Functional Form and Shape Restrictions in Discrete Choice Models*

Julien Monardo

Discussion Paper 25/813

9th April 2025

School of Economics

University of Bristol
Priory Road Complex
Bristol
BS8 1TU
United Kingdom

University of Bristol
School of Economics

Functional Form and Shape Restrictions in Discrete Choice Models*

Julien Monardo†

April 9, 2025

Abstract

Discrete choice demand models are commonly used to answer various economic questions. This paper develops a representation theorem that establishes the necessary and sufficient functional form and shape restrictions characterizing a large family of discrete choice demand models extending beyond the traditional additive random utility framework. The representation theorem yields three significant empirical implications. First, it provides economic intuition for (parameter) restrictions commonly imposed on some popular discrete choice models. Second, it offers a specification tool for building demand models that satisfy mild and easily verifiable properties while being consistent with utility maximization and accommodating rich substitution patterns, including complementarity in demand. Third, it provides an efficient numerical algorithm for demand inversion, a crucial step in the demand estimation procedure.

KEYWORDS. Choice Probability; Demand Inversion; Discrete Choice Models; Functional Form Restrictions; Representation Theorem; Shape Restrictions.

JEL CLASSIFICATION. D11, D12, L00.

*I am grateful to Steven Berry, Debopam Bhattacharya, Xavier D'Haultfoeuille, André de Palma, Federico Echenique, Alfred Galichon, Laura Grigolon, Alessandro Iaria, Grégory Jolivet, Hugo Molina, Lars Neishem, David Pacini, and Pietro Spini for helpful comments and suggestions.

†School of Economics, University of Bristol; julien.monardo@bristol.ac.uk. The Priory Road Complex, Priory Road, Clifton BS8 1TU, United Kingdom

1 Introduction

Discrete choice demand models are commonly used to answer various economic questions.[1] As is well known, the answers to these questions crucially depend on the functional form and shape restrictions imposed on the demand model. However, when derived from an additive random utility model (ARUM), discrete choice demand models restrict substitution patterns due to technical constraints. Notably, these models rule out complementarity in demand by assumption.[2] This limitation prevents the application of these models to increasingly prevalent multi-channel sales markets where complementarity may occur, particularly between online and offline sales channels, such as in the hotel market between online travel agencies and direct booking channels.

Motivated by these observations, this paper derives the necessary and sufficient functional form and shape restrictions characterizing a large family of discrete choice demand models that satisfy mild regularity properties. This results in a representation theorem that provides a unifying framework for discrete choice modeling extending beyond the ARUM framework. This representation theorem yields three significant empirical implications. First, it elucidates the economic intuition underlying (parameter) restrictions in many widely-used discrete choice demand models. Second, it offers a general specification tool for building demand models that accommodate rich substitution patterns, including complementarity in demand. This specification tool is particularly valuable for analyzing multi-channel markets such as the hotel market, where complementarity between online travel agencies and direct booking channels may occur.[3] Third, it provides an efficient numerical algorithm for demand inversion.

To obtain the representation theorem, I consider a consumer facing a choice set of differentiated products whose choice is described by a choice probability function satisfying mild regularity properties. The representation theorem states that any such choice probabil-

[1]For examples, see Table 1 in Berry and Haile (2016).

[2]In the ARUM, the consumer chooses the product that provides her with the highest indirect utility given by the sum of deterministic and random utility terms. Discrete choice demand models derived from such an ARUM rule out complementarity in demand, which arises when increasing the desirability of a product increases the probability of a competing product, due to a technical condition that ensures the vector of random utility terms has a valid cumulative distribution function. See Section 2.1 for details and Anderson, de Palma, and Thisse (1992) for a comprehensive treatment of discrete choice models. This demand-based complementary is different from Gentzkow (2007)'s utility-based complementarity, which occurs when the utility of consuming products together exceeds the sum of their standalone utilities.

[3]The French (Decision 15-D-06 of April 21, 2015) and German (Decision B9-66/10) competition authorities mention this possibility.

ity function exhibits the same functional form as McFadden (1978)'s generalized extreme value (GEV) choice probability function, but with less restrictive shape restrictions, which translates into richer substitution patterns and an expanded parameter space in parametric cases. This result leads to the first two empirical implications. First, the proof of the representation theorem helps understand the economic intuition behind the restrictions satisfied by the GEV models. Indeed, the discrete choice literature has recognized the good properties of the GEV models while also deploring their mathematical complexity (Daly and Bierlaire, 2006) and lack of economic intuition (Train, 2009). By deriving rather than merely stating the functional form and shape restrictions, the representation theorem provides economic intuition to the GEV restrictions. It also involves shape restrictions that are more straightforward to verify, thereby addressing the mathematical complexity of the GEV models while retaining their good properties.

Second, the representation theorem provides a general specification tool for discrete choice demand models that extends beyond the ARUM framework while maintaining consistency with utility maximization. This parallels McFadden (1978) who introduces the GEV family as a specification tool for building ARUM-consistent discrete choice demand models. However, this departs from McFadden (1978) who assumes, rather than derives, the GEV functional form and shape restrictions. To illustrate the empirical usefulness of the representation theorem as a specification tool, I apply Bresnahan et al. (1997)'s product differentiation logit (PDL) model to study the demand relationship between online travel agencies and direct booking channels in the hotel market, segmented by hotel brand and sales channel. Importantly, the PDL model can capture complementarity between these channels when the corresponding nesting parameter exceeds one, a condition that violates the GEV restrictions but satisfies the restrictions established in the representation theorem.

The third and final empirical implication deals with demand inversion, a crucial step in the demand estimation procedure based on individual-level and/or market-level data which involves numerically inverting the choice probability (market share) function. The numerical inversion can be numerically challenging, which has led the literature to propose different algorithms.[4] I show that all demand models encompassed by the representation theorem are invertible. Furthermore, I show that the proof of invertibility provides a numerical algorithm for demand inversion that only requires solving an unconstrained convex optimization problem, alleviating the numerical challenge.[5]

[4]See Conlon and Gortmaker (2020) or Pál and Sándor (2023) for a review of these algorithms.

[5]Any convex problem is guaranteed to converge from any starting values to the unique optimum and

Lastly, I extend the framework to incorporate unobserved heterogeneity in consumer preferences through random coefficients. This extension encompasses all random coefficient discrete choice models that add a nesting structure (i.e., an allocation of products into nests or groups) to the random coefficient logit model. As highlighted by Grigolon and Verboven (2014), this approach provides a tractable method for capturing sources of market segmentation that continuous product characteristics cannot adequately represent. Importantly, this extension enables to develop discrete choice demand models that simultaneously accommodate unobserved heterogeneity in preferences and complementarity in demand. I further demonstrate that these models are all invertible and can be numerically inverted by solving an unconstrained convex optimization problem. Lastly, using simulations of the random coefficient logit model performed by Pál and Sándor (2023), I find that the proposed numerical algorithm delivers superior performance compared to the widely-used SQUAREM algorithm developed by Reynaerts, Varadha, and Nash (2012).

Related literature. This paper connects to several strands of literature. First, it builds upon the theoretical literature on discrete choice demand models that analyze the properties required for a choice probability function to be ARUM consistent (Daly and Zachary, 1979; Börsch-Supan, 1990; Koning and Ridder, 2003; Jaffe and Kominers, 2012; Armstrong and Vickers, 2015). I extend the analysis by investigating properties of choice probability function that satisfy mild regularity conditions extending beyond the ARUM framework while maintaining consistency with utility maximization. This theoretical investigation yields three empirical implications: a specification tool for discrete choice modeling that allows for complementarity in demand, economic intuition for commonly assumed functional form and shape restrictions, and an efficient numerical algorithm for demand inversion. Relatedly, this paper relates to the literature on the integrability problem, which characterizes the set of demand functions derivable from utility maximization (Hurwicz and Uzawa, 1971; Jackson, 1986; Nocke and Schutz, 2017; Fally, 2022; Bhattacharya, 2025). I establish that the representation theorem characterizes the set of discrete choice demand functions derived from Allen and Rehbeck (2019b)'s utility model of heterogeneous, utility-maximizing consumers satisfying mild regularity assumptions.

Second, this paper relates to the discrete choice literature focusing on the GEV family. While the logit and nested logit models are the most commonly employed GEV models,

benefits from efficient, robust algorithms available in many software packages and supported by well-defined mathematical properties.

they have been criticized for generating restrictive substitution patterns. This has led the literature to use more complex GEV models, including the ordered logit (Small, 1987), the product differentiation logit (Bresnahan, Stern, and Trajtenberg, 1997), the paired combinatorial logit (Koppelman and Wen, 2000), the generalized nested logit (Wen and Koppelman, 2001), the cross nested logit (see, e.g., Bierlaire, 2006), the flexible coefficient multinomial logit (Davis and Schiraldi, 2014), and the ordered nested logit (Grigolon, 2021). The representation theorem derived in this paper demonstrates the assumed GEV functional form and certain shape restrictions emerge directly from mild regularity assumptions imposed on discrete choice demand models. This finding both validates the assumed restrictions and provides economic intuition behind them. Furthermore, it allows to build demand models that share the GEV functional form but with weaker shape restrictions, effectively expanding the demand parameter space in parametric cases.

Third, this paper connects to papers building discrete choice demand models that allow for complementarity in demand. One stream of literature follows Gentzkow (2007) by incorporating complementarity in utility through discrete choice models for baskets of products (see e.g., Ershov et al., 2023; Iaria and Wang, 2025). These papers accommodate complementarity in demand through positive demand synergies, which differs from complementarity in demand. Another stream of research uses discrete choice demand models with nesting structure to generate complementarity (Hortaçsu et al., 2024; Fosgerau et al., 2024; Iaria and Wang, 2025). Fosgerau et al. (2024) and Hortaçsu et al. (2024) allow for complementarity by specifying an inverse market share model consistent with Allen and Rehbeck (2019b)'s utility model, while Iaria and Wang (2025) construct a nesting-based model for baskets of products building on Gentzkow (2007).

Organization of the paper. The remainder of the paper is organized as follows. Section 2 provides established facts about discrete choice models. Section 3 provides the main results of this paper, and, in particular, the representation theorem and its empirical implications. Section 4 extends the framework to incorporate unobserved heterogeneity in consumer preferences through random coefficients. Section 5 concludes.

2 Some Facts about Discrete Choice Models

Consider a consumer facing a choice set $\mathcal{J} = \{0, 1, \dots, J\}$ of $J + 1$ differentiated products. Let $\delta_j \in \mathbb{R}$ be the quality index for product $j \in \mathcal{J}$, which can be seen as a price-adjusted quality index.[6] Let $\boldsymbol{\delta} = (\delta_0, \dots, \delta_J)^\intercal$ be the vector of quality indices. The consumer's choice is described by a choice probability (or demand) function $\boldsymbol{\sigma}(\boldsymbol{\delta}) = (\sigma_0(\boldsymbol{\delta}), \dots, \sigma_J(\boldsymbol{\delta}))$ that maps from quality indices to choice probabilities. The consumer can be an individual or represent a population of consumers. In the latter case, the choice probability function can be interpreted as a market share function.

The focus on a single unit, discrete choice demand model (i.e., on choice probabilities or market shares) rather than on a standard continuous choice demand model (i.e., on raw quantities) can be motivated in several ways. First, the empirical literature typically relies on a single unit, discrete-choice demand model that describes a vector of choice probabilities or market shares depending on whether data are individual- or market-level. Furthermore, empirical researchers commonly use a single unit, discrete choice demand model to answer various economic questions. They also typically use data that come in the form of market shares, data that describe consumers making single unit, discrete choices, or data on quantities that they convert into shares. Lastly, an extensive theoretical literature uses single unit, discrete choice models.[7]

Moreover, the absence of income effect (i.e., that income does not affect the choice probability function) can be motivated by the fact that this assumption is usually maintained for categories of products for which expenditure is a small fraction of income (e.g., beers, breakfast cereals, etc.).

The most famous families of choice probability functions are the family of choice probability functions consistent with an additive random utility model (McFadden, 1981; Anderson et al., 1992), hereafter referred to as ARUM choice probabilities, and McFadden (1978)'s family of generalized extreme value (GEV) choice probability functions. In the remainder of this section, I review key established facts of these two families of choice probability functions.

2.1 ARUM Choice Probabilities

In the ARUM, the consumer chooses the product $j \in \mathcal{J}$ that provides her with the highest indirect utility given by

$$u_j = \delta_j + \varepsilon_j,$$

[6] In the discrete choice literature, quality indices are also referred to as mean utilities or utility indices.

[7] See, e.g., Anderson, de Palma, and Thisse (1992), Hofbauer and Sandholm (2002), Fudenberg, Iijima, and Strzalecki (2015), Allen and Rehbeck (2023), and Garrido (2024).

where δ_j is often referred to as the deterministic utility term and ε_j is a random utility term. The vector of random utility terms $\boldsymbol{\varepsilon} = (\varepsilon_0, \ldots, \varepsilon_J)$ follows a joint distribution with finite means that is absolutely continuous, fully supported on $\mathbb{R}^{J+1}$, and independent of the vector of deterministic utility terms. These assumptions are standard in the discrete choice literature. They imply that utility ties occur with probability zero, choice probabilities are all everywhere positive, and there is no random coefficient.

Daly and Zachary (1979) establish the necessary and sufficient conditions for a choice probability function to be consistent with an ARUM. They show that a choice probability function $\boldsymbol{\sigma}(\boldsymbol{\delta})$ is consistent with an ARUM if and only if it satisfies the following properties known as the Daly-Zachary conditions: for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$,[8]

(P·1) For all $j \in \mathcal{J}$, $\sigma_j(\boldsymbol{\delta}) > 0$, and $\sum_{k \in \mathcal{J}} \sigma_k(\boldsymbol{\delta}) = 1$.

(P·2) For all $j \in \mathcal{J}$ and for any fixed finite values of $\{\delta_k\}_{k \neq j}$, $\lim_{\delta_j \to +\infty} \sigma_j(\boldsymbol{\delta}) = 1$.

(P·3) For all $j \in \mathcal{J}$ and $\mathbf{c} = (c, \ldots, c)^\intercal \in \mathbb{R}^{J+1}$, $\sigma_j(\boldsymbol{\delta} + \mathbf{c}) = \sigma_j(\boldsymbol{\delta})$.

(P·4) For all $j \in \mathcal{J}$, the mixed partial derivatives $\partial^k \sigma_j(\boldsymbol{\delta}) / \partial \delta_{i_1} \partial \delta_{i_2} \ldots \delta_{i_k}$ of σ_j with respect to k distinct elements $i_1, \ldots, i_k \in \mathcal{J} \backslash \{j\}$ of the vector $\boldsymbol{\delta}$ of quality indices other than its own exist, are nonnegative if k is odd, nonpositive if k is even, and are independent of the order of differentiation.

(P·5) The substitution matrix $\left[\frac{\partial \sigma_i(\boldsymbol{\delta})}{\partial \delta_j}\right]_{0 \leq i,j \leq J}$ is symmetric.[9]

This result provides a complete characterization of the ARUM choice probabilities. As established by (P·1), probabilities must be positive and sum to one. According to (P·2), the possibility exists for any product to dominate others when it becomes infinitely attractive. Translation invariance, described in (P·3), implies that only differences in quality indices, not their absolute values, determine the choice probabilities. The technical condition (P·4) ensures that the vector of random utility terms $\boldsymbol{\varepsilon}$ has a valid cumulative distribution

[8]See also Anderson, de Palma, and Thisse (1992, Theorem 3.1, p. 69) and Koning and Ridder (2003, Section 3). In their original formula, Daly and Zachary (1979) do not require the joint distribution of the vector of random utility terms to be fully supported on $\mathbb{R}^{J+1}$ and thus allow for zero probabilities. I impose the full support assumption as it is satisfied by the vast majority of ARUM choice probabilities used in the theoretical and enpirical literature. In their original formula, Daly and Zachary (1979) further assume that the mixed partial derivatives are continuous, but this assumption has not been maintained in the subsequent reformulations of the result.

