbf{q}) = q_g^{\mathbf{1}\{j \in g\}}$, and let $T_0^{(j)}(\mathbf{q}) = q_j$. Then the Jacobian of $\ln \mathbf{T}_g$ has elements jk given by $\frac{\mathbf{1}\{j \in g\}\mathbf{1}\{k \in g\}}{q_g}$, and thus $\mathbf{J}_{\ln \mathbf{T}_g} = \frac{\mathbf{1}_g \mathbf{1}_g^\intercal}{q_g}$ where $\mathbf{1}_g = (\mathbf{1}\{1 \in g\}, \ldots, \mathbf{1}\{J \in g\})^\intercal$. Each $\mathbf{T}_g, g \in \mathscr{G}$ is an almost generator while $\mathbf{T}_0$ is the logit generator. Lastly, $\sum_{\{g \in \mathcal{G} | j \in g\}} \mu_g + \mu_0 = 1$. Then the conditions for application of Proposition 6 are fulfilled. $\square$

The following corollary provides another application of Proposition 6, which allows to build models with analytic formulae for both the demand functions and their inverse, as it is the case for the logit and NL models. Let un-normalized demands $\tilde{\mathbf{q}}$ be demands obtained before normalizing their sum to 1, i.e., $\mathbf{q} = \tilde{\mathbf{q}}/|\tilde{\mathbf{q}}|$.

Corollary 1 (Invertible nesting)**.** Let $\mathscr{G} = \{g_0, \ldots, g_J\}$ be a finite set of $J+1$ nests (i.e., the number of nests is equal to the number of products). Let $\mu_g > 0$, for all $g \in \mathscr{G}$, be the associated nesting parameters where $\sum_{\{g \in \mathscr{G} | j \in g\}} \mu_g = 1$ for all $j \in \mathscr{J}$, and $q_g = \sum_{i \in g} q_i$. Let $\mathbf{S}$ be given by

$$S^{(j)}(\mathbf{q}) = \prod_{\{g \in \mathscr{G} | j \in g\}} q_g^{\mu_g}. \tag{29}$$

Let $\mathbf{W} = \text{diag}(\mu_{g_0}, \ldots, \mu_{g_J})$ and let $\mathbf{M} \in \mathbb{R}^{(J+1)\times(J+1)}$ with entries $M_{jk} = \mathbf{1}_{\{j \in g_k\}}$ (where rows correspond to products and columns to nests). If $\mathbf{M}$ is invertible, then $\mathbf{S}$ is a generator, and the un-normalized demands satisfy

$$\boldsymbol{\delta} = \ln \mathbf{S}(\tilde{\mathbf{q}}) \Leftrightarrow \tilde{\mathbf{q}} = \left(\mathbf{M}^T\right)^{-1} \exp\left(\mathbf{W}^{-1}\mathbf{M}^{-1}\boldsymbol{\delta}\right).$$

The generator (29) satisfies Assumption 1 only when there is at least one degenerate nest (i.e., a nest with a single product). This means that Corollary 1 allows for zero demands when there is no degenerate nest. Note that zero demands may also arise in an ARUM where the error terms have bounded support.

Proof of Corollary 1. Following the proof of Proposition 3, the (candidate) generator $\mathbf{S}$ given by (29) is clearly an almost generator. Thus, it remains to show that, if $\mathbf{M}$ is invertible, then the Jacobian of $\ln \mathbf{S}$ is positive definite.

Note that

$$\ln S^{(j)}(\mathbf{q}) = \sum_{k=0}^{J} \mu_{g_k} \mathbf{1}\{j \in g_k\} \ln(q_{g_k}),$$
$$= \sum_{k=0}^{J} \mu_{g_k} \mathbf{1}\{j \in g_k\} \ln\left(\sum_{i=0}^{J} \mathbf{1}\{i \in g_k\} q_i\right),$$

and, in turn,

$$\frac{\partial \ln S^{(j)}(\mathbf{q})}{\partial q_l} = \sum_{k=0}^{J} \mu_{g_k} \frac{\mathbf{1}\{j \in g_k\} \mathbf{1}\{l \in g_k\}}{q_{g_k}},$$

which can be expressed in matrix form as

$$\mathbf{J}_{\ln \mathbf{S}}(\mathbf{q}) = \mathbf{M}\mathbf{V}\mathbf{M}^{\intercal},$$

where $\mathbf{V} = \operatorname{diag}\left(\frac{\mu_{g_0}}{q_{g_0}}, \ldots, \frac{\mu_{g_J}}{q_{g_J}}\right)$. This is positive definite since all μ_g are strictly positive and $\mathbf{M}$ is invertible.