[9]That is, $\frac{\partial \sigma_i(\boldsymbol{\delta})}{\partial \delta_j} = \frac{\partial \sigma_j(\boldsymbol{\delta})}{\partial \delta_i}$, for all $i, j \in \mathcal{J}$.

function. Moreover, (P·4) implies that the substitution matrix is everywhere positive semi-definite, but the converse does not hold.[10] When quality indices are decreasing in prices, (P·4) thus implies that the choice probability functions obey the law of demand, i.e., the choice probability function of each product is nonincreasing in its own price. Hence, (P·4) and (P·5) ensure that the choice probabilities are consistent with quasi-linear utility maximization,[11] and in turn, that they can be used to perform correct welfare analysis. Lastly, (P·4) rules out complementarity in demand, which arises when increasing the desirability of a product (i.e., its quality index) increases the probability a competing product is chosen, i.e., the cross-quality index derivative of demand is positive.

2.2 GEV Choice Probabilities

McFadden (1978) introduces the family of GEV choice probability functions through a generating function $G(e^{\boldsymbol{\delta}})$ defined on $\mathbb{R}^{J+1}_{++}$ with the following properties:

(G·1) $G(e^{\boldsymbol{\delta}}) > 0$.[12]

(G·2) $G(e^{\boldsymbol{\delta}})$ is homogeneous of degree $\mu > 0$ in $e^{\boldsymbol{\delta}}$.[13]

(G·3) For all $j \in \mathcal{J}$, $G(e^{\boldsymbol{\delta}}) \to +\infty$ as $e^{\delta_j} \to +\infty$.

(G·4) The mixed partial derivatives of G with respect to k distinct elements of the vector $e^{\boldsymbol{\delta}}$ of exponentiated quality indices exist and are continuous. Moreover, they are nonnegative if k is odd and nonpositive if k is even.

If $G(e^{\boldsymbol{\delta}})$ is a generating function, then the GEV choice probabilities are given by

$$\sigma_j(\boldsymbol{\delta}) = \frac{e^{\delta_j} \dfrac{\partial G\left(e^{\boldsymbol{\delta}}\right)}{\partial e^{\delta_j}}}{\mu G\left(e^{\boldsymbol{\delta}}\right)}, \quad j \in \mathcal{J}, \tag{1}$$

and are consistent with an ARUM.

[10] See, e.g., Koning and Ridder (2003, Appendix C).

[11] See, e.g., Nocke and Schutz (2017).

[12] In his original formulation, McFadden (1978) assumes that $G(e^{\boldsymbol{\delta}}) \geq 0$. However, as highlighted by Daly and Bierlaire (2006), it is also required that $G(e^{\boldsymbol{\delta}}) \neq 0$.

[13] A function $f : \mathbb{R}^{J+1}_{++} \to \mathcal{Y}_f \subseteq \mathbb{R}$ is homogeneous of degree $\mu > 0$ if $f(\lambda \mathbf{x}) = \lambda^{\mu} f(\mathbf{x})$ for all $\lambda > 0$ and $\mathbf{x} \in \mathbb{R}^{J+1}_{++}$. In his original formulation, McFadden (1978) uses $\mu = 1$. Ben-Akiva and Francois (1983) generalize to any $\mu > 0$.

This result serves as a general specification tool: by specifying a function G satisfying (G·1) through (G·4), we can build an ARUM choice probability function, which can be used to perform welfare analysis and counterfactual analyses to answer many economic questions. As the discrete choice literature highlights, the GEV choice probabilities possess desirable properties: they have closed-form expressions, are ARUM consistent, and accommodate a wide range of consumer behavior (Daly and Bierlaire, 2006). However, they lack economic intuition (Train, 2009), are often considered mathematically complex (Daly and Bierlaire, 2006), and rule out complementarity in demand by assumption. Properties (G·1) – (G·3) are easy to verify but (G·4) is definitely burdensome to verify as it requires checking that all mixed partial derivatives of G, which are non-trivial to derive, have the appropriate sign.

3 Functional Form and Shape Restrictions

As established earlier, (P·4) represents a technical assumption that restricts products to be substitutes in demand. Below, I further show that this property imposes limitations on substitution patterns even within the space of substitutable products. Besides (P·4) implies that the substitution matrix is everywhere positive semi-definite. In the remainder of this paper, I replace (P·4) with a weaker property that imposes positive semi-definiteness and continuous differentiability. This modification is formalized in the following assumption.

Assumption 1. The choice probability function $\boldsymbol{\sigma}(\boldsymbol{\delta})$ satisfies the following properties: for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$, properties (P·1) – (P·5) hold, with (P·4) replaced by

(P·4') $\boldsymbol{\sigma}(\boldsymbol{\delta})$ is continuously differentiable, and its substitution matrix $\left[\frac{\partial \sigma_i(\boldsymbol{\delta})}{\partial \delta_j}\right]_{0 \leq i,j \leq J}$ is positive semi-definite.[14]

Two comments are in order. First, (P·4') and (P·5) together implies the existence of a

[14] A $(J+1)$-dimensional square matrix $\mathbf{A}$ is positive semi-definite if $\mathbf{v}^\intercal \mathbf{A} \mathbf{v} \geq 0$ for all $(J+1)$-dimensional vectors $\mathbf{v} \neq \mathbf{0}$, where $\mathbf{0}$ is the $(J+1)$-dimensional vector of zeros. For computing second-order demand derivatives and making statements about demand curvature, we should require the choice probability function to be twice continuously differentiable rather than merely continuously differentiable. This strengthened condition does not affect the other aspects of the result.

twice continuously convex consumer surplus function $CS(\boldsymbol{\delta})$ satisfying Roy's identity,[15]

$$\frac{\partial CS(\boldsymbol{\delta})}{\partial \delta_j} = \sigma_j(\boldsymbol{\delta}), \quad \boldsymbol{\delta} \in \mathbb{R}^{J+1}, \quad j \in \mathcal{J}. \tag{2}$$

Second, (P·4') does not rule out complementarity in demand, which represents a significant departure from the ARUM and GEV frameworks.

The primary goal of this paper is to derive the functional form and shape restrictions satisfied by any choice probability function meeting the conditions specified in Assumption 1. I develop a representation theorem that characterizes any such choice probability function through a generating function H with well-defined properties. Then, I compare this derived family of choice probability functions with existing families, with particular focus on how it extend the GEV family. Lastly, I address two practical questions: first, I employ a member of the derived family to investigate some economic insights that existing models cannot accommodate; second, I discuss identification using market-level data. The proofs and details for this section are provided in Appendix B.

3.1 Main Result

The following theorem constitutes the main result of this paper.

Theorem 1. If the choice probability function $\boldsymbol{\sigma}(\boldsymbol{\delta})$ satisfies Assumption 1, then there exist a constant $\mu > 0$ and a twice continuously differentiable function $H(e^{\boldsymbol{\delta}})$ defined over $\mathbb{R}^{J+1}_{++}$ such that

$$\sigma_j(\boldsymbol{\delta}) = \frac{e^{\delta_j} \dfrac{\partial H\left(e^{\boldsymbol{\delta}}\right)}{\partial e^{\delta_j}}}{\mu H\left(e^{\boldsymbol{\delta}}\right)}, \quad j \in \mathcal{J}. \tag{3}$$

Moreover, the function $H(e^{\boldsymbol{\delta}})$ satisfies the following properties for all $e^{\boldsymbol{\delta}} \in \mathbb{R}^{J+1}_{++}$:

(H·1) $H(e^{\boldsymbol{\delta}}) > 0$.

(H·2) $H(e^{\boldsymbol{\delta}})$ is homogeneous of degree $\mu > 0$ in $e^{\boldsymbol{\delta}}$.

(H·3) For all $j \in \mathcal{J}$ and for any fixed finite values of $\{e^{\delta_k}\}_{k \neq j}$, $\lim_{e^{\delta_j} \to +\infty} H(e^{\boldsymbol{\delta}}) = +\infty$.

(H·4) For all $j \in \mathcal{J}$, $\frac{\partial H(e^{\boldsymbol{\delta}})}{\partial \delta_j} > 0$.

[15]See Appendix B.1 for details. Roy's identity in the context of discrete choice models is typically referred to as the Williams-Daly-Zachary theorem (McFadden, 1981).

(H•5) $H(e^{\boldsymbol{\delta}})$ is log-convex in $\boldsymbol{\delta}$.[16]

Furthermore, the consumer surplus function is given by

$$CS(\boldsymbol{\delta}) = \frac{1}{\mu} \ln H\left(e^{\boldsymbol{\delta}}\right). \tag{4}$$

Conversely, if there exist a constant $\mu > 0$ and a twice continuously differentiable function $H(e^{\boldsymbol{\delta}})$ defined over $\mathbb{R}^{J+1}_{++}$ satisfying properties (H•1) – (H•5) for all $e^{\boldsymbol{\delta}} \in \mathbb{R}^{J+1}_{++}$, then the choice probability function $\boldsymbol{\sigma}(\boldsymbol{\delta})$ defined by (3) satisfies Assumption 1.

This representation theorem provides a complete characterization of choice probabilities satisfying Assumption 1. It establishes that any such choice probability function exhibits the same functional form as the GEV choice probability while requiring weaker restrictions on the generating function. As shown in Subsection 3.2, this allows the parametric models to accommodate larger parameter space than the GEV generating functions, yielding richer substitution patterns. The expression (4) for the consumer surplus function shows that H encapsulates all the information needed to study consumer behavior, i.e., to compute demand and consumer surplus, and to conduct counterfactual analysis.

The contribution of Theorem 1 is twofold. First, it offers researchers a general specification tool for specifying demand models consistent with utility maximization (via Roy's identity) while relaxing some of the ARUM restrictions. Second, as demonstrated in the next sections, it provides a unifying framework that encompasses not only existing discrete choice models but also models that accommodate richer substitution patterns, including various forms of complementarity in demand.

The proof of Theorem 1 is interesting in itself and can be sketched as follows. I first employ Roy's identity to translate properties (P•1) – (P•5) of the choice probability function into restrictions on the consumer surplus function. Then, I exploit the structure imposed by the single unit assumption to demonstrate that the consumer surplus function satisfies homotheticity-like restrictions. Then, applying Euler's theorem for homothetic functions (McElroy, 1969), I derive an ordinary differential equation whose solution yields the logarithmic functional form of the consumer surplus (4), along with its shape restrictions. The functional form and shape restrictions of the choice probability function then follow directly from Roy's identity. This proof strategy appears to be the first application of homotheticity-like restrictions in discrete choice models and represents a novel use of Euler's theorem for homothetic functions in consumer theory, rather than producer theory.

[16]A function $f : \mathbb{R}^{J+1}_{++} \to \mathbb{R}_{++}$ is log-convex if $\ln f$ is convex.

3.2 Relation to Existing Discrete Choice Models

This section relates the family of choice probability functions covered by the representation theorem to existing families of choice probability functions.

ARUM. The ARUM choice probabilities and the choice probabilities satisfying Assumption 1 only differ in their respective properties (P·4) and (P·4'). (P·4) implies that the substitution matrix is everywhere positive semi-definite. (P·4') additionally requires continuous differentiability of the choice probability, which is not mandated by the Daly-Zachary conditions. However, this distinction is practically insignificant, as most ARUM choice probabilities used in the literature are continuously differentiable. Consequently, the set of choice probability functions satisfying Assumption 1 is practically strictly larger than the set of ARUM choice probability functions.

Furthermore, for any choice probability function $\boldsymbol{\sigma}(\boldsymbol{\delta})$ satisfying Assumption 1, there exists a function $\mathbf{h}\left(e^{\boldsymbol{\delta}}\right)=\left(h_0\left(e^{\boldsymbol{\delta}}\right),\ldots,h_J\left(e^{\boldsymbol{\delta}}\right)\right)$ defined by $h_j\left(e^{\boldsymbol{\delta}}\right)=\frac{1}{\mu}\partial H\left(e^{\boldsymbol{\delta}}\right)/\partial\delta_j, j\in\mathcal{J}$ such that $\sigma_j(\boldsymbol{\delta})=h_j\left(e^{\boldsymbol{\delta}}\right)/\sum_{k\in\mathcal{J}}h_k\left(e^{\boldsymbol{\delta}}\right)$, $j\in\mathcal{J}$, where $\mathbf{h}(e^{\boldsymbol{\delta}})$ satisfies specific properties.[17] This relates closely to Fosgerau, Melo, de Palma, and Shum (2020) and Fosgerau, Monardo, and de Palma (2024) who demonstrate that any ARUM choice probability function can be expressed through a function $\mathbf{h}$ by using an ad hoc transformation of the gradient vector of the exponentiated consumer surplus and Roy's identity. The key contribution of Theorem 1 is to establish that this functional form emerges from a broader family of discrete choice demand models and to provide intuition for functional form and shape restrictions that are typically assumed.

GEV model. GEV models constitute a subset of ARUMs where the vector of random utility terms follows a multivariate extreme value distribution (McFadden, 1978; Anderson et al., 1992). As noted earlier, the family of choice probability functions satisfying Assumption 1 is larger than the family of GEV choice probability functions. Both families impose identical functional forms; however, the generating function of the former satisfies (H·1) – (H·5), which impose weaker constraints than (G·1) – (G·4) that the latter satisfies. Specifically, (H·2) and (H·3) coincide with (G·2) and (G·3). Restriction (G·4) applied only to first-order derivatives coincides with (H·4).[18] Restrictions (G·2) and (G·4) imply (H·2)

[17] See Appendix B.3 for details

[18] Strictly speaking, (G·4) applied only to the first-order derivatives leads to $\partial G(e^{\boldsymbol{\delta}})/\partial e^{\delta_j}\geq 0$. However, GEV models are ARUMs for which the vector of random utility terms has a distribution that is fully supported

and (H·5) but the converse does not hold.[19]

Furthermore, (G·1) – (G·4) lack economic intuition and are mathematically complex. The representation theorem addresses both issues. First, (H·1) – (H·5) are more straightforward to verify than (G·1) – (G·4), which addresses the mathematical complexity of the GEV models. Verifying log-convexity of a function (H·5) is considerably simpler than verifying the mixed partial derivatives sign property (G·4), which requires confirming that higher-order partial derivatives have appropriate signs. Second, by deriving rather than merely stating the properties, the representation theorem provides economic intuition to the GEV properties and thus helps reconcile the good properties and economic intuition in the GEV models. Properties (G·2) and (G·4), by implying that G is log-convex, ensure that the consumer surplus function is everywhere convex and thus consistency with quasi-linear utility maximization. The limit property (G·3) ensures that every product is potentially dominated. The homogeneity property (G·2) ensures that GEV choice probabilities are probabilities (i.e., they satisfy the single-unit assumption). The positivity property satisfied by the generating function (G·1) and its first-order partial derivatives in (G·4) guarantee positive choice probabilities. The mixed partial derivative sign property (G·4) relates specifically to the ARUM setting, confirming the previously established fact that it is required for random utility terms to have a valid cumulative distribution function.

For parametric GEV choice models, the weaker restrictions (H·1) – (H·5) translate into expanded parameter spaces compared to (G·1) – (G·4). Table 1 provides these expanded parameter spaces for some well-known GEV models.[20] For example, in Bresnahan et al. (1997)'s product differentiation logit model, the GEV restrictions require nesting parameters between zero and one, whereas the weaker restrictions only require nesting parameters greater than zero. As demonstrated in the next section, this expansion enables substitution patterns that the GEV restrictions cannot accommodate.

Allen and Rehbeck (2019b)'s utility model of heterogenous consumers. Let $\Delta_J = \{\mathbf{x} = (x_0, \ldots, x_J) \in \mathbb{R}_+^{J+1} : \sum_{k \in \mathcal{J}} x_k = 1\}$ denote the set of choice probabilities, with Δ_J^+ its relative interior and Δ_J° its boundary. Roy's identity (2) is equivalent to the existence of a function $u : \mathbb{R}_+^{J+1} \to \mathbb{R} \cup \{+\infty\}$ of the share vector $\mathbf{s} \in \Delta_J$ satisfying (i) u is finite for all $\mathbf{s} \in \Delta_J^+$ and infinite otherwise; (ii) u is twice continuously, convex for all $\mathbf{s} \in \Delta_J^+$; and

on $\mathbb{R}^{J+1}$. Therefore, the GEV choice probabilities are strictly positive and, therefore, $\partial G(e^{\boldsymbol{\delta}})/\partial e^{\delta_j} > 0$.

[19] See Appendix B.4.4 for the proof of these statements.

[20] See Appendix B.4.3 for the proofs.

Table 1: GEV vs. expanded set of restrictions

Generating function	(G·1) – (G·4)	(H·1) – (H·5)
Nested Logit $\sum_{g=1}^{G}\left(\sum_{j\in g} e^{\delta_j/\mu_g}\right)^{\mu_g}$	$0<\mu_g\leq 1\,\forall g$	$\mu_g>0\,\forall g$
Product Differentiation Logit $\sum_{h=1}^{H} a_h\left(\sum_{g=1}^{G_h}\left(\sum_{j\in\mathcal{G}_{hg}} e^{\delta_j/\mu_g}\right)^{\mu_g}\right)$	$a_h>0,\ \sum_{h=1}^{H} a_h=1,$ $0<\mu_g\leq 1\,\forall g$	$a_h>0,\ \sum_{h=1}^{H} a_h=1,$ $\mu_g\geq 0\,\forall g$
Paired Combinatorial Logit $\sum_{j=0}^{J-1}\sum_{k=i+1}^{J}\left(e^{\delta_j/\mu_{jk}}+e^{\delta_k/\mu_{jk}}\right)^{\mu_{jk}}$	$0<\mu_{jk}\leq 1\,\forall j\neq k$	$\mu_{jk}>0\,\forall j\neq k$
Cross Nested Logit $\sum_{m\in\mathcal{M}}\left(\sum_{j\in\mathcal{J}}\alpha_{jm}e^{\sigma_m\delta_j}\right)^{\frac{\sigma}{\sigma_m}}$	$\alpha_{jm}\geq 0\,\forall j,m,$ $\sum_{m\in\mathcal{M}}\alpha_{jm}>0\,\forall j,$ $\sigma>0,\ \sigma\leq\sigma_m\,\forall m$	$\alpha_{jg}\geq 0\,\forall j,m,$ $\sum_{m\in\mathcal{M}}\alpha_{jm}>0\,\forall j,$ $\sigma>0,\ \sigma_m>0\,\forall m$

(iii) $\lim_{\mathbf{s}\to\Delta_J^\circ}\left|[\partial u(\mathbf{s})/\partial s_j]_{j\in\mathcal{J}}\right|\to+\infty$, such that[21]

$$\boldsymbol{\sigma}(\boldsymbol{\delta})=\underset{\mathbf{s}\in\Delta_J}{\operatorname{argmax}}\left\{\sum_{j\in\mathcal{J}}\delta_j s_j-u(\mathbf{s})\right\}. \tag{5}$$

This utility model of single unit, discrete choice describes a consumer choosing a vector $\mathbf{s}\in\Delta_J$ of shares of the differentiated products to maximize her quasi-linear concave utility function $\sum_{j\in\mathcal{J}}\delta_j s_j-u(\mathbf{s})$ (Hofbauer and Sandholm, 2002; Allen and Rehbeck, 2023). The consumer may be an individual or represent a population of heterogeneous, utility-maximizing consumers (Allen and Rehbeck, 2019b).[22] This utility model general-

[21]This follows from Theorems 23.5 and 26.1 in Rockafellar (1970) once we identify f with CS, f^* with u, $\mathbf{x}$ with $\boldsymbol{\delta}$, and $\mathbf{x}^*$ with $\mathbf{s}$.

[22]For the latter interpretation, the utility that each individual consumer derives from choosing a vector $\mathbf{s}\in\Delta_J$ can be written as $\sum_{j\in\mathcal{J}}\delta_j s_j-\tilde{u}(\mathbf{s},\boldsymbol{\varepsilon})$ where $\boldsymbol{\varepsilon}$ is a vector of random utility terms that captures consumer heterogeneity in preferences. The representative consumer's utility function is then obtained by integrating out the distribution of $\boldsymbol{\varepsilon}$. Allen and Rehbeck (2019b, Theorem 1) provide the conditions under which this aggregation holds and the relation between u, $\tilde{u}$ and the distribution of $\boldsymbol{\varepsilon}$.

izes the ARUM (Hofbauer and Sandholm, 2002; Allen and Rehbeck, 2019b) by allowing for complementarity in demand (Allen and Rehbeck, 2019a) and a wider range of consumer behavior (Allen and Rehbeck, 2023), including taste for variety (Fosgerau et al., 2024), deliberate stochastic choice (Fudenberg et al., 2015), and stochastic choice due to rational inattention (Matějka and McKay, 2015; Fosgerau et al., 2020) or due to costly optimization (Mattsson and Weibull, 2002).

Therefore, the choice probability functions covered by the representation theorem are those consistent with the model of utility-maximizing, heterogeneous consumers satisfying mild regularity assumptions, studied by Allen and Rehbeck (2019b).

3.3 Substitution Patterns

This section investigates the implications of expanding the GEV demand parameter space on substitution patterns, as presented in Table 1. Since these implications depend on the specific GEV model chosen, I focus on the product differentiation logit (PDL) model.[23]

PDL model. The PDL model describes markets that exhibit product segmentation according to several discrete product characteristics, called sources of segmentation. Each source of segmentation defines a finite number of groups such that each product belongs to exactly one group for each source of segmentation. For example, Bresnahan, Stern, and Trajtenberg (1997) analyze the personal computer market segmented by brand status (groups are "branded" versus "non-branded") and technological position (groups are "frontier" versus "non-frontier"). Products are said to be of the same type (respectively, of a completely different type) if they belong to the same group according to all (respectively, any) sources of product segmentation. The outside good is in a group by itself.

Formally, assume that there are H sources of product segmentation indexed by h, with each source h defining G_h groups indexed by g such that products belong to groups $\mathcal{G}_{hg}$. Following Bresnahan, Stern, and Trajtenberg (1997), the PDL generating function is

$$H(e^{\boldsymbol{\delta}}) = \sum_{h=1}^{H} a_h \left(\sum_{g=1}^{G_h} \left(\sum_{k \in \mathcal{G}_{hg}} e^{\delta_k/\mu_h} \right)^{\mu_h} \right) + e^{\delta_0}.$$

The PDL generating function H satisfies the GEV restrictions (G·1) – (G·4) under $a_h > 0$ for all $h = 1, \ldots, H$, $\sum_{h=1}^{H} a_h = 1$, and $0 < \mu_h \leq 1$ for all $h = 1, \ldots, H$. It satisfies the

[23] Appendix B.5 provides the expressions for elasticities and diversion ratios used in this section.

weaker restrictions (H•1) – (H•5) under $a_h > 0$ for all $h = 1, \dots, H$, $\sum_{h=1}^{H} a_h = 1$, and $0 < \mu_h$ for all $h = 1, \dots, H$.

For simplicity, assume that the market exhibits product segmentation according to only two sources of product segmentation labeled 1 and 2, each one defining two groups: $\mathcal{G}_1$ and $\overline{\mathcal{G}}_1$ for 1, and $\mathcal{G}_2$ and $\overline{\mathcal{G}}_2$ for 2. This results in four product types: type 12 ($\mathcal{G}_1 \cup \mathcal{G}_2$), type $1\bar{2}$ ($\mathcal{G}_1 \cup \overline{\mathcal{G}}_2$), type $\bar{1}2$ ($\overline{\mathcal{G}}_1 \cup \mathcal{G}_2$), and type $\bar{1}\bar{2}$ ($\overline{\mathcal{G}}_1 \cup \overline{\mathcal{G}}_2$).

Consider the derivatives of demand for a product $j \in \mathcal{G}_1 \cup \mathcal{G}_2$ with respect to the quality index of a product $k \in \{\mathcal{G}_1 \cup \mathcal{G}_2, \mathcal{G}_1 \cup \overline{\mathcal{G}}_2, \overline{\mathcal{G}}_1 \cup \mathcal{G}_2, \overline{\mathcal{G}}_1 \cup \overline{\mathcal{G}}_2\}$

$$
\begin{aligned}
\frac{\partial \sigma_j(\boldsymbol{\delta})}{\partial \delta_k} &= -\sigma_j(\boldsymbol{\delta})\sigma_k(\boldsymbol{\delta}), && k \in \overline{\mathcal{G}}_1 \cap \overline{\mathcal{G}}_2,\\
\frac{\partial \sigma_j(\boldsymbol{\delta})}{\partial \delta_k} &= -\sigma_j(\boldsymbol{\delta})\sigma_k(\boldsymbol{\delta}) + \frac{\mu_1 - 1}{\mu_1}\sigma_{j|\mathcal{G}_1}(\boldsymbol{\delta})\sigma_{k|\mathcal{G}_1}(\boldsymbol{\delta})\sigma_{\mathcal{G}_1}(\boldsymbol{\delta}), && k \in \mathcal{G}_1 \cap \overline{\mathcal{G}}_2,\\
\frac{\partial \sigma_j(\boldsymbol{\delta})}{\partial \delta_k} &= -\sigma_j(\boldsymbol{\delta})\sigma_k(\boldsymbol{\delta}) + \frac{\mu_2 - 1}{\mu_2}\sigma_{j|\mathcal{G}_2}(\boldsymbol{\delta})\sigma_{k|\mathcal{G}_2}(\boldsymbol{\delta})\sigma_{\mathcal{G}_2}(\boldsymbol{\delta}), && k \in \overline{\mathcal{G}}_1 \cap \mathcal{G}_2\\
\frac{\partial \sigma_j(\boldsymbol{\delta})}{\partial \delta_k} &= -\sigma_j(\boldsymbol{\delta})\sigma_k(\boldsymbol{\delta}) + \frac{\mu_1 - 1}{\mu_1}\sigma_{j|\mathcal{G}_1}(\boldsymbol{\delta})\sigma_{k|\mathcal{G}_1}(\boldsymbol{\delta})\sigma_{\mathcal{G}_1}(\boldsymbol{\delta}) \\
&\quad + \frac{\mu_2 - 1}{\mu_2}\sigma_{j|\mathcal{G}_2}(\boldsymbol{\delta})\sigma_{k|\mathcal{G}_2}(\boldsymbol{\delta})\sigma_{\mathcal{G}_2}(\boldsymbol{\delta}), && k \in \mathcal{G}_1 \cap \mathcal{G}_2.
\end{aligned}
$$

Substitution patterns. The parameters μ_h, commonly called nesting parameters, capture the degree of substitutability among products within the same group according to the sources h of product segmentation. These parameters determine substitution patternss.

To better understand substitution patterns in the PDL model, assume that all products have equal market shares and prices. Let $\varepsilon_{A,B}$ denote the cross-price elasticity of demand for a product of type A with respect to the price of a product of type B. Then, as highlighted by Bresnahan, Stern, and Trajtenberg (1997), when $0 < \mu_1, \mu_2 \leq 1$, the following holds

$$
\varepsilon_{12,12} \geq (\varepsilon_{12,\bar{1}2}, \varepsilon_{12,1\bar{2}}) \geq \varepsilon_{12,\bar{1}\bar{2}}, \text{ and } \varepsilon_{12,\bar{1}2} \lesseqgtr \varepsilon_{12,1\bar{2}}, \tag{6}
$$

This inequality chain indicates that when both nesting parameters μ_1 and μ_2 are between zero and one, consumers are most likely to substitute between products of similar types. Products of the same type exhibit the highest degree of substitution, while products of completely different types show the lowest. Products sharing only one group show an intermediate degree of substitution. The relative degree of substitution between products sharing only one group according to the first source of product segmentation versus the second source of product segmentation depends on the relative values of the nesting parameters μ_1 and μ_2: when $\mu_1 > \mu_2$, the degree of substitution is higher between products sharing only one group according to the first source of segmentation than those sharing only one group according to the second source of segmentation.

When $\mu_1 > 1$ and $0 < \mu_2 \leq 1$, the ordering changes to

$$\varepsilon_{12,\bar{1}2} \geq (\varepsilon_{12,12}, \varepsilon_{12,\bar{1}\bar{2}}) \geq \varepsilon_{12,1\bar{2}}, \text{ and } \varepsilon_{12,12} \lesseqqgtr \varepsilon_{12,\bar{1}\bar{2}}, \tag{7}$$

In this case, products that differ along the first source of segmentation become stronger substitutes than products of the same type. This represents a significant departure from the PDL model under the GEV restrictions, reflecting more complex substitution patterns than under the GEV restrictions.

Finally, when $\mu_1, \mu_2 > 1$, the ordering reverses from the $0 < \mu_1, \mu_2 \leq 1$ case

$$\varepsilon_{12,\bar{1}\bar{2}} \geq (\varepsilon_{12,1\bar{2}}, \varepsilon_{12,\bar{1}2}) \geq \varepsilon_{12,12}, \text{ and } \varepsilon_{12,1\bar{2}} \lesseqqgtr \varepsilon_{12,\bar{1}2}, \tag{8}$$

which implies that when both nesting parameters exceed one, consumers prefer to substitute between products of completely different types rather than between more similar products.

Between-group and within-group substitution. The seemingly counterintuitive substitution patterns in (7) and (8) represent a manifestation of complementarity.

To analyze this phenomenon, consider the effects of an increase in the quality index of a product $j \in \mathcal{G}_1 \cup \mathcal{G}_2$ on the choice probability of competing products k of different types, $k \in \{\mathcal{G}_1 \cup \mathcal{G}_2, \mathcal{G}_1 \cup \overline{\mathcal{G}}_2, \overline{\mathcal{G}}_1 \cup \mathcal{G}_2, \overline{\mathcal{G}}_1 \cup \overline{\mathcal{G}}_2\}$. These effects are formalized in the derivatives of demand for product $j \in \mathcal{G}_1 \cup \mathcal{G}_2$ with respect to the quality index of product k.

Two distinct effects operate. First, a negative between-group effect, indicating between-group substitution and represented by $-\sigma_j(\boldsymbol{\delta})\sigma_k(\boldsymbol{\delta})$, decreases the choice probability of product k, regardless of group membership. This effect operates consistently across all products and constitutes the baseline level of substitution that exists between any two products in the market. It thus represents the market-wide substitution pattern that would exist without the segmentation structure. Second, a within-group effect operates exclusively through the nesting structure between products sharing at least one group membership. This effect equals zero when products j and k are of completely different types. Its magnitude and direction depend on the values of the nesting parameters μ_1 and μ_2 and on group membership. For each source of segmentation h where products j and k share group membership, the within-group effect is negative, which indicates within-group substitutability, when μ_h is between zero and one, further decreasing product k's choice probability and thus intensifying substitution. It becomes positive, which indicates within-group complementarity, when μ_h exceeds one, increasing product k's choice probability and thus offsetting the between-group effect.

When nesting parameters μ_h exceed one, they transform the nature of substitution patterns by creating complementarity effects, causing dissimilar products to become stronger substitutes than similar ones. When μ_h becomes greater than one, the within-group effect along the source of segmentation h shifts from negative to positive, indicating that an improvement in product k's quality index increases rather than decreases the conditional probability of choosing product j within the same group. This constitutes a complementarity effect along the source of segmentation h where products within the same group complement rather than substitute for each other. Complementarity in demand occurs when the within-group complementarity effect exceeds the between-group substitution effect.[24] This occurs when μ_1 and/or μ_2 are sufficiently larger one.