Lastly, with $\mathbf{M}$ invertible, un-normalized demands solve $\ln \mathbf{S}(\mathbf{q}) = \mathbf{M}\mathbf{W} \ln(\mathbf{M}^{\intercal}\mathbf{q}) = \boldsymbol{\delta}$ and are given by

$$\mathbf{q} = (\mathbf{M}^{\intercal})^{-1} \exp\left(\mathbf{W}^{-1}\mathbf{M}^{-1}\boldsymbol{\delta}\right).$$

□

Example 4. Define nests from the symmetric incidence matrix $\mathbf{M}$ with entries $M_{ij} = \mathbf{1}_{\{i \neq j\}}$, so that each product belongs to J nests. The inverse of the incidence matrix has entries ij equal to $\frac{1}{J} - 1_{\{i=j\}}$.

Let $\mu_g = 1/J$ for each nest $g = 1, \ldots, J$. Then the un-normalized demands are given by $\tilde{\mathbf{q}} = (\mathbf{M})^{-1} \exp[J\mathbf{M}^{-1}\boldsymbol{\delta}]$ which leads to the following demands

$$q_i = \frac{\tilde{q}_i}{\sum_{j=0}^{J} \tilde{q}_j} = \frac{\sum_{j=0}^{J} e^{-J\delta_j} - Je^{-J\delta_i}}{\sum_{j=0}^{J} e^{-J\delta_j}}. \tag{30}$$

These demands are non-negative only for values of $\boldsymbol{\delta}$ within some set. To ensure positive demands, it is possible to average with the simple logit generator, since then Assumption 1 is satisfied and Theorem 1 applies.

Demands (30) are not consistent with any ARUM since they do not exhibit the (restrictive) feature of the ARUM that the mixed partial derivatives of q_j alternate in sign

(McFadden, 1981). Indeed, alternatives are substitutes

$$\frac{\partial q_1}{\partial \delta_2} = -J^2 e^{-J(\delta_1+\delta_2)} / \left(\sum_{j=0}^{J} e^{-J\delta_j}\right)^2 < 0,$$

but

$$\frac{\partial^2 q_1}{\partial \delta_2 \partial \delta_3} = -2J^3 e^{-J(\delta_1+\delta_2+\delta_3)} / \left(\sum_{j=0}^{J} e^{-J\delta_j}\right)^3 < 0.$$

The following proposition shows how a generator can be transformed into a new generator by application of a location shift and a bistochastic matrix (i.e., a matrix with non-negative elements that sum to 1 across rows and columns).

Proposition 7 (Transformation)**.** Let $\mathbf{T}$ be a generator and $\mathbf{m} \in \mathbb{R}^{J+1}$ be a location shift vector. Let $\mathbf{A} \in \mathbb{R}^{(J+1)\times(J+1)}$ be an invertible bistochastic matrix, so that $a_{ij} \geq 0$ and $\sum_{i=0}^{J} a_{ij} = \sum_{j=0}^{J} a_{ij} = 1$. Then $\mathbf{S}$ given by

$$\mathbf{S}(\mathbf{q}) = \exp\left(\mathbf{A}^T \left[\ln\left(\mathbf{T}\left(\mathbf{A}\mathbf{q}\right)\right)\right] + \mathbf{m}\right) \tag{31}$$

is a generator, and the un-normalized demands are given by

$$\tilde{\mathbf{q}} = \mathbf{A}^{-1}\mathbf{T}^{-1}\left(\exp\left[\left(\mathbf{A}^{\intercal}\right)^{-1}\left(\boldsymbol{\delta} - \mathbf{m}\right)\right]\right).$$