A compelling case for within-group complementarity is the halo effect. A market exhibits a channel halo effect when a product's presence or enhanced performance in one distribution channel generates positive spillover effects that strengthen its performance across other channels rather than cannibalizing them. This effect manifests prominently in the relationship between online (digital) and offline (physical) channels. In the hotel industry, a channel halo effect occurs when a hotel's presence on online travel agencies (OTAs) boosts direct bookings. This outcome can be attributed to awareness generation, with OTA listings functioning as discovery platforms that enhance visibility and consumer trust.

The PDL model can capture channel halo effect in the hotel industry by defining two sources of segmentation, where source $h = 1$ corresponds to hotel brand or property, and sources $h = 2$ represents booking channels (OTAs versus direct booking channels). The model exhibits a channel halo effect between booking channels if $\mu_2 > 1$, indicating com-

[24] The empirical literature has mainly relied on two definitions of complementarity (Berry, Khwaja, Kumar, Musalem, Wilbur, Allenby, Anand, Chintagunta, Hanemann, Jeziorski, and Angelo, 2014): for a differentiable demand function $\boldsymbol{\sigma}(\boldsymbol{\delta})$ consistent with a twice differentiable direct utility function $u(\mathbf{s})$, where $\mathbf{s} = (s_0, \dots, s_J)$ denotes the vector of market shares, products j and k are complements in demand if $\partial\sigma_j(\boldsymbol{\delta})/\partial\delta_k > 0$ and they are complements in utility if $\partial^2 u(\mathbf{s})/\partial s_j \partial s_k > 0$. For $J > 2$, the two definitions do not coincide (Samuelson, 1974). In this paper, I use the first definition, which contrasts with Gentzkow (2007) who uses the second definition. The fact that the PDL model allows for complementarity relates to Feng, Li, and Wang (2018) and Allen and Rehbeck (2019a) and can be the manifestation of different effects, including attraction, compromise, and halo effects.

plementarity between channels. Moreover, $\mu_1 < 1$ captures traditional substitution between different hotel brands, reflecting inter-brand competition.

Another manifestation of the halo effect is the brand halo effect, where a brand's enhanced reputation or a successful product generates positive spillover effects across other products within the same brand portfolio.[25] In the beer market segmented by beer types (ales and lagers) and brands (Budweiser and Miller), if Budweiser advertises its lager and the nesting parameter along the type segmentation exceeds one, this campaign boosts sales of Budweisers American ale, thereby creating a brand halo effect.

Mechanism. An increase in the quality index of product $k \in \mathcal{G}_1 \cup \overline{\mathcal{G}}_2$ has two opposite effects on the probability of choosing product $j \in \mathcal{G}_1 \cup \mathcal{G}_2$.[26] A positive spillover effect operates via the probability of group $\mathcal{G}_1$. Since a group's attractiveness depends on the attractiveness of all products within that group, an increase in product k's quality index enhances the attractiveness of group $\mathcal{G}_1$, which increases the probability of choosing any product in group $\mathcal{G}_1$, including product j. A negative within-group substitution effect operates via the probability of choosing conditional on $\mathcal{G}_1$. An increase in product k's quality index makes it more attractive in its group $\mathcal{G}_1$, which decreases the probability of choosing any other product of group $\mathcal{G}_1$, including product j. Complementarity in demand occurs when the positive spillover effect exceeds the negative within-group substitution effect.

3.4 Identification

This section discusses identification using market-level data, i.e., market shares s_{jt}, prices p_{jt}, product/market characteristics $\mathbf{x}_{jt}$, and instruments $\mathbf{z}_{jt}$ for products $j = 1, \dots, J$ and markets $t = 1, \dots, T$.[27]

Let the quality index for product j in market t be specified as

$$\delta_{jt} = \mathbf{x}_{jt}\boldsymbol{\beta} - \alpha p_{jt} + \xi_{jt}, \quad j = 1, \dots, J, \quad t = 1, \dots, T, \tag{9}$$

where $\xi_{jt} \in \mathbb{R}$ is the unobserved characteristics term for product-market jt, which captures product differentiation not explained by observed characteristics. Assume that $\delta_{0t} = 0$ for

[25] For example, a successful brand extension or a well-received advertisement for one product of a given brand can create a brand halo effect on the other products of the brand.

[26] See Feng et al. (2018, Section 2) and Allen and Rehbeck (2023, Section 4.3) for a similar discussion.

[27] Market shares can be computed from observed quantities, volumes, or individual choice data depending on data availability.

all $t = 1, \ldots, T$.

Following the literature, product characteristics $\mathbf{x}_{jt}$ are considered exogenous, i.e., uncorrelated with ξ_{jt}, while prices p_{jt} and market shares s_{jt} are treated as endogenous.

Identification. Following Berry and Haile (2014), a key requirement for demand identification is the invertibility of the system of market share equations for each market t

$$\sigma_j(\boldsymbol{\delta}_t) = s_{jt}, \quad j = 1, \ldots, J.$$

The following lemma, which builds on Eriksson (1986, Proposition 2), establishes the invertibility of theses market share equations for all market share functions encompassed by the representation theorem.

Lemma 1. Consider any vector of market shares $(s_{1t}, \ldots, s_{Jt})^\intercal$ such that $s_{jt} > 0$ for all $j = 1, \ldots, J$ and $\sum_{k=1}^{J} s_{kt} < 1$. Let $\sigma_j(\boldsymbol{\delta}_t)$ be defined by (3), with $H(e^{\boldsymbol{\delta}_t})$ being a twice continuously differentiable function defined over $\mathbb{R}_{++}^{J+1}$ and satisfying (H$\cdot$1) – (H$\cdot$5) for all $e^{\boldsymbol{\delta}_t} \in \mathbb{R}_{++}^{J+1}$. Then, there exists a unique vector $\boldsymbol{\delta}_t$ such that $\sigma_j(\boldsymbol{\delta}_t) = s_{jt}$ for all $j = 1, \ldots, J$.

Invertibility ensures the existence of an inverse market share function $\boldsymbol{\sigma}^{-1} = (\sigma_0^{-1}, \ldots, \sigma_J^{-1})$ such that $\sigma_j^{-1}(\mathbf{s}_t) = \delta_{jt}$ for all $j = 1, \ldots, J$, which, combined with (9), yields

$$\xi_{jt} = -\mathbf{x}_{it}\boldsymbol{\beta} + \alpha p_{jt} + \sigma_j^{-1}(\mathbf{s}_t).$$

In addition to invertibility, demand identification requires the existence of instruments $\mathbf{z}_t$ for prices and market shares that satisfy standard exclusion and completeness restrictions. Following Berry and Haile (2014, Theorem 1), the following establishes the identification of the choice probability functions encompassed by the representation theorem.

Proposition 1. Assume there exist instruments $\mathbf{z}_t$ such that (i) $\mathbb{E}[\xi_{jt}|\mathbf{x}_t, \mathbf{z}_t]$ for all $j = 1, \ldots, J$ (exclusion restriction); and (ii) for all functions $B(\mathbf{s}_t, \mathbf{p}_t)$ with finite expectation, if $\mathbb{E}[B(\mathbf{s}_t, \mathbf{p}_t)] = 0$ almost surely, then $B(\mathbf{s}_t, \mathbf{p}_t) = 0$ almost surely (completeness restriction). Let $\sigma_j(\boldsymbol{\delta}_t)$ be defined by (3), with $H(e^{\boldsymbol{\delta}_t})$ being a twice continuously differentiable function defined over $\mathbb{R}_{++}^{J+1}$ and satisfying (H$\cdot$1) – (H$\cdot$5) for all $e^{\boldsymbol{\delta}_t} \in \mathbb{R}_{++}^{J+1}$. Then, for all $j = 1, \ldots, J$, $\sigma_j(\boldsymbol{\delta}_t)$ is identified on $\{(\delta_{0t}, \ldots, \delta_{Jt}) \in \mathbb{R}^{J+1} : \delta_{0t} = 0\}$.

Discussion. An intuitive discussion about the type of variation in the data that helps identify demand parameters is helpful. For concreteness, I focus on the PDL model.

As is well known, exogenous variables that shift costs and/or markups are candidates as instruments for identification of the price parameter (see, e.g., Berry and Haile, 2021, Section 4.2). These include Hausman instruments (prices in other markets), BLP-type instruments (functions of the characteristics of competing products), exogenous shocks in market structure, and Waldfogel-Fan instruments (characteristics of nearby markets such as average demographic measures).

Furthermore, in the PDL model, as formalized in (6) – (8), the nesting parameters determine substitution patterns in a specific way: the relative magnitudes of different substitution types, captured by the elasticities types $\varepsilon_{A,B}$ where $A, B \in \{12, \bar{1}, 1\bar{2}, \bar{1}\bar{2}\}$, inform us about the values of the nesting parameters. Intuitively, by observing how market shares respond to exogenous variation in the choice set across different product types, we can learn about the values of the nesting parameters. Therefore, variables that generate exogenous variation in the choice set serve as appropriate instruments for the identification of the nesting parameters. In particular, these include the cost and markup shifters used for the identification of the price parameter. The difference is that these instruments identify the price parameter through the overall magnitude of market share responses to exogenous price variation, whereas they identify the nesting parameters through the relative changes in market share across different product types.

Consider again the hotel sector segmented by hotel brand (brand 1/brand 2) and sales channels (OTA/direct channel). When a hotel brand 1 room booked through an OTA becomes unavailable or changes price, consumers might switch to (i) other hotel brand 1 rooms on the OTA; (ii) hotel brand 1 rooms through the direct channel; (iii) brand 2 rooms on the OTA; or (iv) to brand 2 rooms through the direct channel. The relative strength of these substitution types (i) to (iv) provides crucial variation needed to separately identify μ_1 (the hotel brand parameter) from μ_2 (the sales channel parameter). For instance, if substitution type (i) is strongest, (iv) is weakest, and (ii) is stronger than (iii), this suggests both μ_1 and μ_2 should be between zero and one, with μ_1 exceeding μ_2.

3.5 Algorithm for Demand Inversion

The leading estimation methods in the literature, based on individual-level and/or market-level data, require numerically inverting the choice probability (or market share) function, a crucial step in the estimation procedure referred to as demand inversion.[28]

Demand inversion requires finding the unique vector $\boldsymbol{\delta}_t^* = (0, \delta_{1t}^*, \dots, \delta_{Jt}^*)^\intercal$ of net qualities for each market t such that

$$\sigma_j(\boldsymbol{\delta}_t^*) = s_{jt} \quad \Leftrightarrow \quad \delta_{jt}^* = \sigma_j^{-1}(\mathbf{s}_t), \quad j = 1, \dots, J. \tag{10}$$

Except for the logit and nested logit models, the choice probability function $\boldsymbol{\sigma}(\boldsymbol{\delta}_t)$ does not generally yield a closed-form expression for its inverse $\boldsymbol{\sigma}^{-1}(\mathbf{s}_t)$. Consequently, $\boldsymbol{\delta}_t^*$ cannot be computed analytically as a function of $\mathbf{s}_t$ (and demand parameters) based on the equality on the right-hand side of the equivalence (10), thus requiring a numerical inversion. Thus, estimation requires the invertibility of the function $\boldsymbol{\sigma}(\boldsymbol{\delta}_t)$ and an algorithm to numerically compute $\boldsymbol{\delta}_t^*$.

Lemma 1 establishes that the choice probability functions (3) encompassed by the representation theorem are invertible in $\boldsymbol{\delta}_t$. Moreover, the proof of invertibility is constructive, providing a numerical algorithm to compute the vector $\boldsymbol{\delta}_t^*$ of quality indices. Specifically, for the choice probability function (3), $\boldsymbol{\delta}_t^*$ can be computed as the unique solution of the unconstrained convex optimization problem

$$\min_{\boldsymbol{\delta}_t \in \{\boldsymbol{\delta}_t \in \mathbb{R}^{J+1} : \delta_{0t} = 0\}} \left\{ \frac{1}{\mu} \ln H(e^{\boldsymbol{\delta}_t}) - \sum_{k \in \mathcal{J}} \delta_{kt} s_{kt} \right\}.$$

The use of convex programming for numerical demand inversion is not novel; it has been explored in the context of ARUM choice probability functions with or without random coefficients (Davis and Schiraldi, 2014; Li, 2018; Fosgerau, Melo, Shum, and Sørensen, 2021). However, the contribution here lies in demonstrating that this approach extends beyond the ARUM framework to models where complementarity arises. In the following section, I provide a comparative analysis of the proposed convex programming algorithm against the widely-used SQUAREM acceleration technique for the random coefficient logit model, evaluating performance in terms of computational efficiency and convergence properties across various market configurations.

[28]See Conlon and Gortmaker (2020, 2025) for a review of these methods.

4 Extension

This section extends the framework of the previous section. For this, I consider a population of heterogenous consumers $i = 1, \ldots, N$ facing the choice set $\mathcal{J}$. First, I assume that the population of consumers is segmented into a finite number R of types $r = 1, \ldots, R$, with each segment r characterized by its own generating function $H^r(e^{\boldsymbol{\delta}})$. Then, I represent this setting as a discrete-type random coefficient model, where the vector of coefficients capturing unobserved heterogeneity in taste takes a finite number of possible values (see, e.g., Berry and Jia, 2010). Lastly, I further generalize this setting by allowing $R = N$, thus allowing each consumer to represent a unique type. This generalization yields a discrete approximation of the conventional random coefficient demand function, wherein random coefficients follow a joint continuous distribution. This formulation leads to an algorithm for demand inversion that is applicable to all models presented in this section. Proofs of this section are provided in Appendix C.

4.1 Multiple-Segment Choice Model

Assume that the population of consumers is segmented into a finite number R of types $r = 1, \ldots, R$. Each consumer of type r is described by a choice probability function $\boldsymbol{\sigma}^r(\boldsymbol{\delta}) = (\sigma_0^r(\boldsymbol{\delta}), \ldots, \sigma_J^r(\boldsymbol{\delta}))$ satisfying Assumption 1. That is, the individual choice probability function of a consumer of type r is given by

$$\sigma_j^r(\boldsymbol{\delta}) = \frac{e^{\delta_j} \dfrac{\partial H^r\left(e^{\boldsymbol{\delta}}\right)}{\partial e^{\delta_j}}}{\mu H^r\left(e^{\boldsymbol{\delta}}\right)}, \quad j \in \mathcal{J},$$

where the twice continuously differentiable function $H^r(e^{\boldsymbol{\delta}})$ defined over $\mathbb{R}_{++}^{J+1}$ satisfies (H·1) – (H·5) for all $e^{\boldsymbol{\delta}} \in \mathbb{R}_{++}^{J+1}$.

Besides, her preferences are represented by a surplus function given by

$$CS^r(\boldsymbol{\delta}) = \frac{1}{\mu} \ln H^r\left(e^{\boldsymbol{\delta}}\right).$$

Let w^r denote the percentage of type r consumers in the population, such that $w^r > 0$ for all $r = 1, \ldots, R$, and $\sum_{r=1}^{R} w^r = 1$. Then, the market share of product j is computed by aggregating the individual choice probabilities over consumer types

$$\sigma_j(\boldsymbol{\delta}) = \sum_{r=1}^{R} w^r \sigma_j^r(\boldsymbol{\delta}) = \sum_{r=1}^{R} w^r \frac{e^{\delta_j} \dfrac{\partial H^r\left(e^{\boldsymbol{\delta}}\right)}{\partial e^{\delta_j}}}{\mu^r H^r\left(e^{\boldsymbol{\delta}}\right)}, \quad j \in \mathcal{J}, \tag{11}$$

and the aggregate consumer surplus function is

$$CS(\boldsymbol{\delta}) = \sum_{r=1}^{R} w^r CS^r(\boldsymbol{\delta}) = \sum_{r=1}^{R} \frac{w^r}{\mu^r} \ln H^r\left(e^{\boldsymbol{\delta}}\right).$$

There are two polar cases: first, $R = 1$, which corresponds to the model of the previous section; second, $R = N$, where each consumer represents a unique type.