Proof of Proposition 7. $\mathbf{S}$ defined by (31) is twice continuously differentiable. It is also linearly homogeneous since for $\lambda > 0$,

$$\begin{aligned}
\mathbf{S}(\lambda\mathbf{q}) &= \exp\left(\mathbf{A}^{\intercal} \ln \mathbf{T}\left(\mathbf{A}\left(\lambda\mathbf{q}\right)\right) + \mathbf{m}\right), \\
&= \exp\left(\mathbf{A}^{\intercal} \ln \lambda + \mathbf{A}^{\intercal} \ln T\left(\mathbf{A}\mathbf{q}\right) + \mathbf{m}\right), \\
&= \exp\left(\ln \lambda + \mathbf{A}^{\intercal} \ln T\left(\mathbf{A}\mathbf{q}\right) + \mathbf{m}\right) = \lambda\mathbf{S}(\mathbf{q}),
\end{aligned}$$

where the second equality comes from the linear homogeneity of $\mathbf{T}$, and the third equality comes from the fact that columns of $\mathbf{A}$ sum to 1.

By Lemma 5, the Jacobian of $\ln \mathbf{S}$ is $\mathbf{J}_{\ln S} = \mathbf{A}^{\intercal}\mathbf{J}_{\ln \mathbf{T}}\mathbf{A}$, which is symmetric and positive definite.

The final conclusion follows from solving $\ln \mathbf{S}(\tilde{\mathbf{q}}) = \boldsymbol{\delta}$. □

By Proposition 7, if $\mathbf{A}$ is an invertible bistochastic matrix and Ω a GE, then $\mathbf{q} \rightarrow$

$\Omega(\mathbf{Aq})$ is also a GE, since $\Omega(\mathbf{Aq}) = -\mathbf{q}^\intercal \mathbf{A}^\intercal \ln \mathbf{S}(\mathbf{Aq})$. This construction may be useful if choice alternatives can be viewed as mixtures of another level of choice alternatives. In addition, similarly to Corollary 1, Proposition 7 allows the construction of models with analytic formulae for both their demand functions and their inverse. Lastly, it allows for zero demands: this may arise when the generator $\mathbf{T}$ does not satisfy Assumption 1. We illustrate Proposition 7 with a generator that leads to demands where products may be complements.

Example 5. Let $J + 1 = 3$, $\mathbf{m} = \mathbf{0}$, and $\mathbf{T}(\mathbf{q}) = \mathbf{q}$, and

$$\mathbf{A} = \begin{pmatrix} p & 1-p & 0 \\ 1-p & p & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with $p < 0.5$. Then we obtain

$$\tilde{\mathbf{q}} = \mathbf{A}^{-1}\left(\exp\left[\left(\mathbf{A}^T\right)^{-1}\boldsymbol{\delta}\right]\right) = \begin{pmatrix} \frac{p}{2p-1} e^{\frac{p}{2p-1}\delta_1 - \frac{1-p}{2p-1}\delta_2} - \frac{1-p}{2p-1} e^{\frac{p}{2p-1}\delta_2 - \frac{1-p}{2p-1}\delta_1} \\ \frac{p}{2p-1} e^{\frac{p}{2p-1}\delta_2 - \frac{1-p}{2p-1}\delta_1} - \frac{1-p}{2p-1} e^{\frac{p}{2p-1}\delta_1 - \frac{1-p}{2p-1}\delta_2} \\ e^{\delta_3} \end{pmatrix},$$

so that

$$q_3 = \frac{e^{\delta_3}}{e^{\frac{p}{2p-1}\delta_1 - \frac{1-p}{2p-1}\delta_2} + e^{\frac{p}{2p-1}\delta_2 - \frac{1-p}{2p-1}\delta_1} + e^{\delta_3}},$$

and $\frac{\partial q_3}{\partial \delta_1} > 0$ if and only if $\delta_2 - \delta_1 > (2p-1)\ln\left(\frac{1-p}{p}\right)$.