It is easy to show that if each $\boldsymbol{\sigma}^r(\boldsymbol{\delta})$ satisfies Assumption 1, then so does $\boldsymbol{\sigma}(\boldsymbol{\delta})$. Furthermore, the aggregate consumer surplus function can be conveniently rewritten as

$$CS(\boldsymbol{\delta}) = \ln\left(\prod_{r=1}^{R}\left[H^r\left(e^{\boldsymbol{\delta}}\right)\right]^{\dfrac{w^r}{\mu^r}}\right)$$

to show that $H(e^{\boldsymbol{\delta}}) = \prod_{r=1}^{R}[H^r\left(e^{\boldsymbol{\delta}}\right)]^{w^r/\mu^r}$ can be considered as a generating function satisfying (H·1) – (H·5).

The following proposition summarizes this.

Proposition 2. Let $\boldsymbol{\sigma}(\boldsymbol{\delta}) = \sum_{r=1}^{R} w^r \boldsymbol{\sigma}^r(\boldsymbol{\delta})$, where $w^r > 0$ for all $r = 1, \ldots, R$, and $\sum_{r=1}^{R} w^r = 1$, and where each $\sigma^r(\boldsymbol{\delta})$, $r = 1, \ldots, R$, satisfies Assumption 1. Then, $\boldsymbol{\sigma}(\boldsymbol{\delta})$ satisfies Assumption 1.

Furthermore, let $H(e^{\boldsymbol{\delta}}) = \prod_{r=1}^{R}[H^r\left(e^{\boldsymbol{\delta}}\right)]^{w^r/\mu^r}$, where $w^r > 0$ for all $r = 1, \ldots, R$, and $\sum_{r=1}^{R} w^r = 1$, and where each $H^r(e^{\boldsymbol{\delta}})$, $r = 1, \ldots, R$, defined over $\mathbb{R}_{++}^{J+1}$, is a twice continuously differentiable function satisfying (H·1) – (H·5) for all $e^{\boldsymbol{\delta}} \in \mathbb{R}_{++}^{J+1}$. Then, $H(e^{\boldsymbol{\delta}})$, defined over $\mathbb{R}_{++}^{J+1}$, is a twice continuously differentiable function satisfying (H·1) – (H·5) for all $e^{\boldsymbol{\delta}} \in \mathbb{R}_{++}^{J+1}$.

Identification using market-level data. Building on Eriksson (1986, Theorem 3) and Berry and Haile (2014, Theorem 1), the following proposition establishes the invertibility of these market share equations and the identification of the choice probability functions (11), respectively.

Proposition 3. Consider any vector of market shares $(s_{1t}, \ldots, s_{Jt})^\intercal$ such that $s_{jt} > 0$ for all $j = 1, \ldots, J$ and $\sum_{k=1}^{J} s_{kt} < 1$, and any vector $\mathbf{w} = (w^1, \ldots, w^R)^\intercal$ such that $w^r > 0$ for all $r = 1, \ldots, R$ and $\sum_{r=1}^{R} w^r = 1$. Let $\sigma_j(\boldsymbol{\delta}_t)$ be defined by (11), with each $H^r(e^{\boldsymbol{\delta}_t})$, $r = 1, \ldots, R$ a twice continuously differentiable function defined over $\mathbb{R}^{J+1}_{++}$ satisfying properties (H·1) – (H·5) for all $e^{\boldsymbol{\delta}_t} \in \mathbb{R}^{J+1}_{++}$. Assume that there exists at least one $H^r(e^{\boldsymbol{\delta}_t})$ that is strictly log-convex in $\boldsymbol{\delta}$ for all $e^{\boldsymbol{\delta}_t} \in \mathbb{R}^{J+1}$. Then, there exists a unique vector $\sigma_j(\boldsymbol{\delta}_t) = s_{jt}$ for all $j = 1, \ldots, J$.

Furthermore, assume there exist instruments $\mathbf{z}_t$ such that (i) $\mathbb{E}[\xi_{jt}|\mathbf{x}_t, \mathbf{z}_t]$ for all $j = 1, \ldots, J$ (exclusion restriction); and (ii) for all functions $B(\mathbf{s}_t, \mathbf{p}_t)$ with finite expectation, if $\mathbb{E}[B(\mathbf{s}_t, \mathbf{p}_t)] = 0$ almost surely, then $B(\mathbf{s}_t, \mathbf{p}_t) = 0$ almost surely (completeness restriction). Then, for all $j = 1, \ldots, J$, $\sigma_j(\boldsymbol{\delta}_t)$ is identified on $\{(\delta_{0t}, \ldots, \delta_{Jt}) \in \mathbb{R}^{J+1} : \delta_{0t} = 0\}$.

4.2 Random Coefficient Model

4.2.1 Discrete Types.

Let the utility that consumer i of type r derives from product j be equal to $\delta_j + \mu_{rj}$, where δ_j is the component of utility that is common across consumers and μ_{ij} is the component that captures heterogeneity across consumer types. Let $\boldsymbol{\mu}_r = (\boldsymbol{\mu}_{r0}, \ldots, \boldsymbol{\mu}_{rJ})$.

The discrete-type random coefficient model can be obtained by replacing $H^r(e^{\boldsymbol{\delta}})$ with $H(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_r})$ in (11), so the market share of product j is given by

$$\sigma_j(\boldsymbol{\delta}, \{\boldsymbol{\mu}_r\}_{r=1}^{R}) = \sum_{r=1}^{R} w^r \frac{\frac{\partial H\left(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_r}\right)}{\partial e^{\delta_j + \mu_{rj}}}}{\mu H\left(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_r}\right)}, \tag{12}$$

and the consumer surplus function is given by $CS(\boldsymbol{\delta}, \{\boldsymbol{\mu}_r\}_{r=1}^{R}) = \ln(\prod_{r=1}^{R}[H\left(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_r}\right)]^{w^r/\mu})$, where $w^r > 0$ for all $r = 1, \ldots, R$, and $\sum_{r=1}^{R} w^r = 1$. Therefore, the corresponding generating function is $H(\boldsymbol{\delta}, \{\boldsymbol{\mu}_r\}_{r=1}^{R}) = \prod_{r=1}^{R}[H\left(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_r}\right)]^{w^r/\mu}$ and the market share function (12) is invertible.

4.2.2 Random Coefficient Model with Continuous Mixing Distribution

Let the net utility that consumer i derives from product j be equal to $\delta_j + \mu_{ij}$, where δ_j is the component of the net utility that is common across consumers and μ_{ij} is the component that captures consumer heterogeneity. Let $\boldsymbol{\mu}_i = (\boldsymbol{\mu}_{i0}, \ldots, \boldsymbol{\mu}_{iJ})$. Each consumer represents a unique type ($R = N$).

The market share function is computed by aggregating the individual choice probabilities over consumers, i.e.,

$$\sigma_j(\boldsymbol{\delta}; \boldsymbol{\theta}_2) = \int \sigma_{ij}(\boldsymbol{\delta}, \boldsymbol{\mu}_i) f(\boldsymbol{\mu}_i|\boldsymbol{\theta}_2) \mathrm{d}\boldsymbol{\mu}_i, \tag{13}$$

where $f(\boldsymbol{\mu}_i|\boldsymbol{\theta}_2)$ is the continuous mixing distribution over consumers i parametrized by $\boldsymbol{\theta}_2$, justifying why these models are usually referred to as mixed or random coefficients discrete choice demand models. This setting encompasses the popular random coefficient logit and random coefficient nested logit models, as well as any mixture of GEV models.

The random coefficient demand (13) is usually approximated at a finite set of I nodes v_i with associated weights w_i as follows

$$\sigma_j(\boldsymbol{\delta}; \boldsymbol{\theta}_2) \approx \sum_{i=1}^{I} w^i \frac{\dfrac{\partial H\left(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_i(v_i;\boldsymbol{\theta}_2)}\right)}{\partial e^{\delta_j+\mu_{ij}(v_i;\boldsymbol{\theta}_2)}}}{\mu H\left(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_i(v_i;\boldsymbol{\theta}_2)}\right)}, \tag{14}$$

where the right-hand side can be obtained by replacing $H(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_r})$ with $H(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_i(v_i;\boldsymbol{\theta}_2)})$ in (12), so the consumer surplus function is $CS(\boldsymbol{\delta}; \boldsymbol{\theta}_2) = \ln(\prod_{i=1}^{N}[H\left(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_i(v_i;\boldsymbol{\theta}_2)}\right)]^{w^i/\mu})$, where $w^i > 0$ for all $i = 1, \ldots, N$, and $\sum_{i=1}^{N} w^i = 1$. Therefore, the corresponding generating function is $H(e^{\boldsymbol{\delta}}; \boldsymbol{\theta}_2) = \prod_{i=1}^{N}[H\left(e^{\boldsymbol{\delta}+\boldsymbol{\mu}_i(v_i;\boldsymbol{\theta}_2)}\right)]^{w^i/\mu}$ and the market share function (14) is invertible.

4.3 Algorithm for Demand Inversion

Algorithm. The vector $\boldsymbol{\delta}_t^*$ that solve the market share equations can be computed as the unique solution of the unconstrained convex optimization problem

$$\min_{\boldsymbol{\delta}_t \in \{\boldsymbol{\delta}_t \in \mathbb{R}^{J+1}: \delta_{0t}=0\}} \left\{ \sum_{r=1}^{R} \frac{w^r}{\mu^r} \ln H^r\left(e^{\boldsymbol{\delta}_t}\right) - \sum_{k \in \mathcal{J}} \delta_{kt} s_{kt} \right\},$$

which applies to the multiple-segment choice market share (11) when $H^r\left(e^{\boldsymbol{\delta}_t}\right) = H^r\left(e^{\boldsymbol{\delta}_t}\right)$ and $\mu^r = \mu^r$, to the discrete-type random coefficient market share (12) when $H^r\left(e^{\boldsymbol{\delta}_t}\right) = H\left(e^{\boldsymbol{\delta}_t+\boldsymbol{\mu}_{rt}}\right)$ and $\mu^r = \mu$, and to the conventional random coefficient market share (14) when $H^r\left(e^{\boldsymbol{\delta}}\right) = H\left(e^{\boldsymbol{\delta}_t+\boldsymbol{\mu}_i(v_{it};\boldsymbol{\theta}_2)}\right)$ and $\mu^r = \mu$.

Demand inversion can be numerically challenging. This has led the literature to propose different algorithms: a fixed-point algorithm (Berry, Levinsohn, and Pakes, 1995), accelerated fixed-point algorithms known as the SQUAREM and spectral algorithms (Reynaerts, Varadha, and Nash, 2012), and a mathematical program with equilibrium constraints algorithm (Dubé, Fox, and Su, 2012).

The recent literature often uses on the SQUAREM algorithm (Conlon and Gortmaker, 2020; Pál and Sándor, 2023; Conlon and Gortmaker, 2025). The remainder of this subsection is used to compare the performance of the algorithm based on convex optimization, referred to as the convex algorithm, with that of the SQUAREM algorithm.

Theoretically speaking, the convex algorithm only requires solving the unconstrained convex optimization problem that directly arises from the chosen demand model, without the need to consider the appropriate problem to solve. As a result, the convex algorithm is guaranteed to converge from any starting values, benefits from efficient, robust algorithms available in many software packages, and relies on well-defined mathematical properties. By contrast, the SQUAREM algorithm is not guaranteed to converge and requires taking a stand on which fixed-point iteration mapping to use.[29]

Simulations. To compare the performance of the convex and SQUAREM algorithms, I use Pál and Sándor (2023)'s exact same simulations of the random coefficient logit model with normally distributed random coefficients. For the estimation, I use their same implementation of Berry, Levinsohn, and Pakes (1995)'s GMM minimization procedure, except that I parallelize the demand inversions across markets and do not constrain the standard deviation of the normally distributed random coefficients to be non-negative.[30] See Pál and Sándor (2023) for details.

I run all the estimations on a personal computer with 12th Gen Intel(R) Core(TM) i7-12700, 2100 Mhz, 12 Cores CPU, and 32 GB RAM, Windows operating system (Windows 11), Matlab R2023b. The central optimizer is Knitro 14.0. For both Berry, Levinsohn, and Pakes (1995)'s GMM minimization problem and the unconstrained convex optimization problem of the convex algorithm, I use the interior-point algorithm from KNITRO while providing the analytic gradient and hessian of the objective functions.

The data generating processes (DGP) vary with the number of markets T, the number

[29] See e.g., Berry and Jia (2010) and Grigolon (2021) who modify the original BLP contraction mapping for the random coefficient nested logit model and the ordered nested logit model, respectively.

[30] Indeed, due to the symmetry of the normal distribution, the sign of the standard deviation of the normal distribution is not identified. Furthermore, in several instances, when constraining the standard deviation to be non-negative, the GMM minimization problem reached the zero boundary.

of products per market J, and the value of the parameter on the constant β_1 that controls for the size of the outside good market share s_{0t}.

Tables 2, 3, and 4 below replicate Tables 2, 4, and 6 in Pál and Sándor (2023), respectively. For each DGP, results are based on 20 replications estimated with 5 different starting values, providing a total of 100 executions of the estimation procedure. The tables report the following indicators: the number of convergent cases out of 100 (Conv), the mean CPU time in seconds with the standard deviation in parentheses (CPU(std)), the number of cases out of 100 in which the global minima are found (GlobOpt), the number of cases out of 20 when none of the 5 starting values lead to the global minimum (FailsNr) and an estimate of the computing time in seconds needed for obtaining the global minima (CPU-Gopt).

Table 2 presents the results for DGPs with 50 markets, 25 products per market, and with a varying parameter value on the constant $\beta_1 \in \{-4, -2, 0, 2, 4\}$. According to the Conv, GlobOpt, and FailsNr indicators, the SQUAREM and convex algorithms perform equally well. For the CPU and CPU-Gopt indicators, the convex algorithm performs slightly better than SQUAREM.

Table 2: Results for different β_1 values; 50 markets, 25 products.

β_1	s_0	Method	Conv	CPU(std)	GlobOpt	FailsNr	CPU-Gopt
-4	0.94	SQUAREM	100	11(3)	66	0	70
		Convex	100	11(3)	66	0	70
-2	0.86	SQUAREM	100	11(2)	62	0	79
		Convex	100	10(2)	62	0	71
0	0.73	SQUAREM	100	11(2)	65	0	72
		Convex	100	10(2)	65	0	66
2	0.54	SQUAREM	100	10(2)	70	0	57
		Convex	100	9(2)	70	0	52
4	0.34	SQUAREM	100	10(3)	74	0	51
		Convex	100	10(3)	74	0	51

Table 3 presents the results for DGPs with 10 markets, 125 products per market, and with a varying parameter value on the constant $\beta_1 \in \{-7, -4, 0\}$. According to the Conv, GlobOpt, and FailsNr indicators, the SQUAREM and convex algorithms perform equally well. For the CPU and CPU-Gopt indicators, the convex algorithm performs better than SQUAREM.

Table 3: Results for different β_1 values; 10 markets, 125 products.

β_1	s_0	Method	Conv	CPU(std)	GlobOpt	FailsNr	CPU-Gopt
-7	0.90	SQUAREM	100	31(9)	75	0	154
		Convex	100	25(8)	75	0	125
-4	0.70	SQUAREM	100	28(8)	77	0	132
		Convex	100	21(6)	77	0	99
0	0.25	SQUAREM	100	28(7)	79	0	124
		Convex	100	18(5)	79	0	80

Table 4 presents the results for DGPs with a varying number of markets $T \in \{5, 10, 25\}$, a varying number of products per market $J \in \{50, 125, 250\}$ and with a fixed parameter value on the constant $\beta_1 = 0$. According to the Conv, and FailsNr indicators, the SQUAREM and convex algorithms perform equally well. According to the GlobOpt indicator, they also perform equally well, except for the last DGP with $T = 5$ and $J = 250$, where there is a replication/starting value combination where the SQUAREM algorithm failed to converge to the global optimum while the convex algorithm converged to the global optimum. For the CPU and CPU-Gopt indicators, the convex algorithm performs better than SQUAREM for large $J \in \{125, 250\}$ but equally well for small $J = 50$.

Table 4: Results for different J and T values; $\beta_1 = 0$.

T	J	s_0	Method	Conv	CPU(std)	GlobOpt	FailsNr	CPU-Gopt
25	50	0.49	SQUAREM	100	9(2)	73	0	47
			Convex	100	9(2)	73	0	47
10	125	0.25	SQUAREM	100	31(8)	79	0	137
			Convex	100	18(5)	79	0	80
5	250	0.12	SQUAREM	100	40(14)	69	0	236
			Convex	100	30(9)	70	0	172

Overall, the simulations show that the convex algorithm slightly outperforms the SQUAREM algorithm. This is especially true for DGPs that have a large number of products per market.

5 Concluding Remarks

In this paper, I derive the necessary and sufficient functional form and shape restrictions characterizing a large family of discrete choice demand models that satisfy mild regularity properties. This leads to a representation theorem for discrete choice demand models.

I illustrate three empirical implications of this representation theorem. First, I show that it helps understand the economic intuition behind the assumed GEV properties, reconciling their good properties and their economic interpretation. Second, I show that the representation theorem provides a specification tool for discrete choice modeling that extends beyond the ARUM framework while maintaining consistency with utility maximization, in particular by allowing for complementarity in demand. Furthermore, I find that the derived functional form is exactly the same as those of the assumed GEV models. Besides, the derived restrictions are less restrictive, much easier to check, and expand the GEV parameter space in parametric cases. Third, I propose a new numerical algorithm for demand inversion based on an unconstrained convex optimization problem and have shown that it achieves better computational performance than the often-used SQUAREM algorithm.

In this paper, I rely on the single unit assumption and assume the absence of income effect. The recent literature has shown how relaxing the single-unit assumption allows us to obtain greater flexibility (Birchall, Mohapatra, and Verboven, 2024) and that the no income effect assumption may restrict demand slope and curvature as well as pass-through and merger price effects (Griffith, Nesheim, and O'Connell, 2018; Miravete, Seim, and Thurk, 2023; Kaneko and Toyama, 2025). It would be interesting to relax these two assumptions.

Appendix A Mathematical Appendix

This appendix provides mathematical notation and definitions used in the appendices, as well as fundamental properties of homogeneous and homothetic functions.

The following sets, vectors, and matrices are used throughout: $\mathbf{x} \in \mathbb{R}^{J+1}$ refers to the vector $(x_0, \ldots, x_J)^\intercal$; $\Delta_J \subset \mathbb{R}^{J+1}$ denotes the probability simplex $\Delta_J = \{\mathbf{x} \in \mathbb{R}_+^{J+1} : \sum_{k=0}^{J} x_k = 1\}$; Δ_J^+ and Δ_J° represent the relative interior and boundary of Δ_J, respectively; $\mathbf{0} \in \mathbb{R}^{J+1}$ and $\mathbf{1} \in \mathbb{R}^{J+1}$ denote the vectors of zeros and ones, respectively; $\mathbf{I} \in \mathbb{R}^{(J+1)\times(J+1)}$ denotes the identity matrix.

For functions, I use the following notation: f refers to a real-valued function; $\mathbf{F} = (F_0, \ldots, F_J)$ refers to a vector-valued function, where each F_j, $j = 0, \ldots, J$, is a real-valued function; $\nabla_{\mathbf{x}} f(\mathbf{x}) = [\partial f(\mathbf{x}) / \partial x_i]_{0 \leq i \leq J}$ denotes the gradient of f with respect to $\mathbf{x}$; $\nabla_{\mathbf{x}}^2 f(\mathbf{x}) = [\partial^2 f(\mathbf{x}) / \partial x_i \partial x_j]_{0 \leq i,j \leq J}$ denotes the Hessian matrix of f with respect to $\mathbf{x}$; $\mathbf{J}_{\mathbf{F}}^{\mathbf{x}}(\mathbf{x}) = [\partial F_i(\mathbf{x}) / \partial x_j]_{0 \leq i,j \leq J}$ denotes the Jacobian matrix of $\mathbf{F}$ with respect to $\mathbf{x}$; f is continuously differentiable or twice continuously differentiable if its first or second derivatives exist and are continuous, respectively. $\mathbf{F}$ is continuously differentiable or twice continuously differentiable if each of its components F_j is continuously differentiable or twice continuously differentiable, respectively.

Let $f : \mathbb{R}^{J+1}_{++} \rightarrow \mathcal{Y}_f \subseteq \mathbb{R}$ and $\mathbf{F} : \mathbb{R}^{J+1}_{++} \rightarrow \mathcal{Y}_F \subseteq \mathbb{R}^{J+1}$. f is *homogeneous* of degree $\mu > 0$ if $f(\lambda \mathbf{x}) = \lambda^{\mu} f(\mathbf{x})$ for all $\lambda > 0$ and $\mathbf{x} \in \mathbb{R}^{J+1}_{++}$. $\mathbf{F}$ is homogeneous of degree $\mu > 0$ if each of its components F_j is. f is *homothetic* if $f(\mathbf{x}) = f(\mathbf{y}) \Rightarrow f(\lambda \mathbf{x}) = f(\lambda \mathbf{y})$ for all $\lambda > 0$ and $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{J+1}_{++}$. $f : \mathbb{R}^{J+1}_{++} \rightarrow \mathbb{R}_{++}$ is *log-convex* if $\ln f$ is convex.

The following lemma states key properties of homogeneous and homothetic functions.

Lemma 2. Let $f : \mathbb{R}^{J+1}_{++} \rightarrow \mathcal{Y}_f \subseteq \mathbb{R}$ be an increasing differentiable function.

1. f is homogeneous of degree $\mu > 0$ if and only if
$$\sum_{k \in \mathcal{J}} \nabla_{x_k} f(\mathbf{x}) x_k = \mu f(\mathbf{x}), \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^{J+1}_{++}.$$
2. f is homothetic if and only if there exist a function g homogeneous of some degree $\mu > 0$ and an increasing function F such that $f(\mathbf{x}) = F\left(g(\mathbf{x})\right)$ for all $\mathbf{x} \in \mathbb{R}^{J+1}_{++}$.
3. f is homothetic if and only if there exist a function g homogeneous of some degree $\mu > 0$ and an increasing function F such that
$$\sum_{k \in \mathcal{J}} \nabla_{x_k} f(\mathbf{x}) x_k = \mu \frac{F^{-1}(y)}{(F^{-1})'(y)}, \quad \text{for all} \quad \mathbf{x} \in \mathbb{R}^{J+1}_{++},$$
where $y = f(\mathbf{x}) = F\left(g(\mathbf{x})\right)$.

Proof. For Part 1, see proof of Proposition 4.3 in Carter (2001, p. 491). For Part 2, see proof of Proposition 3.13 in Carter (2001, p. 357), which can be found in the book resources available at the MIT Press webpage for the book (solutions of exercises 3.174 and 3.175 in Carter_SM.pdf). For Part 3, see proof of Theorem 1 in McElroy (1969). $\square$

Part 1 states that any homogeneous function satisfies Euler's theorem. Parts 2 and 3 assert that any homothetic function can be represented as an increasing transformation of a homogeneous function and satisfies another Euler's theorem, respectively.

Appendix B Proofs for Section 3

This appendix provides complete proofs for the representation theorem established in Section 3.1 and its corollary discussed in Subsection 3.2. For clarity, Appendix B.1 presents the idea and main outlines of the proof, while Appendix B.2 provides detailed proofs of all intermediate lemmas. Then, Appendix B.3 presents and proves the corollary of the representation theorem. Furthermore, Appendix B.4 offers detailed explanations and proofs specific to the GEV models, Appendix B.5 provides the expressions for demand elasticities and diversion ratios. Appendix B.6 provide the proofs related to identification.

B.1 Representation Theorem: Idea and Main Outlines

First, I note that (P·4') and (P·5) imply the existence of a twice continuously differentiable function $CS : \mathbb{R}^{J+1} \to \mathbb{R}$, referred to as consumer surplus, such that

$$\boldsymbol{\sigma}(\boldsymbol{\delta}) = \nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta}), \quad \text{for all} \quad \boldsymbol{\delta} \in \mathbb{R}^{J+1}. \tag{15}$$

Indeed, any continuously differentiable vector-valued function with a symmetric Jacobian matrix can be expressed as the gradient of some real-valued function. If $F : \mathbb{R}^{J+1} \to \mathbb{R}^{J+1}$ is continuously differentiable with a Jacobian matrix that is everywhere symmetric, then there exists a twice continuously differentiable function $f : \mathbb{R}^{J+1} \to \mathbb{R}$ such that $\nabla_{\mathbf{x}} f(\mathbf{x}) = F(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^{J+1}$ (Sandholm, 2010, Section 3.A.9, p. 108). When $F : \mathbb{R}^{J+1} \to \Delta_J^+$, f satisfies additional properties to guarantee that F maps to Δ_J^+.

Then, I show that the consumer surplus function satisfies the following properties.

Lemma 3. Any consumer surplus function $CS : \mathbb{R}^{J+1} \to \mathbb{R}$ defined by (15), where the choice probability function $\boldsymbol{\sigma} : \mathbb{R}^{J+1} \to \Delta_J^+$ is a continuously differentiable function that satisfies Assumption 1, possesses the following properties: for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$

(CS·1) CS is convex function in $\boldsymbol{\delta}$.

(CS·2) $\nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta}) \in \Delta_J^+$.

(CS·3) For all $j \in \mathcal{J}$, $\lim_{\delta_j \to +\infty} \nabla_{\delta_j} CS(\boldsymbol{\delta}) = 1$

(CS·4) For all $c \in \mathbb{R}$, $CS\left(\boldsymbol{\delta} + c\mathbf{1}\right) = CS\left(\boldsymbol{\delta}\right) + c$.

Furthermore, I express the consumer surplus $CS(\boldsymbol{\delta})$ as a function of the exponentiated vector $e^{\boldsymbol{\delta}}$ of net qualities. That is, let $\widetilde{CS} : \mathbb{R}_{++}^{J+1} \to \mathbb{R}$ be such that

$$\widetilde{CS}\left(e^{\boldsymbol{\delta}}\right) = CS\left(\boldsymbol{\delta}\right), \quad \text{for all} \quad \boldsymbol{\delta} \in \mathbb{R}^{J+1}. \tag{16}$$

The following lemma states that any consumer surplus is homothetic function in the exponentiated vector of net qualities.

Lemma 4. Any function $\widetilde{CS} : \mathbb{R}^{J+1}_{++} \to \mathbb{R}$ defined by (16), where $CS : \mathbb{R}^{J+1} \to \mathbb{R}$ satisfies (CS·1) – (CS·4) for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$, is homothetic in $e^{\boldsymbol{\delta}}$.

By Lemma 2 (part 2), the homotheticity of $\widetilde{CS}$ implies the existence of a function H homogeneous of some degree $\mu > 0$ and an increasing function ϕ such that

$$\widetilde{CS}\left(e^{\boldsymbol{\delta}}\right) = \phi\left[H(e^{\boldsymbol{\delta}})\right], \quad \text{for all} \quad e^{\boldsymbol{\delta}} \in \mathbb{R}^{J+1}_{++}.$$

Further, by Lemma 2 (part 3), $\widetilde{CS}$ satisfies Euler's theorem for homothetic functions

$$\sum_{k\in\mathcal{J}} \nabla_{e^{\delta_k}} \widetilde{CS}\left(e^{\boldsymbol{\delta}}\right) e^{\delta_k} = \mu \frac{\phi^{-1}(y)}{\left(\phi^{-1}\right)'(y)} = 1, \quad \text{for all} \quad e^{\boldsymbol{\delta}} \in \mathbb{R}^{J+1}_{++},$$

where $y = \widetilde{CS}(e^{\boldsymbol{\delta}})$ and where the 2nd equality uses (CS·2).

This last equality can be conveniently rewritten as $\left(\phi^{-1}\right)'(y) = \mu\phi^{-1}(y)$, which is a well-known first-order differential equation whose general solution is $\phi^{-1}(y) = ce^{\mu y}$, where $c > 0$ is an arbitrary constant. Note that $c > 0$ because $\phi^{-1}(y) = (\phi^{-1})'(u)/\mu$ must be positive as both μ and $(\phi^{-1})'$ are positive.

Consequently, the function ϕ is given by $\phi(v) = \frac{1}{\mu}(\ln(v) - \ln(c))$, where c can be normalized to 1 without loss of generality, since the consumer surplus function is unique up to an additive constant. This shows that any twice continuously differentiable consumer surplus function satisfying (CS·1) – (CS·4) for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$ can be expressed as (4), and, by Roy's identity, that any implied choice probability function is characterized by (3), where the generating function $H(e^{\boldsymbol{\delta}})$ satisfies some properties stated in the lemma below. This concludes the proof as Theorem 1 follows from Lemmas 3, 4 and 5.

Lemma 5. Any function $CS : \mathbb{R}^{J+1} \to \mathbb{R}$ satisfying (CS·1) – (CS·4) for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$ can be expressed as (4) with an implied choice probability function $\boldsymbol{\sigma} : \mathbb{R}^{J+1} \to \Delta^+_J$ characterized by (3) where the function $H : \mathbb{R}^{J+1}_{++} \to \mathbb{R}_{++}$ is a twice continuously differentiable function satisfying, for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$, the properties (H·1) – (H·5) specified in Theorem 1.

B.2 Representation Theorem: Proofs

Proof of Lemma 3.

(CS·1). Differentiating $\nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta}) = \boldsymbol{\sigma}(\boldsymbol{\delta})$ with respect to $\boldsymbol{\delta}$ leads to $\nabla^2_{\boldsymbol{\delta}} CS(\boldsymbol{\delta}) = \mathbf{J}^{\boldsymbol{\delta}}_{\boldsymbol{\sigma}}(\boldsymbol{\delta})$.

By (P·4'), $\mathbf{J}_{\boldsymbol{\sigma}}^{\boldsymbol{\delta}}(\boldsymbol{\delta})$ is positive semi-definite. Therefore, $\nabla_{\boldsymbol{\delta}}^2 CS(\boldsymbol{\delta})$ is positive semi-definite and thus $CS(\boldsymbol{\delta})$ is convex in $\boldsymbol{\delta}$.

(CS·2) and (CS·3). This directly follows from (P·1) and (P·2), respectively.