E Appendix for Section 5

E.1 Properties of the GNE model

Proof of Proposition 5. Using the relation (5) between $\boldsymbol{\delta}$ and $\mathbf{q}$, we get, for any pair of products j and k,

$$\frac{q_j(\boldsymbol{\delta})}{q_k(\boldsymbol{\delta})} = \exp\left(\frac{\delta_j - \delta_k}{\mu_0} + \sum_{c=1}^{C} \frac{\mu_c}{\mu_0} \ln\left(\frac{q_{\sigma_c(k)}(\boldsymbol{\delta})}{q_{\sigma_c(j)}(\boldsymbol{\delta})}\right)\right). \tag{32}$$

Then, for products j and k of the same type (i.e., with $\sigma_c(k) = \sigma_c(j)$ for all c), Expression (32) reduces to $\frac{q_j}{q_k} = \exp\left(\frac{\delta_j - \delta_k}{\mu_0}\right)$, and, in turn, the ratio q_j/q_k is independent of the characteristics or existence of all other products, i.e., IIA holds for products of the same type. However, for any two products of different types, this ratio can depend on the characteristics of other products, so that IIA does not hold in general for products of different types. □

E.2 Numerical properties of the GNE model

Using Proposition 2, the matrix of own- and cross-price derivatives for the GNE model is given by

$$\mathbf{J_q} = \text{-}\alpha\mathbf{\Psi}\left(\mathbf{\Theta};\boldsymbol{\mu}\right)\text{diag}\left(\mathbf{q}\right)\left[\mathbf{I} - \mathbf{1}_{J+1}\mathbf{q}^{\intercal}\right],$$

where

$$\mathbf{\Psi}\left(\mathbf{\Theta};\boldsymbol{\mu}\right) = \left[\mu_0\mathbf{I}_{J+1} + \sum_{c=1}^{C}\mu_c\mathbf{\Theta}_c\mathbf{Q}_{\sigma_c}\right]^{-1},$$

with $\mathbf{\Theta}_c$ given by (19) and $\mathbf{Q}_{\sigma_c}$ being the diagonal matrix of the market shares of products within their nest $\sigma_c(j)$, i.e., $\left(\mathbf{Q}_{\sigma_c}\right)_{jj} = \frac{q_j}{q_{\sigma_c(j)}}$. This means that we cannot obtain an analytic formula for each entry of the matrix of own- and cross-price derivatives independently. We therefore perform simulations to better understand substitution and complementarity patterns the GNE model can accommodate.

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 $\boldsymbol{\mu} = (\mu_0, \ldots, \mu_C)$, and NS different vectors of market shares $\mathbf{q} = (q_0, \ldots, q_J)$. We set $NS = 20$, $C = 3$, $M = 3$, and $J = 30$, and we end up with $8,000$ market structures by combining these dimensions. We obtain (i) a nesting structure by simulating a $NS \times C$ matrix of binomial random numbers; (ii) a vector of nesting parameters by simulating a $(C+1)$ vector of uniformly distributed random numbers where the first element is μ_0, then by normalizing the vector of the other nesting parameters to get a unit vector $\boldsymbol{\mu}$; (iii) a vector of market shares by simulating a $(J+1)$ vector of uniformly distributed random numbers where the first element is q_0, then by normalizing the vector of market shares of inside products to get a unit vector $\mathbf{q}$ (the normalization is to simulate markets with very low and very high values for μ_0 and q_0).

The following table gives summary statistics on the simulated data:

TABLE 2: SUMMARY STATISTICS ON THE SIMULATED DATA

Variable	Mean	Min	Max
q_0	0.5253	0.0064	0.9906
q_j	0.0158	3e-06	0.0697
μ_0	0.4662	0.0697	0.9532
μ_1	0.2014	0.0135	0.8480
μ_2	0.1420	0.0175	0.4036
μ_3	0.1904	0.0059	0.5212

Nesting structure. Table 3 shows the distribution of the own- and cross-price derivatives for the simulated data according to the proximity of the products in the characteristics space used to form product types.