(CS·4). Let $f(\boldsymbol{\delta}, c) = CS(\boldsymbol{\delta} + \mathbf{1}c) - CS(\boldsymbol{\delta}) - c$ for all $\boldsymbol{\delta} \in \mathbb{R}^{J+1}$ and $c \in \mathbb{R}$. Then,

$$\frac{\partial f(\boldsymbol{\delta}, c)}{\partial \delta_j} = \frac{\partial CS(\boldsymbol{\delta} + \mathbf{1}c)}{\partial \delta_j} - \frac{\partial CS(\boldsymbol{\delta})}{\partial \delta_j} = \sigma_j(\boldsymbol{\delta} + \mathbf{1}c) - \sigma_j(\boldsymbol{\delta}) = 0, \quad j \in \mathcal{J},$$
$$\frac{\partial f(\boldsymbol{\delta}, c)}{\partial c} = \sum_{j \in \mathcal{J}} \frac{\partial CS(\boldsymbol{\delta} + \mathbf{1}c)}{\partial(\delta_j + c)} \frac{\partial(\delta_j + c)}{\partial c} - 1 = \sum_{j \in \mathcal{J}} \sigma_j(\boldsymbol{\delta} + \mathbf{1}c) - 1 = \sum_{j \in \mathcal{J}} \sigma_j(\boldsymbol{\delta}) - 1 = 0,$$

where, for the derivatives with respect to δ_j, the second equality uses (15) and the third equality uses (P·3), and, for the derivative with respect to c, the second equality uses (15), the third equality uses (P·3), and the fourth equality uses (P·1).

As a result, the function f is equal to a constant $\tilde{c}$, i.e., $f(\boldsymbol{\delta}, c) = CS(\boldsymbol{\delta} + \mathbf{1}c) - CS(\boldsymbol{\delta}) - c = \tilde{c}$, which, by setting $c = 0$, shows that $\tilde{c} = 0$, and thus $f(\boldsymbol{\delta}, c) = 0$, i.e., $CS(\boldsymbol{\delta} + \mathbf{1}c) = CS(\boldsymbol{\delta}) + c$. □

Proof of Lemma 4. Let $\widetilde{CS}\left(e^{\boldsymbol{\delta}}\right) = \widetilde{CS}\left(e^{\boldsymbol{\delta}'}\right)$ for all $e^{\boldsymbol{\delta}}, e^{\boldsymbol{\delta}'} \in \mathbb{R}_{++}^{J+1}$. Then, $\widetilde{CS}$ is homothetic in $e^{\boldsymbol{\delta}}$, as for any $\lambda > 0$ and $e^{\boldsymbol{\delta}}, e^{\boldsymbol{\delta}'} \in \mathbb{R}_{++}^{J+1}$,

$$\begin{aligned}
\widetilde{CS}\left(\lambda e^{\boldsymbol{\delta}}\right) &= \widetilde{CS}\left(e^{\boldsymbol{\delta} + (\ln \lambda)\mathbf{1}}\right) = CS\left(\boldsymbol{\delta} + (\ln \lambda)\mathbf{1}\right) = CS\left(\boldsymbol{\delta}\right) + \ln \lambda = \widetilde{CS}\left(e^{\boldsymbol{\delta}}\right) + \ln \lambda \\
&= \widetilde{CS}\left(e^{\boldsymbol{\delta}'}\right) + \ln \lambda = CS\left(\boldsymbol{\delta}'\right) + \ln \lambda = CS\left(\boldsymbol{\delta}' + (\ln \lambda)\mathbf{1}\right) \\
&= \widetilde{CS}\left(e^{\boldsymbol{\delta}' + (\ln \lambda)\mathbf{1}}\right) = \widetilde{CS}\left(\lambda e^{\boldsymbol{\delta}'}\right),
\end{aligned}$$

where the 3rd and 7th equalities use (CS·4). □

Proof of Lemma 5.

(H·1). This follows from the derivation of the differential equation in Subsection B.1.

(H·5). By (CS·1), $\widetilde{CS}$ is a twice continuously differentiable convex function of $\boldsymbol{\delta}$. Then, because $\mu > 0$, $\mu\widetilde{CS}(e^{\boldsymbol{\delta}}) = \ln H(e^{\boldsymbol{\delta}})$ is twice continuously differentiable convex in $\boldsymbol{\delta}$, which implies that $H(e^{\boldsymbol{\delta}})$ is twice continuously differentiable log-convex in $\boldsymbol{\delta}$.

(H·2). As explained in Subsection B.1, this follows from applying Lemma 2 (part 2) to $\widetilde{CS}$, which, by Lemma 4, is homothetic in $e^{\boldsymbol{\delta}}$.

(H·3). By (CS·2), $\lim_{\delta_j \to +\infty} \nabla_{\delta_j} CS(\boldsymbol{\delta}) = \lim_{\delta_j \to +\infty} \frac{e^{\delta_j} \nabla_{e^{\delta_j}} H(e^{\boldsymbol{\delta}})}{H(e^{\boldsymbol{\delta}})} = 1$, i.e., $\lim_{e^{\delta_j} \to +\infty} \nabla_{\delta_j} CS(\boldsymbol{\delta}) = 1$. By (H·1) and (H·5), $\nabla_{e^{\delta_j}} H(e^{\boldsymbol{\delta}})$ is positive and increasing, so $\lim_{e^{\delta_j} \to +\infty} \nabla_{e^{\delta_j}} H(e^{\boldsymbol{\delta}})$ cannot equal zero. This fact implies that $\lim_{e^{\delta_j} \to +\infty} H(e^{\boldsymbol{\delta}}) = +\infty$.

(H·4). $\nabla_{e^{\delta_j}} H(e^{\boldsymbol{\delta}}) = \mu \nabla_{\delta_j} CS(\boldsymbol{\delta}) e^{\mu CS(\boldsymbol{\delta})} e^{-\delta_j} > 0$ because of (CS·2) and $\mu > 0$. □

Proof of Theorem 1. The "Then" and "Moreover" part follow directly from Lemmas 3, 4 and 5. The "conversely" part is straightforward. □

B.3 Representation Theorem: Corollary

B.3.1 Corollary

Define a function $\mathbf{h}\left(e^{\boldsymbol{\delta}}\right) = \left(h_0\left(e^{\boldsymbol{\delta}}\right), \ldots, h_J\left(e^{\boldsymbol{\delta}}\right)\right)$, with h_j given by

$$h_j\left(e^{\boldsymbol{\delta}}\right) = \frac{1}{\mu} \frac{\partial H\left(e^{\boldsymbol{\delta}}\right)}{\partial \delta_j}, \quad j \in \mathcal{J}. \tag{17}$$

The following lemma provides the restrictions satisfied by $\mathbf{h}(e^{\boldsymbol{\delta}})$.

Lemma 6. The function $\mathbf{h}\left(e^{\boldsymbol{\delta}}\right)$ defined over $\mathbb{R}_{++}^{J+1}$ is continuously differentiable and satisfies the following properties: for all $e^{\boldsymbol{\delta}} \in \mathbb{R}_{++}^{J+1}$,

(h·1) For all $j \in \mathcal{J}$, $h_j(e^{\boldsymbol{\delta}}) > 0$.

(h·2) For all $j \in \mathcal{J}$, $\lim_{e^{\delta_j} \to +\infty} h_j(e^{\boldsymbol{\delta}}) = +\infty$.

(h·3) $\mathbf{h}(e^{\boldsymbol{\delta}})$ is homogeneous of degree $\mu > 0$ in $e^{\boldsymbol{\delta}}$.

(h·4) The derivative matrix $\left[\frac{\partial h_i\left(e^{\boldsymbol{\delta}}\right)}{\partial \delta_j}\right]_{0 \leq i,j \leq J}$ is symmetric and positive semi-definite.

This transformation yields an alternative representation result, which states that any choice probability function satisfying Assumption 1 can be expressed in a form that directly generalizes the logit choice probability function through the function $\mathbf{h}$.

Corollary 1. If the choice probability function $\boldsymbol{\sigma}(\boldsymbol{\delta})$ satisfies Assumption 1, then there exist a constant $\mu > 0$ and a continuously differentiable function $\mathbf{h}\left(e^{\boldsymbol{\delta}}\right) = \left(h_0\left(e^{\boldsymbol{\delta}}\right), \ldots, h_J\left(e^{\boldsymbol{\delta}}\right)\right)$ defined over $\mathbb{R}_{++}^{J+1}$ such that

$$\sigma_j(\boldsymbol{\delta}) = \frac{h_j\left(e^{\boldsymbol{\delta}}\right)}{\sum_{k \in \mathcal{J}} h_k\left(e^{\boldsymbol{\delta}}\right)}, \quad j \in \mathcal{J}. \tag{18}$$

Moreover, the function $\mathbf{h}\left(e^{\boldsymbol{\delta}}\right)$ satisfies properties (h·1) – (h·4), for all $e^{\boldsymbol{\delta}} \in \mathbb{R}_{++}^{J+1}$. Furthermore, the consumer surplus function is given by

$$CS(\boldsymbol{\delta}) = \frac{1}{\mu} \ln \left(\sum_{k \in \mathcal{J}} h_k \left(e^{\boldsymbol{\delta}}\right) \right). \tag{19}$$

Conversely, if there exist a constant μ and a continuously differentiable function $\mathbf{h}\left(e^{\boldsymbol{\delta}}\right)$ defined over $\mathbb{R}_{++}^{J+1}$ that satisfies properties (h·1) – (h·4) for all $e^{\boldsymbol{\delta}} \in \mathbb{R}_{++}^{J+1}$, then the choice probability function $\boldsymbol{\sigma}(\boldsymbol{\delta})$ defined by (18) satisfies Assumption 1.

B.3.2 Proofs

Proof of Lemma 6.

(h·1). $h_j(e^{\boldsymbol{\delta}}) = \frac{1}{\mu} \nabla_{\delta_j} H(e^{\boldsymbol{\delta}}) = \frac{1}{\mu} e^{\delta_j} \nabla_{e^{\delta_j}} H(e^{\boldsymbol{\delta}})$, so that, by (H·4), $h_j(e^{\boldsymbol{\delta}}) > 0$.

(h·2). This follows directly from $\nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta}) = \mathbf{h}(\boldsymbol{\delta})/H(e^{\boldsymbol{\delta}})$, i.e., $\mathbf{h}(\boldsymbol{\delta}) = \nabla_{\boldsymbol{\delta}} CS(\boldsymbol{\delta}) H(e^{\boldsymbol{\delta}})$ and thus $h_j(\boldsymbol{\delta}) = \nabla_{\delta_j} CS(\boldsymbol{\delta}) H(e^{\boldsymbol{\delta}})$ where $\lim_{e^{\delta_j} \to +\infty} \nabla_{\delta_j} CS(\boldsymbol{\delta}) = 1$ (by (CS·2)) and $\lim_{e^{\delta_j} \to +\infty} H(e^{\boldsymbol{\delta}}) = +\infty$ (by (H·3)).

(h·3). h_j is homogeneous of degree $\mu > 0$ since, for all $\lambda > 0$ and all $e^{\boldsymbol{\delta}} \in \mathbb{R}_{++}^{J+1}$,

$$h_j\left(\lambda e^{\boldsymbol{\delta}}\right) = \frac{1}{\mu} \nabla_{\delta_j} H\left(\lambda e^{\boldsymbol{\delta}}\right) = \frac{1}{\mu} \nabla_{\delta_j} \left[\lambda^{\mu} H\left(\lambda e^{\boldsymbol{\delta}}\right)\right] = \lambda^{\mu} \left(\frac{1}{\mu} \nabla_{\delta_j} H\left(e^{\boldsymbol{\delta}}\right)\right) = \lambda^{\mu} h_j\left(e^{\boldsymbol{\delta}}\right),$$

where the 2nd equality uses that H is homogeneous of degree μ (by (H·3)).

(h·4). This follows from (H·1) and the fact that $\mathbf{J}_{\mathbf{h}}^{\boldsymbol{\delta}}(e^{\boldsymbol{\delta}}) = \nabla_{\boldsymbol{\delta}}^2 H(e^{\boldsymbol{\delta}})$. Indeed, by Young's theorem (Carter, 2001, Theorem 4.2), $\mathbf{J}_{\mathbf{h}}^{\boldsymbol{\delta}}(e^{\boldsymbol{\delta}})$ is symmetric as it is the Hessian matrix of the twice continuously differentiable function $H(e^{\boldsymbol{\delta}})$. $\mathbf{J}_{\mathbf{h}}^{\boldsymbol{\delta}}(e^{\boldsymbol{\delta}})$ is positive semi-definite as it is the Hessian matrix of the function $H(e^{\boldsymbol{\delta}})$, which, by (CS·1), is convex in $\boldsymbol{\delta}$. □

Proof of Corollary 1. The functional forms for the choice probability function (18) and consumer surplus functions (19) directly follows from

$$H(e^{\boldsymbol{\delta}}) = \frac{1}{\mu} \sum_{k \in \mathcal{J}} e^{\delta_k} \nabla_{e^{\delta_k}} H\left(e^{\boldsymbol{\delta}}\right) = \sum_{k \in \mathcal{J}} \frac{1}{\mu} \nabla_{\delta_k} H\left(e^{\boldsymbol{\delta}}\right) = \sum_{k \in \mathcal{J}} h_k\left(e^{\boldsymbol{\delta}}\right),$$

where the 1st equality applies Euler's theorem to the homogeneous function H (Lemma 2, part 1). The properties of $\mathbf{h}$ then follows from Lemma 6. The "conversely" part is straigthforward. $\square$

B.4 Proofs and Details about GEV Models

B.4.1 (G·2) and (G·4) imply (H·2) and (H·5)

By Lemma 2 (part 2), the homogeneity of $G\left(e^{\boldsymbol{\delta}}\right)$ (G·2) in $e^{\boldsymbol{\delta}}$, implies that $\ln G\left(e^{\boldsymbol{\delta}}\right)$ is homothetic in $e^{\boldsymbol{\delta}}$. Then, by Lemma 2 (part 3), $\ln G(e^{\boldsymbol{\delta}})$ satisfies

$$\sum_{j\in\mathcal{J}}\frac{\partial \ln G(e^{\boldsymbol{\delta}})}{\partial \delta_j}=\mu.$$

Differentiating this equation with respect to δ_k, $k\in\mathcal{J}$, and rearranging it, we obtain

$$\frac{\partial^2 \ln G(e^{\boldsymbol{\delta}})}{\partial \delta_k^2}=-\sum_{j\in\mathcal{J}\setminus\{k\}}\frac{\partial^2 \ln G(e^{\boldsymbol{\delta}})}{\partial \delta_j\partial\delta_k},\quad k\in\mathcal{J}.$$

Furthermore, the second-order derivatives of the right-hand side of the above equation are negative (G·4). This implies that the Hessian of $\ln G(e^{\boldsymbol{\delta}})$ is a weakly diagonally dominant matrix with non-negative diagonal entries, which, in turn, implies that $\ln G(e^{\boldsymbol{\delta}})$ is convex in $\boldsymbol{\delta}$, that is, $G(e^{\boldsymbol{\delta}})$ is log-convex in $\boldsymbol{\delta}$ (H·4).

B.4.2 (H·2) and (H·5) does not necessarily imply (G·2) and (G·4)

I show this using example of H satisfying (H·2) and (H·5) but not (G·4). Consider

$$H(e^{\boldsymbol{\delta}})=\left(e^{\delta_0}+e^{\delta_1}\right)+\left(e^{\delta_0}+e^{\delta_2}\right)+\left(e^{\delta_1/\mu_{12}}+e^{\delta_2/\mu_{12}}\right)^{\mu_{12}}.$$

This generating function satisfies (H·1) – (H·5) for all $e^{\boldsymbol{\delta}}\in\mathbb{R}^{J+1}_{++}$ when $\mu_{12}>0$.[31] But,

$$\frac{\partial H(e^{\boldsymbol{\delta}})}{\partial e^{\delta_1}\partial e^{\delta_2}}=\frac{\mu_{12}-1}{\mu_{12}}\left(e^{\delta_1/\mu_{12}}+e^{\delta_2/\mu_{12}}\right)^{\mu_{12}-2},$$

is positive when $\mu_{12}>1$. This shows that a generating function satisfying (H·2) and (H·5) may not necessarily satisfy (G·4).