Table 3: DISTRIBUTION OF PRICE DERIVATIVES ACCORDING TO THE NUMBER OF COMMON NESTS

Same nests	$\nabla_{\mathbf{q}} > 0$	Median	Min	Max	Freq.
Own-price derivatives					
–	0.00%	-0.0222	-0.7781	-3e-06	100.00%
Cross-price derivatives					
0 (None)	45.33%	-7e-07	-0.1539	0.0251	25.09%
1	90.38%	0.0002	-0.1114	0.2082	43.59%
2	100.00%	0.0006	-1e-09	0.2641	26.47%
3 (All)	100.00%	0.0009	-1e-09	0.3100	4.85%
Total	82.09%	0.0002	-0.1539	0.3100	100.00%

Notes: Column "$\nabla_{\mathbf{q}} > 0$" gives the percentage of positive cross-price elasticities (i.e., the percentage of substitutable products). Column "Freq." gives the frequencies (in percentage) of the cross-price elasticities (e.g., 4.85 percent of the cross-price elasticities involve products of the same type).

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, products are less likely to be substitutable. Products that are very similar (i.e., 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 of different types, while products of the same type are always substitutable.

Size of the outside option (q_0). Table 4 shows the percentage of positive cross-price derivatives (substitutes) according to the size of the outside option.

Table 4: PERCENTAGE OF SUBSTITUTES ACCORDING TO THE SIZE OF THE OUTSIDE OPTION

q_0	$\nabla_{\mathbf{q}} > 0$	q_0	$\nabla_{\mathbf{q}} > 0$
0.64%	87.07%	53.42%	84.06%
4.80%	87.06%	63.39%	83.66%
8.99%	86.65%	77.40%	81.82%
13.37%	86.65%	78.09%	81.83%
14.22%	86.39%	79.78%	81.39%
16.65%	86.44%	88.89%	78.30%
17.97%	86.16%	94.35%	76.42%
47.75%	84.31%	96.94%	73.05%
48.13%	84.20%	98.33%	70.94%
48.33%	84.50%	99.06%	70.90%

One concern is that a large outside option could generate complementarity. Table 4 shows that the size of the outside option does not matter for the purpose considered here: with the simulated data, there are 87 (resp. 84 and 76) percent of substitutable products for $q_0 < 1$ (resp. $q_0 \simeq 53$ and $q_0 \simeq 94$) percent. However, at the extremes, a higher value value of the outside option is associated with a higher proportion of complementary products.

Nesting parameters. Table E.2 shows the distribution of cross-price derivative according to the level of the closeness of products, as measured by the sum of nesting parameters $\mu_{jk} = \sum_{c=1}^{3} \mu_i \mathbf{1}\left\{j \in \sigma_c\left(k\right)\right\}$ for two products j and k.

Table 5: PERCENTAGE OF SUBSTITUTES ACCORDING TO THE VALUE OF μ_{jk}

μ_{jk}	$\nabla_q > 0$	Median	Min	Max
$[0, 0.1[$	65.60%	0.0000	-0.1539	0.0286
$[0.1, 0.2[$	96.37%	0.0002	-0.0538	0.1462
$[0.2, 0.3[$	93.52%	0.0003	-0.1114	0.1670
$[0.3, 0.4[$	94.16%	0.0007	-0.0673	0.2082
$[0.4, 0.5[$	93.89%	0.0009	-0.0432	0.2049
$[0.5, 0.6[$	100.00%	0.0020	1e-08	0.2295
$[0.6, 0.7[$	100.00%	0.0026	3e-08	0.2339
$[0.7, 0.8[$	100.00%	0.0032	3e-08	0.2641
$[0.8, 0.9[$	100.00%	0.0041	6e-08	0.1615
$[0.9, 1[$	100.00%	0.0130	2e-07	0.3100

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.

E.3 Data

Data. We use data from the Dominick's Database made available by the James M. Kilts Center, University of Chicago Booth School of Business. 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 weekly stores traffic.