[31] This is the paired combinatioral logit generating function with $J=2$, $\mu_{01}=\mu_{02}=0$ and $\mu_{12}>1$. This can be shown by applying the proof below for the cross nested logit model to the paired combinatorial logit model, as the latter is a special case of the former

B.4.3 Table 1: Models

Cross nested logit (CNL) model. The CNL model is a specific GEV model, which admits a nesting structure allowing for all possible nests and partial nest membership. Following Bierlaire (2006), the CNL generating function is as given in Table 1, where $\mathcal{M}$ is the set of all possible nests, m is the index for nests with nesting parameter σ_m, α_{jm} are membership parameters, and σ is an additional demand parameter. Bierlaire (2006) shows the CNL generating function satisfies the GEV restrictions under the parameter restrictions: $\alpha_{jm} \geq 0$ for all $j \in \mathcal{J}$ and $m \in \mathcal{M}$; $\sum_{m \in \mathcal{M}} \alpha_{jm} > 0$ for all $j \in \mathcal{J}$; $\sigma > 0$; $\sigma_m > 0$ and $\sigma \leq \sigma_m$ for all $m \in \mathcal{M}$.

Paired combinatorial logit (PCL) model. The PCL model admits a nesting structure that allocates every pair of products into a nest. Following Koppelman and Wen (2000), the PCL generating function is as given in Table 1. The PCL generating function satisfies the GEV restrictions if $0 < \mu_{jk} \leq 1$ for all $j \neq k$.

Product differentiation logit (PDL) model. See Subsection 3.3.

B.4.4 Table 1: Proofs

The CNL generating function is given by $H(e^{\boldsymbol{\delta}}) = \sum_{m \in \mathcal{M}} f_m(\boldsymbol{\delta})$, where

$$f_m(\boldsymbol{\delta}) = \left(\sum_{j \in \mathcal{J}} \alpha_{jm} e^{\sigma_m \delta_j} \right)^{\frac{\sigma}{\sigma_m}} \tag{20}$$

I here show that it satisfies (H·1) – (H·5) under the parameter restrictions: $\alpha_{jm} \geq 0$ for all $j \in \mathcal{J}$ and $m \in \mathcal{M}$; $\sum_{m \in \mathcal{M}} \alpha_{jm} > 0$ for all $j \in \mathcal{J}$; $\sigma > 0$; $\sigma_m > 0$ for all $m \in \mathcal{M}$. This implies that $\sigma \leq \sigma_m$ for all $m \in \mathcal{M}$ is required to satisfy the GEV restrictions (G·1) – (G·4) but not to satisfy (H·1) – (H·5).

(H·1). This follows from the fact that H is equal to a sum of zero and positive terms.

(H·2). This follows from Bierlaire (2006)'s proof of Theorem 2 [point 2] as it does not use the property that $\sigma \leq \sigma_m$ for all $m \in \mathcal{M}$.

(H·3). By Bierlaire (2006)'s proof of Theorem 2 [point 3], this follows from $\sum_{m \in \mathcal{M}} \alpha_{jm} > 0$ for all $j \in \mathcal{J}$.

(H·4). This follows from Bierlaire (2006)'s proof of Theorem 2 [point 4] as it only uses that $\sigma > 0$.

(H·5). As the sum of log-convex functions is log-convex (Boyd and Vandenberghe, 2004, Subsection 3.5.2), it suffices to show that the function f_m in (20) is log-convex in $\boldsymbol{\delta}$ under the above parameter restrictions, that is, that the following function is convex in $\boldsymbol{\delta}$

$$g_m(\boldsymbol{\delta}) = \ln f_m(\boldsymbol{\delta}) = \frac{\sigma}{\sigma_m} \ln \left(\sum_{j \in \mathcal{J}} \alpha_{jm} e^{\sigma_m \delta_j} \right). \tag{21}$$

$$h^2(\boldsymbol{\delta}) = \ln \left(\sum_{j \in \mathcal{J}} e^{\sigma_m \delta_j + \ln(\alpha_{jm})} \right) = \ln \left(\sum_{j \in \mathcal{J}} \alpha_{jm} e^{\sigma_m \delta_j} \right) = \frac{\sigma_m}{\sigma} g_m(\boldsymbol{\delta})$$

is convex in $\boldsymbol{\delta}$. Lastly, use that if h^2 is a convex function and $\alpha \geq 0$, then αh^2 is convex. With $\alpha = \sigma/\sigma_j$, it implies that $g_m(\boldsymbol{\delta})$ is convex provided that $\sigma/\sigma_j \geq 0$, i.e., provided that $\sigma_m > 0$ as $\sigma > 0$ is required for (H·4) to hold. This shows that $\sigma \leq \sigma_m$ is not required for consistency with (H·1) – (H·5).

The PCL model is the specific CNL model for which $\mathcal{M} = \{(j,k), j = 0, \ldots, J-1, k = j+1, \ldots, J\}$, $\alpha_{(j,k)} = 1$ for all $j = 0, \ldots, J-1, k = j+1, \ldots, J$, $\sigma = 1$, and $\sigma_{(j,k)} = \mu_{jk}$. The PCL model satisfies the expanded set of restrictions even when $\mu_{jk} > 1$.

The PDL model is the specific CNL model for which $\mathcal{M} = \{\mathcal{G}_{hg}, h = 1, \ldots, H, g = 1, \ldots, G_h\}$, $\alpha_{j\mathcal{G}_{hg}} = a_h$ for all $h = 1, \ldots, H$, $g = 1, \ldots, G_h$ $j = 1, \ldots, J$, $\sigma = 1$, and $\sigma_h = 1/\mu_h$ for all $h = 1, \ldots, H$. The PDL model satisfies the expanded set of restrictions even when $\mu_h > 1$.

B.5 Demand Elasticities and Diversion Ratios

To summarize substitution, own- and cross-price demand elasticities are typically reported. Based on (3), they are given by

$$\begin{aligned}
\left[\nabla_{p_k} \sigma_j(\boldsymbol{\delta})\right] \frac{p_k}{\sigma_j(\boldsymbol{\delta})} &= \left(1\{k = j\} + \frac{e^{\delta_k} \nabla^2_{e^{\delta_j}, e^{\delta_k}} H(e^{\boldsymbol{\delta}})}{\nabla_{e^{\delta_j}} H(e^{\boldsymbol{\delta}})} - \mu \sigma_k(\boldsymbol{\delta}) \right) \left[\nabla_{p_k} \delta_k\right] p_k, \\
\left[\nabla_{p_k} \sigma_j(\boldsymbol{\delta})\right] \frac{p_k}{\sigma_j(\boldsymbol{\delta})} &= \left(\frac{\nabla_{\delta_k} h_j(e^{\boldsymbol{\delta}})}{h_j(e^{\boldsymbol{\delta}})} - \mu \sigma_k(\boldsymbol{\delta}) \right) \left[\nabla_{p_k} \delta_k\right] p_k.
\end{aligned}$$

Sometimes, the diversion ratios $100 \nabla_{p_k} \sigma_j(\boldsymbol{\delta}) / \left| \nabla_{p_j} \sigma_j(\boldsymbol{\delta}) \right|$ are reported. They can be directly obtained from the expression of the price elasticities just above.

B.6 Proofs for Subsection 3.4

Proof of Lemma 1. This follows from Proposition 2 in Eriksson (1986) once we note that the proof does not use stronger properties than (H·1) – (H·5). □

Proof of Proposition 1. This follows from Theorem 1 in Berry and Haile (2014) because the choice probability function (3) satisfies the index restriction and is invertible. □

Appendix C Proofs for Section 4

Proof of Proposition 2. The first part is straightforward as (11) can be viewed as a mixture of (3). For (H·1) – (H·4), this follows from Daly and Bierlaire (2006, Theorem 7). For (H·5), this follows from the same arguments as for the proof of (H·5) in the CNL model in Appendix B.4.4. □

Proof of Proposition 3. The invertibility part follows from Proposition 3 in Eriksson (1986) once we note that the proof does not use stronger properties than (H·1) – (H·5). The identification part follows from Theorem 1 in Berry and Haile (2014) because the choice probability function (11) satisfies the index restriction and is invertible. □

References

ALLEN, R. AND J. REHBECK (2019a): "Hicksian complementarity and perturbed utility models," *Economic Theory Bulletin*, 1–17.

——— (2019b): "Identification With Additively Separable Heterogeneity," *Econometrica*, 87, 1021–1054.

——— (2023): "Revealed stochastic choice with attributes," *Economic Theory*, 75, 91–112.

ANDERSON, S. P., A. DE PALMA, AND J. F. THISSE (1992): *Discrete choice theory of product differentiation*, Cambridge, MA: MIT press.

ARMSTRONG, M. AND J. VICKERS (2015): "Which demand systems can be generated by discrete choice?" *Journal of Economic Theory*, 158, 293–307.

Ben-Akiva, M. and B. Francois (1983): "Mu-homogenous generalized extreme value model," *Working Paper, Department of Civil Engineering, MIT*.

Berry, S. and P. Haile (2016): "Identification in differentiated products markets," *Annual review of Economics*, 8, 27–52.

Berry, S. and P. Jia (2010): "Tracing the woes: An empirical analysis of the airline industry," *American Economic Journal: Microeconomics*, 2, 1–43.

Berry, S., A. Khwaja, V. Kumar, A. Musalem, K. C. Wilbur, G. Allenby, B. Anand, P. Chintagunta, W. M. Hanemann, P. Jeziorski, and M. Angelo (2014): "Structural models of complementary choices," *Marketing Letters*, 25, 245–256.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

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

——— (2021): "Foundations of demand estimation," in *Handbook of industrial organization*, Elsevier, vol. 4, 1–62.

Bhattacharya, D. (2025): "Integrability and identification in multinomial choice models," *Journal of Economic Theory*, 223, 105938.

Bierlaire, M. (2006): "A theoretical analysis of the cross-nested logit model," *Annals of operations research*, 144, 287–300.

Birchall, C., D. Mohapatra, and F. Verboven (2024): "Estimating substitution patterns and demand curvature in discrete-choice models of product differentiation," *Review of Economics and Statistics*, 1–40.

Börsch-Supan, A. (1990): "On the compatibility of nested logit models with utility maximization," *Journal of Econometrics*, 43, 373–388.

Boyd, S. and L. Vandenberghe (2004): *Convex optimization*, Cambridge, U.K.: Cambridge University Press.

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.

CARTER, M. (2001): *Foundations of mathematical economics*, MIT Press.

CONLON, C. AND J. GORTMAKER (2020): "Best practices for differentiated products demand estimation with pyblp," *The RAND Journal of Economics*, 51, 1108–1161.

——— (2025): "Incorporating micro data into differentiated products demand estimation with PyBLP," *Journal of Econometrics*, 105926.

DALY, A. AND M. BIERLAIRE (2006): "A general and operational representation of generalised extreme value models," *Transportation Research Part B: Methodological*, 40, 285–305.

DALY, A. AND S. ZACHARY (1979): "Improved multiple choice models," in *Determinants of Travel Choice*, ed. by D. Hensher and Q. Dalvi (eds), London: Teakfield, 337–357.

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.

ERIKSSON, E. A. (1986): "Generalized extreme value discrete choice demand models: existence and uniqueness of market equilibria," *Regional Science and Urban Economics*, 16, 547–572.

ERSHOV, D., J.-W. LALIBERTÉ, AND S. ORR (2023): "Mergers in a Model with Complementarity," *Accepted at RAND Journal of Economics*.

FALLY, T. (2022): "Generalized separability and integrability: Consumer demand with a price aggregator," *Journal of Economic Theory*, 203, 105471.

FENG, G., X. LI, AND Z. WANG (2018): "On substitutability and complementarity in discrete choice models," *Operations Research Letters*, 46, 141–146.

FOSGERAU, M., E. MELO, A. DE PALMA, AND M. SHUM (2020): "Discrete choice and rational inattention: A general equivalence result," *International economic review*, 61, 1569–1589.

FOSGERAU, M., E. MELO, M. SHUM, AND J. R.-V. SØRENSEN (2021): "Some remarks on CCP-based estimators of dynamic models," *Economics Letters*, 204, 109911.

FOSGERAU, M., J. MONARDO, AND A. DE PALMA (2024): "The inverse product differentiation logit model," *American Economic Journal: Microeconomics*, 16, 329–370.

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

GARRIDO, F. (2024): "An aggregative approach to pricing equilibrium among multi-product firms with nested demand," *The RAND Journal of Economics*, 55, 359–374.

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

GRIFFITH, R., L. NESHEIM, AND M. O'CONNELL (2018): "Income effects and the welfare consequences of tax in differentiated product oligopoly," *Quantitative Economics*, 9, 305–341.

GRIGOLON, L. (2021): "Blurred boundaries: A flexible approach for segmentation applied to the car market," *Quantitative Economics*, 12, 1273–1305.

GRIGOLON, L. AND F. VERBOVEN (2014): "Nested logit or random coefficients logit? A comparison of alternative discrete choice models of product differentiation," *Review of Economics and Statistics*, 96, 916–935.

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

HORTAÇSU, A., J. LIEBER, J. MONARDO, AND A. DE PAULA (2024): "Estimating Nesting Structures," *Working paper*.

HURWICZ, L. AND . UZAWA, H (1971): "On the integrability of demand functions," *In: Chipman, J.S., Hurwicz, L., Richter, M.K., Sonnenschein, H.F. (Eds.), Preferences, Utility, and Demand. Harcourt Brace Jovanovich, New York.*

IARIA, A. AND A. WANG (2025): "An Empirical Model of Quantity Discounts with Large Choice Sets," *Conditionally Accepted at the Journal of the European Economic Association*.

JACKSON, M. O. (1986): "Integration of demand and continuous utility functions," *Journal of Economic Theory*, 38, 298–312.

JAFFE, S. AND S. D. KOMINERS (2012): "Discrete choice cannot generate demand that is additively separable in own price," *Economics Letters*, 116, 129–132.

KANEKO, S. AND Y. TOYAMA (2025): "Demand Estimation with Flexible Income Effect: An Application to Pass-Through and Merger Analysis," *The Journal of Industrial Economics*, 73, 186–233.

KONING, R. H. AND G. RIDDER (2003): "Discrete choice and stochastic utility maximization," *The Econometrics Journal*, 6, 1–27.

KOPPELMAN, F. S. AND C.-H. WEN (2000): "The paired combinatorial logit model: properties, estimation and application," *Transportation Research Part B: Methodological*, 34, 75–89.

LI, L. (2018): "A general method for demand inversion," *arXiv preprint arXiv:1802.04444*.

MATĚJKA, F. AND A. MCKAY (2015): "Rational inattention to discrete choices: A new foundation for the multinomial logit model," *American Economic Review*, 105, 272–98.

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. (1978): "Modeling the Choice of Residential Location," *Spatial Interaction Theory and Planning Models*, 75–96.

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

MIRAVETE, E., K. SEIM, AND J. THURK (2023): "Elasticity and Curvature of Discrete Choice Demand Models," Tech. rep., Working paper.

NOCKE, V. AND N. SCHUTZ (2017): "Quasi-linear integrability," *Journal of Economic Theory*, 169, 603–628.

PÁL, L. AND Z. SÁNDOR (2023): "Comparing procedures for estimating random coefficient logit demand models with a special focus on obtaining global optima," *International Journal of Industrial Organization*, 88, 102950.

REYNAERTS, J., R. VARADHA, AND J. C. NASH (2012): "Enhancing the convergence properties of the BLP (1995) contraction mapping," .

ROCKAFELLAR, R. T. (1970): *Convex analysis*, Princeton university press.

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.

SANDHOLM, W. H. (2010): *Population games and evolutionary dynamics*, MIT press.

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

TRAIN, K. E. (2009): *Discrete choice methods with simulation*, New York, NY: Cambridge University Press.

WEN, C.-H. AND F. KOPPELMAN (2001): "The Generalized Nested Logit Model," *Transportation Research B*, 35, 627–641.