We supplement the data with the nutrient content of the cereals using the USDA Nutrient Database for Standard Reference. This dataset is made available by the United States Department of Agriculture and provides the nutrient content of more than 8,500 different foods including ready-to-eat cereals (in particular, we use releases SR11 and SR16 for sugar). We use six characteristics: fiber, sugar, lipid, protein in grams/100g of cereals, energy in Kcal/100g of cereals, and sodium in mg/100 g of cereals. We convert each characteristics into g/serve, Kcal/serve, and mg/serve, respectively.

We supplement the data with the sugar monthly price in dollars/kg. We use this variable to form a cost-based instrument: the price of the cereal's sugar content (i.e., sugar content in g times the sugar monthly price in dollars/g).

Market shares and prices. 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 (i) an individual in a household consumes around 15 servings per month, and (ii) consumers visit stores twice a week.[29] Indeed, according to USDA's Economic Research Service, per capita consumption of RTE cereals was equal to around 14 pounds (that is, about 6350 grammes) in 1992, which is equivalent to serving 15 servings per month (without loss of generality, we assume that a serving weight is equal to 35 grammes). Then, the potential month-store market size (in servings) is computed as the weekly average number of households which visited that store in that given month, times the average household size for that store, divided by two, times the number of servings an individual consumes in a month. Using the weekly average number of households itself allows to take into account the fact that consumers visit stores once a week. 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. The sample we use consists of the six biggest companies mentioned above. Brand names seem to play a non-negligible role: Kellogg's is the biggest company with large market shares in all segments; and General Mills, the second biggest one, is especially present in the family and kids segments. Taken together they account for around 80 percent of the market. As regards market segments, the family and kids segments dominate and account for almost 70 percent of the market.

Table 7 shows the nutrient content of the cereals according to their market segment and brand name. 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 more calories. Nabisco offers cereals with less sugar and less calories, and Quaker and Ralston offer cereals with more calories.

Implementing the GNE model. We must first select the dimensions along which the market is segmented. Then, we estimate the GNE model by 2SLS (or GMM) using cost

[29] As a robustness check, we have also estimated the models with the alternative assumption that consumers visit stores once a week. Results do not change significantly.

shifters and BLP instruments as instruments for prices and nesting terms. Lastly, we get the estimated net utility $\hat{\boldsymbol{\delta}}$, the estimated marginal utility of income $\hat{\alpha}$, and the estimated nesting parameters $\hat{\boldsymbol{\mu}} = (\hat{\mu}_1, \ldots, \hat{\mu}_C)$, which we use to compute the matrix of own- and cross-price elasticities in two steps. The first step computes the predicted market shares $\hat{\mathbf{q}}$ by solving for $\mathbf{q}$ the system of nonlinear equations $\ln \mathbf{S}(\mathbf{q}) = \boldsymbol{\delta} + \mathbf{c}$, with $\mathbf{S}$ defined in Equation (20) and with $\mathbf{c} = -\ln(q_0) = -\ln\left(1 - \sum_{j=1}^{J} q_j\right)$, by normalization $\delta_0 = 0$. To do so, we use the Stata command `solvenl` or the Matlab command `fsolve`. The second step computes the matrix of elasticities $\boldsymbol{\eta}$ at $\hat{\mathbf{q}}$ and $\hat{\boldsymbol{\delta}}$ using (8) and $\boldsymbol{\eta} = \mathrm{diag}(\boldsymbol{\delta})\, \mathbf{J}_{\mathbf{q}} \mathrm{diag}(\mathbf{q})^{-1}$, where $\mathbf{J}_{\mathbf{q}}$ is obtained using Proposition 2.

E.4 Results: Elasticities for the main specifications

Tables 9 and 10 give the estimated average own- and cross-price elasticities of demands for the main specifications, averaged over markets and product types. Product types are defined in Table 6.

Table 6: PRODUCT TYPES

Number	Product type
1	General Mills – Family
2	General Mills – Health/nutrition
3	General Mills – Kids
4	General Mills – Taste enhanced
5	Kelloggs – Family
6	Kelloggs – Health/nutrition
7	Kelloggs – Kids
8	Kelloggs – Taste enhanced
9	Nabisco – Health/nutrition
10	Post – Health/nutrition
11	Post – Kids
12	Post – Taste enhanced
13	Quaker – Family
14	Quaker – Kids
15	Quaker – Taste enhanced
16	Ralston – Family
17	Ralston – Kids

Table 7: SAMPLE STATISTICS BY SEGMENT AND BY BRAND NAME

Dimensions	Sugar g/serve	Energy Kcal/serve	Fiber g/serve	Lipid g/serve	Sodium mg/serve	Protein g/serve	N
Segment							
Family	7.54	130.41	2.22	0.99	269.66	2.88	17
	(5.27)	(9.83)	(2.61)	(0.71)	(88.64)	(1.03)	
Health/nutrition	5.03	122.54	3.16	0.54	168.54	3.84	9
	(3.69)	(5.78)	(1.31)	(0.21)	(133.62)	(1.35)	
Kids	13.40	137.75	1.00	1.35	211.38	2.01	16
	(4.17)	(3.80)	(0.69)	(0.79)	(44.77)	(0.87)	
Taste enhanced	9.70	129.28	3.32	2.22	166.43	3.16	8
	(2.05)	(15.50)	(1.12)	(1.93)	(76.38)	(0.34)	
Brand Name							
General Mills	9.92	132.09	1.99	1.51	230.69	2.65	17
	(4.67)	(7.69)	(0.98)	(0.82)	(60.83)	(0.83)	
Kellogg's	9.58	127.50	2.47	0.85	228.49	2.88	18
	(5.52)	(11.16)	(2.81)	(0.96)	(103.93)	(1.43)	
Nabisco	0.25	125.48	3.43	0.58	2.10	3.83	2
	(0.09)	(0.74)	(0)	(0)	(1.98)	(0.02)	
Post	12.02	130.76	2.09	1.03	212.03	2.49	5
	(4.64)	(14.83)	(2.02)	(0.78)	(22.31)	(1.15)	
Quaker	8.50	139.44	2.26	2.43	159.88	3.59	5
	(4.04)	(9.20)	(0.66)	(1.86)	(94.60)	(1.15)	
Ralston	7.09	138.48	0.58	0.51	305.43	2.04	3
	(6.61)	(1.41)	(0.08)	(0.65)	(71.57)	(0.39)	
Total	9.31	131.16	2.17	1.22	216.29	2.82	50
	(5.21)	(10.21)	(1.92)	(1.08)	(93.53)	(1.15)	

Notes: Standard deviations are reported in parentheses.

Table 8: TOP 50 BRANDS

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

Table 9: AVERAGE PRICE ELASTICITIES FOR THE GNE MODEL

	Own	Cross																
Type		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	-3.165	0.241	0.111	0.085	0.092	0.092	-0.038	-0.064	-0.056	0.003	0.013	-0.013	-0.005	0.133	-0.023	-0.016	0.077	-0.079
2	-3.146	0.072	0.379	0.069	0.074	-0.039	0.269	-0.041	-0.037	0.113	0.238	-0.072	-0.064	-0.003	-0.006	-0.002	-0.004	-0.006
3	-3.427	0.066	0.083	0.316	0.076	-0.026	-0.009	0.225	-0.016	0.002	-0.059	0.174	-0.064	-0.100	0.150	-0.090	-0.044	0.206
4	-2.950	0.067	0.083	0.071	0.384	-0.034	-0.018	-0.030	0.282	0.003	-0.035	-0.047	0.255	-0.051	-0.047	0.266	-0.004	-0.000
5	-2.560	0.097	-0.063	-0.035	-0.049	0.211	0.050	0.078	0.065	-0.007	-0.017	0.011	-0.003	0.125	-0.008	-0.021	0.074	-0.058
6	-2.700	-0.028	0.308	-0.008	-0.018	0.035	0.372	0.056	0.046	0.122	0.251	-0.066	-0.072	-0.015	0.005	-0.005	-0.009	0.011
7	-3.366	-0.037	-0.037	0.169	-0.024	0.044	0.044	0.250	0.057	-0.005	-0.062	0.144	-0.046	-0.084	0.122	-0.071	-0.036	0.170
8	-2.531	-0.042	-0.042	-0.016	0.288	0.046	0.046	0.072	0.376	-0.005	-0.056	-0.029	0.265	-0.063	-0.036	0.268	-0.009	0.017
9	-1.949	0.001	0.074	0.001	0.002	-0.003	0.069	-0.003	-0.003	0.909	0.062	-0.011	-0.009	0.005	0.005	0.006	0.005	0.005
10	-1.524	0.008	0.225	-0.047	-0.028	-0.010	0.208	-0.065	-0.045	0.090	—	0.399	0.396	0.033	-0.022	-0.002	0.016	-0.038
11	-3.175	-0.006	-0.046	0.092	-0.026	0.004	-0.036	0.102	-0.016	-0.011	0.266	0.404	0.272	-0.029	0.069	-0.050	-0.010	0.087
12	-1.949	-0.003	-0.050	-0.042	0.184	-0.001	-0.048	-0.040	0.186	-0.011	0.337	0.344	—	-0.005	-0.044	0.182	0.013	-0.026
13	-2.337	0.050	-0.002	-0.048	-0.027	0.045	-0.008	-0.053	-0.032	0.005	0.020	-0.025	-0.004	—	0.233	0.254	0.040	-0.058
14	-2.273	-0.011	-0.004	0.090	-0.030	-0.003	0.003	0.098	-0.022	0.005	-0.016	0.079	-0.040	0.285	0.387	0.266	-0.013	0.089
15	-1.778	-0.008	-0.001	-0.056	0.177	-0.010	-0.003	-0.059	0.174	0.006	-0.003	-0.058	0.170	0.326	0.277	—	0.012	-0.037
16	-2.607	0.028	-0.002	-0.019	-0.002	0.026	-0.005	-0.021	-0.004	0.005	0.009	-0.008	0.009	0.038	-0.009	0.008	0.808	0.759
17	-3.382	-0.021	-0.003	0.072	-0.000	-0.014	0.004	0.078	0.006	0.003	-0.017	0.057	-0.014	-0.041	0.051	-0.021	0.581	—

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

Table 10: AVERAGE PRICE ELASTICITIES FOR THE THREE-LEVEL NL MODELS

Type	3NL1				3NL2			
	Own	Cross			Own	Cross		
		Same subgroup	Same group	Different group		Same subgroup	Same group	Different group
1	-3.528	0.182	0.137	0.005	-3.524	0.208	0.147	0.007
2	-3.414	0.426	0.257	0.004	-3.530	0.332	0.104	0.005
3	-3.778	0.300	0.228	0.004	-3.876	0.226	0.129	0.006
4	-3.228	0.403	0.302	0.004	-3.398	0.255	0.117	0.006
5	-2.840	0.178	0.145	0.006	-2.868	0.169	0.124	0.008
6	-2.994	0.353	0.282	0.004	-3.181	0.186	0.089	0.006
7	-3.678	0.261	0.172	0.003	-3.763	0.199	0.080	0.005
8	-2.781	0.386	0.296	0.004	-2.971	0.215	0.092	0.006
9	-2.804	0.309	0.169	—	-2.030	1.102	—	0.003
10	-1.930	—	0.244	0.003	-1.605	—	0.507	0.005
11	-3.671	0.229	0.111	0.002	-3.435	0.488	0.329	0.003
12	-2.295	—	0.201	0.003	-2.028	—	0.396	0.004
13	-2.584	—	0.059	0.002	-2.270	—	0.298	0.003
14	-2.685	0.213	0.129	0.002	-2.480	0.434	0.321	0.004
15	-2.064	—	0.203	0.003	-1.845	—	0.357	0.004
16	-3.494	0.226	0.058	0.002	-2.713	1.029	0.800	0.003
17	-3.929	—	0.089	0.002	-3.236	—	0.667	0.002

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